공학박사 학위논문

Data-driven Technology Foresight:
Text Analysis of Emerging Technologies

데이터 기반의 기술 포사이트:

유망기술의 텍스트 분석

2018년 2월

서울대학교 대학원

산업조선공학과

권 희 열

Data-driven Technology Foresight:
Text Analysis of Emerging Technologies

지도 교수 박 용 태

이 논문을 공학박사 학위논문으로 제출함

2017년 10월

서울대학교 대학원

산업조선공학과

권 희 열

권희열의 공학박사 학위논문을 인준함

2017년 12월

위 원 장 ______________________ (인)

부위원장 ______________________ (인)

위 원 ______________________ (인)

위 원 ______________________ (인)

위 원 ______________________ (인)

Abstract

Data-driven Technology Foresight: Text Analysis of Emerging Technologies

Heeyeul Kwon
Department of Industrial Engineering
The Graduate School
Seoul National University

This dissertation argues for new directions in the field of technology foresight. Technology foresight was formulated on the basis of qualitative and participatory research. Initially, most foresight activities were triggered by the prospect of a handful number of experts, but recent studies highlight theoretical paradigm shifts toward a more comprehensive and data-driven approach to creating shared insights on the future of emerging technologies. Much of the research up to now, however, has been descriptive in nature, and a definite method of realizing the notion has not yet been addressed in the existing literature to a large extent. To this end, we have attempted to formalize the concept of data-driven technology foresight by incorporating unconventional data sources – future-oriented web data, Wikipedia data, and scientific publication data – and different analytical tools – Latent Semantic Analysis, IdeaGraph, and Morphological Analysis. Four distinct foresight frameworks were proposed for the proactive management process of emerging technologies: impact identification, impact analysis, plan development, and technology ideation.

The study was guided by the following research questions: (1) what kinds of data sources are available on the web and which of those are considered useful in foresight studies? (2) Where could we incorporate these data sources and which techniques are most suitable for the given purposes? (3) Which foresight-related fields would particularly benefit from applying a data-driven approach and what are the positive effects? The proposals outlined should be considered exploratory and open-ended. It is designed to determine the nature of the problem, rather than to offer definitive and conclusive answers. Nevertheless, the proposed scheme may well provide not just a rationale but a theoretical grounding for this newly introduced notion. This dissertation is expected to yield a foothold for the readers to better comprehend and act on this new shift in the field of technology foresight.

Keywords: Technology foresight; Text analysis; Emerging technology; Technology management; Data-driven approach; Quantitative study

Student Number: 2013-31007

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Emergence of Technology Foresight

Forecasting the future of technology has always been an issue of profound importance. It is vital for dreamers who hope to innovate better products for the future – and for the public who hope to benefit from the new and the advanced. To envision what lies ahead, the notion of *technology forecasting* was commonly adopted. Technology forecasting applies to the systematic attempts for anticipating and understanding future directions, rates, characteristics, and effects of technological shifts (Firat et al., 2008). The techniques of technology forecasting include trend extrapolation (Ayres, 1969), growth curves (Martino, 1993), bibliometrics (Winsor, 1932), analogies (O'Connor, 1971), cross impact analysis (Gordon & Hayward, 1968), and system dynamics (Forrester, 1997). The methods may vary depending on the aim or the context; however, the most fundamental premise behind such a notion is that the future is pre-defined as a linear continuation of present trends (Cuhls, 2003). In result, the present state of a technology was projected into the future based on historical streams of data, assuming the past will continue to hold in the future.

Yet, only recently have managers in public and private organizations realized such a process is quite inapplicable for recently introduced technologies. Today, firms create new values through fostering more game-changing innovations, like radical, discontinuous, and disruptive innovations, and they are often associated with major technological breakthroughs. This has resulted the rise of *emerging technologies*. These technologies in the embryonic stages are not yet fully developed nor indeed extensively diffused. In fact, their very nature is quite different from that of established technologies (Halaweh, 2013; Köhler & Som, 2014). The feature is three-fold. First, emerging technology holds a remarkably high level of uncertainty due to a lagging state of technological development and an insufficient amount of knowledge (Eaton et al., 2014; Stahl, 2011). Second, it holds a complex inter-relationship with diverse actors, such as existing technologies or social norms (Köhler & Som, 2014). Various risks migrate unexpectedly and amplify with one another to surprise us with unforeseen effects (Alcock & Busby, 2006). Finally, the future of new technologies is largely shaped not only by technical but numerous underlying non-technical influences, including social, economic, environmental, and political factors. Since there is no statistics on the future of societal processes, many forecasting methods highlight the significance of effective information on emerging technologies (Coates et al., 2001; Durance & Godet, 2010).

As seen, we can no longer assume that the future is an immutable extension of the present when dealing with such unknown, unpredictable and unstable technologies (Gray, 2008). The future is now open and unpredictable. The conventional theories have highlighted the science of future-tense presents as if something *will* happen in the future. However, it is now the matter of the science of present-tense futures, considering something *would* happen in the future (Schatzmann, 2013). Furthermore, there are no “right” or “wrong” answers in forecasting, especially when the future path of a new technology is shaped by various social forces. Not only cannot such future problems be clearly defined but they also do not hold definite solution principles in the sense of definitive and objective answer (Köhler & Som, 2014; Navarro et al., 2008; Rittel & Webber, 1973; Pang, 2010). They also defy traditional linear problem-solving approaches. Rather than making one accurate prediction, we must open the widest realm of possibilities in making sense of the future through gathering as much and as diverse perspectives as possible.

This is where the notion of *technology foresight* comes into play. Rooted in the field of futures studies but quite differentiated from technology forecasting, technology foresight was introduced in the late 1990s for the purpose of foreseeing potential implications of new technological innovation upon society. One common definition was noted by EU Foren Guide, defining technology foresight as "systematic, participatory, future intelligence gathering and medium-to-long term vision-building process aimed at present-day decisions and mobilizing joint actions" (Harper, 2013). Schatzmann (2013) defined the foresight process as gaining an increased awareness about plausible alternative futures through collecting and creating future-oriented knowledge. Cuhls (2003) also formulated as "the overall process of creating an understanding and appreciation of information generated by looking ahead". However, technology foresight is still an emergent phenomenon, and there still exists a disagreement regarding the definition and the contour of the discipline (Schatzmann, 2013; Smith & Saritas, 2011).

1.2 Towards a Data-driven Technology Foresight

The definitions may seem somewhat different; however, one distinctive feature shared by most technology foresight studies is *collective intelligence.* The notion of *collective intelligence* is an emergent property from synergies among three factors: 1) data/knowledge/information; 2) software/hardware; and 3) experts and other individuals with insight (von der Gracht, 2015). It is of vital importance to ensure sufficiently diverse perspectives, interests, values, information, and knowledge for opening forward into a future of widest possibility and opportunity. Such a collective and inclusive nature is the key characteristic of technology foresight, which distinguishes it from other futures study (Saritas & Nugroho, 2012). In fact, it has always been the fundamental factor stimulating theoretical and methodological paradigm shifts within the field of technology foresight.

Conventionally, foresight activities were performed on the basis of participatory approaches, including scenario workshops, brainstorming, consensus-building, interviews, or horizon scanning (Harper, 2013). The main knowledge source was first-hand information directly collected from a handful number of experts. Such a process may ensure much of reliability and validity; however, it may be subject to many potential cognitive biases and an inefficient information gathering process. In recent years, the utilization of ICT and Web 2.0 has opened the window of opportunity of a new generation of technology foresight: *Foresight 2.0* and *Foresight Support Systems (FSS)* (Gheorghiu et al., 2008; Schatzmann, 2013; Keller et al., 2015). The rise of “interactive online platform” and “community-oriented web” facilitated a better collaboration within the group by enabling a higher interactivity, better focus of expertise, and an accumulative content development (Gheorghiu et al., 2008). According to Raford (2014), several advantages of Foresight 2.0 include an increase in the amount, a large diversity of perspectives, a very time-effective method, a transparent process, and a real-time data collection. As seen, ICT tools have altered the methodological basis of technology foresight by yielding a more efficient, embedded and continuous collaboration tools.

After a monumental buzz of Web 2.0, a recent introduction of Web 3.0 has spurred a whole new form of technology foresight: Foresight 3.0. Web 3.0 corresponds to a *semantic web*, meaning a massive amount of data are inter-linked with one another in the Web and further categorized and stored in a way that a machine can exploit and interpret. The envisioning process was no longer solely depended on a direct knowledge gathering process. The opinions and perspectives of diverse groups of people from different backgrounds are now available on the web, and a thorough analysis and interpretation may surely enrich knowledge heterogeneity and diversity. Moreover, the increasing urgency of handling an abundance of data has also resulted the use of data mining techniques. These theoretical and methodological shifts are expected to stimulate more extensive and efficient knowledge-gathering process, thus leading to an enhanced foresight performance. As illustrated in Figure 1-1, the efforts of expanding the range and scope of knowledge bases have shifted the fundamentals of technology foresight. The field was initially conducted based on a qualitative and participatory approach, but it is now shifted towards a more quantitative and data-driven approach to creating shared insights on future technology-related challenges.

Figure 1-1 An illustration of the paradigm shifts in technology foresight

1.3 Problem Statement

This whole paradigm shift within the field of technology foresight is, however, still a nascent area of research, and there remains a considerable room for reaching the notion of Foresight 3.0. Specifically, the concept of *collective intelligence* has not yet been fully achieved in many foresight activities. Whilst an aggregation of individual intelligence and further formation of a bigger future depiction is vital in foresight studies (Harper, 2013), previous studies have only used a fraction of the globally available information and experience. A variety of knowledge is now disseminated extensively and rapidly on the web, and much of them remain untouched. It is so wide-ranging that it may be categorized based on different values: types of players, perspectives, and orientations. In details, the web encompasses the opinions not just of experts but also of the general public. Some pages include a more retrospective knowledge, and some may include prospective knowledge. Furthermore, the data sources are either relatively more objective or more subjective in nature.

This research particularly highlights three particular data types expected to be served as effective knowledge sources for various foresight processes: *future-oriented web data*, Wikipedia data, and scientific publication data. First, many different future-oriented information is being flooded through the Internet, denoted as *future-oriented web data* in this article. Both experts and the general public are envisaging future technology and its potential impact on the society and further interactively sharing them through the Web. Second, Wikipedia might be the most prominent example of collective intelligence, frequently mentioned by numerous futures studies. It is considered the largest, fastest growing encyclopedia in the world (Wang & Domeniconi, 2008). It holds millions of articles covering not only historical but future-oriented subjects in all areas of human knowledge. Its significant strengths are domain-specific concepts and coherent knowledge structure. Finally, scientific publication data offers a wide variety of professionalized knowledge regarding scientific, technical, societal, or political domains. The "nuggets" of key reviews, model, or developments can be used as the solutions for various issues. In fact, such a data is suitable for deriving strategic information, which is one of the most significant values of technology foresight.

The information is out there. All we need to do is seize this momentous opportunity and properly incorporate them in actual foresight activities. However, not much of scientific investigation was carried out regarding how and where to execute such a foresight process, mainly due to its relative infancy. The research in this dissertation, in response, sought to apply the aforementioned data sources into different areas of futures studies and further demonstrate the value and benefits that data-driven technology foresight approach provide.

The main questions addressed in this dissertation are:

1. What kinds of data sources are available on the web and which of those are considered useful in foresight studies?
2. Where could we incorporate these data sources and which techniques are most suitable for the given purposes?
3. Which future-related fields would particularly benefit from adopting a data-driven foresight approach and what are the positive effects?

1.4 Dissertation Overview

The aim of this dissertation is to better exploit the opportunities presented by data and analytics assets in the field of technology foresight. For the sake of clarity, the following chapter begins by laying out the methodological and theoretical dimensions of the research. It specifically provides the general basis of the aforementioned data sources and discusses their expected strengths in each technology foresight activity. The data types include *future-oriented web data*, Wikipedia data, and scientific publication data. It further reviews theoretical background and general procedures of the methodologies incorporated in this research, including Latent Semantic Analysis (LSA), IdeaGraph, and morphological analysis. Using these data sources and methodologies, four different frameworks of data-driven technology foresight were proposed in the steps within the proactive management process of emerging technologies. The overall framework of this dissertation is illustrated in Figure 1-2.

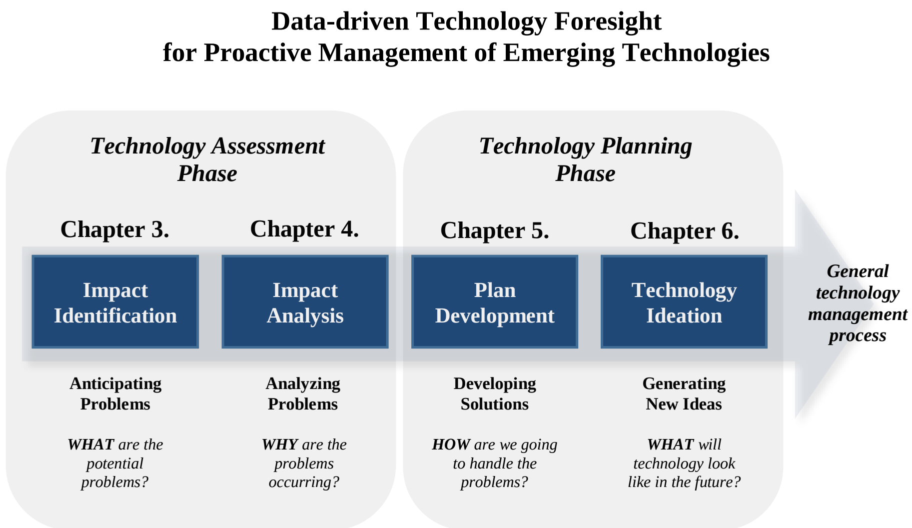


Figure 1-2 Overall framework of this dissertation

In reality, the process of engineering design does not start in isolation, and in practice, several preliminary stages precede it (Carleton & Cockayne, 2009). The proactive management process is one of the most fundamental stages in the development of emerging technologies required prior to the general technology management process, such as identification, selection, acquisition, exploitation, and protection steps (Phaal et al., 2001). The emergence of new technologies results various unforeseen threats and vulnerabilities, and technology foresight is considered vital throughout R&D value chain, particularly in anticipating, assessing, and adapting to those potential impacts (Koivisto et al., 2009). Foresight activities are capable of steering a technology development process to a more proactive and preemptive direction. The proactive management process comprises two major phases: technology assessment phase and technology planning phase. In details, potential problems associated with target emerging technologies are first identified and analyzed ex-ante in technology assessment phase. The issues are then solved proactively, and creative ideas are further generated for a better future value creation before entering into the process of product and service development.

In details, Chapter 3 identifies potential impact-related issues associated with emerging technologies and attempts to answer the question – what are the potential problems? Chapter 4 analyzes and specifies future societal impacts of emerging technologies and seeks to answer the question – why are the problems occurring? Chapter 5 suggests technical solutions to those potential issues and addresses the question – how are we going to handle the problems? Finally, Chapter 6 generates new creative and innovative ideas regarding the future development of emerging technologies and solves the question – what will technology look like in the future? These chapters discuss theoretical backgrounds, proposed frameworks, and illustrative case examples of each step of the proactive management process. Furthermore, each research conducts a comparative analysis to previous findings to define new insights as well as contradictions. Such a linkage with other literature allows to build internal validity and sharpen generalizability of the proposed result (Eisenhardt, 1989). The information regarding specific problem statements and utilized data sources and methodologies are summarized in Table 1-1. The detailed contents of each chapter are elaborated as follows.

Table 1-1 Summary of each chapter in this dissertation

	Problem Statements	**Data Sources**	**Methodologies**
Chapter 3. ***Foresight for Impact Identification***	· Incorporating the voices of the general public when exploring the societal future of an emerging technology · Efficiently gathering and analyzing the opinions of both experts and the public	· *future-oriented web data*	· LSA
Chapter 4. ***Foresight for Impact Analysis***	· Determining the right balance between creativity and plausibility when constructing future scenarios · Efficiently gathering the opinions of the many and of the few in scenario building process	· *future-oriented web data*	· LSA · IdeaGraph
Chapter 5. ***Foresight for Plan Development***	· Closing the gap between philosophizing and existing in responsible development · Offering a clear-cut methodology of efficiently capturing the perspectives of the general public and of scholars	· *future-oriented web data* · scientific publication data	· LSA
Chapter 6. ***Foresight for Technology Ideation***	· Reducing the involvement of human subjectivity, particularly in idea generation using morphological analysis, to increase the quantity of resulting ideas · Incorporating different data sources considering highly complex and uncertain nature of today's matters	· Wikipedia data	· empirical investigation · morphological analysis

Chapter 3. Foresight for Impact Identification

Applying LSA text mining technique in envisioning social impacts of emerging technologies: the case of drone technology

Chapter 3 seeks to define what emerging technologies are and further anticipate future impacts associated with their emergence based on the notion of data-driven technology foresight approach. This issue was normally handled in the field of technology assessment (TA). Technology assessment was established for developing an early warning system to detect and control negative effects exerted by new technological changes and developments (van den Ende et al., 1998). However, following questions posed by the conventional approaches must be addressed: 1) how are the future societal impacts of new technologies envisioned? 2) the general publics have recently played a critical role in shaping the future technology and society: what are the ways of incorporating their opinions in the foresight process? This research proposes a data-driven foresight process of identifying and understanding the holistic overview of emerging technologies' unintended consequences. LSA text mining technique is employed to extract multiple groups of contextually similar terms from *future-oriented web data*, and the resulting term clusters are considered new abstractions of future social impacts. Furthermore, the research acquires greater depth and breadth in conceptualizing social impacts through coupling with previous impact assessment literature. The proposed methodology gained insights into the utilization of *future-oriented web data* and text mining techniques for the foresight activities, expected to mitigate public skepticism and pursue a better social acceptance of emerging technologies.

Chapter 4. Foresight for Impact Analysis

Data-driven scenario building process using text analysis: what are the latent impacts of emerging technologies?

Chapter 4 focuses on analyzing and specifying future impacts of emerging technologies based on the notion of data-driven technology foresight approach. An in-depth visualization of future depictions was realized with scenario building process. Scenarios are alternative dynamic stories that incorporate a variety of detailed ingredients uncertain futures, rather than an accurate prediction of a single outcome (Peterson et al., 2003). However, one intrinsic limitation of scenario-related studies must be highlighted: a difficulty in generating scenarios with a balanced creativity and plausibility. This research proposes a data-driven foresight method of constructing well-balanced storylines with a much of credibility but, at the same time, a fair amount of unexpectedness. The scenarios are particularly about unforeseen ramifications of emerging technologies with much details beyond the obvious. The *future-oriented web data* was the knowledge source utilized as the ingredients for constructing the scenarios. The plausibility was captured based on frequently occurring opinions extracted from LSA. The creativity was obtained based on low-frequency terms extracted from IdeaGraph technique. The study contributed to the field of numerous scenario-related fields, such as scenario analysis, scenario planning, and scenario development.

Chapter 5. Foresight for Plan Development

Proactive development of emerging technologies in a socially responsible manner: Data-driven problem solving process using LSA

Chapter 5 aims to develop technological plans for coping with future consequences associated with emerging technologies based on the notion of data-driven technology foresight. Such an issue has been discussed in the field of responsible development, or so-called responsible innovation. Many scholars emphasize to put "responsibility" at the heart of emerging technology design to ensure public acceptance. The responsibility, herein, points out our neglect of technology's unpredictable consequences on individuals, societies, and environments in favor of immediate commercial success (Grunwald, 2011; Blok & Lemmens, 2015). However, one major drawback of this field is that the research to date had focused only on the vigorous clarification of the conceptual contours rather than the establishment of specific guidelines that could be utilized in actual business activities. This research incorporates *future-oriented web data* and scientific publication data for envisioning a set of eventualities where an emerging technology plays out to be a harmful tool to the society and suggesting essential technical requirements for the right technological preparedness, respectively. LSA text mining technique is applied to extract multiple concepts based on the terms' semantic relations. The proposed framework offered responsive and responsible insights to designers and engineers of emerging technologies.

Chapter 6. Foresight for Technology Ideation

Towards data-driven idea generation: the application of Wikipedia to morphological analysis

Chapter 6 attempts to generate creative and innovative ideas regarding the future structures of emerging technologies based on the notion of data-driven technology foresight. This was generally dealt in the field of idea generation. The generation of new and creative ideas is vital for stimulating innovations, and morphological analysis is one appropriate method in such a circumstance, given its objective, impersonal, and systematic nature. One question that needs to be asked, however, is whether there exist other useful data sources, other than patent or scientific literature, for a better idea generation process. This research, in response, expands the data set to a new boundary to stimulate more systematic process and more specialized and practical idea generation. Given its case-specific characteristics and well-coordinated knowledge structure, Wikipedia data is incorporated for developing two models of developing a morphological matrix. The effectiveness of the data-driven approach was demonstrated with a conventional discussion-based approach. The proposed method, in result, generated more specified and diversified morphological matrix, expected to be an essential supporting tool for developing creative ideas.

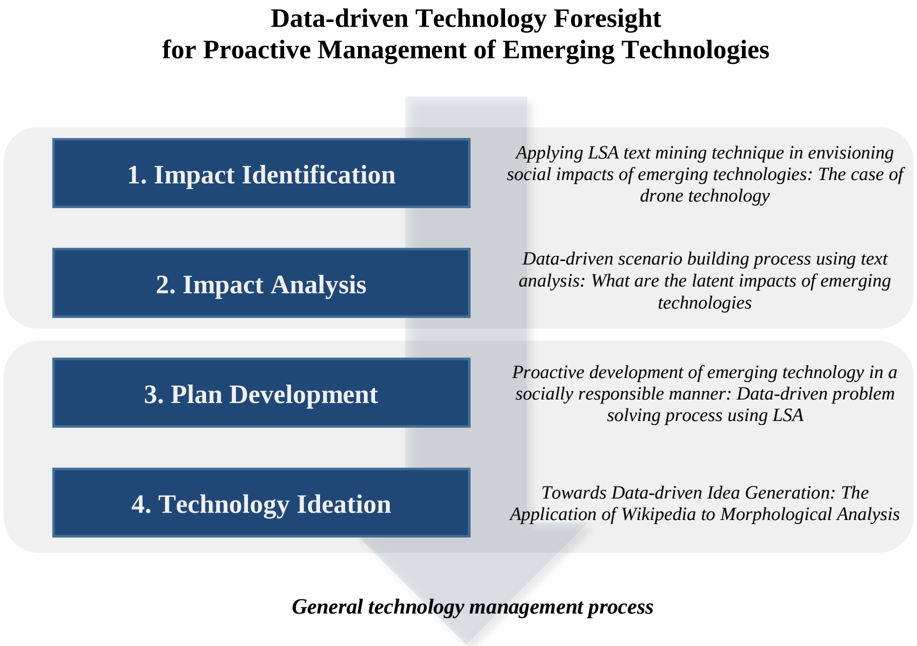


Figure 1-3 Overall structure of this dissertation

Chapter 2

Data Sources and Methodologies

2.1 Data Sources

2.1.1 Future-oriented Web Data

We are living in an era in which people are drowning in big data but craving for valuable knowledge. All sorts of information are being flooded through the Internet; however, this research particularly highlights the knowledge source with prospective and future-oriented perceptions. Every day, unsolicited, people are sharing their thoughts, insights, or even concerns regarding future technology and the society in a distant world via various online platforms, including news articles, editorials review, weblogs, social media, webzine, or online forums. The involved players are the individuals from different expertise and backgrounds, ranging from experts in science or engineering to the broader general public are interested in future invention and innovation. Such a particular knowledge source encapsulates various future-related perspectives about new technologies and socio-technical circumstance, which have not yet happened or experienced. Numerous futures scholars have discussed the emergence and the significance of the future-oriented database for a more effective foresight studies.

As stated earlier, ICT has gained prominent importance for providing a digital platform with a full of knowledge assets, but recently has it played an essential role in foresight activities (Banuls & Salmeron, 2011). Numerous futurists and decision makers consider ICT as a valuable interface. It promotes asynchronous communication and collaborative content development within foresight communities. This apparently has a profound consequence on the nature of foresight processes since foresight process itself requires intensive interaction between stakeholders (Bañuls & Salmeron, 2008; Cuhls, 2003; Van Merkerk & Smits, 2008). A research by Keller and von der Gracht (2014), for instance, identify seven ICT-based drivers which will strengthen the focus of foresight processes, such as accessibility, collaboration, efficiency, linkages, market, progress, and quantitative data-handling, thereby demonstrating the future path of ICT in foresight study (Keller & von der Gracht, 2014). Moreover, particularly in the context of scenario planning, Raford (2014) express web-based crowdsourcing as a methodological innovation for strategic foresight activity (Raford, 2014). He proposes substantial impacts of ICT on the scenario process: increased participation, increased volume and speed of data analysis, increased transparency, and decreased overall cost.

A large and growing body of literature highlights the theoretical background of digital-collaborative foresight applications. Pang (2010) reviews the psychology and neuroscience related literature and stresses the deployment of various sources, such as futurists' blogs and other publicly available data. By introducing the term Futures 2.0, Pang argues that future-oriented data source is capable of improving futures work and its impact (Pang, 2010). Similarly, Schatzmann et al. (2013) suggest transparent, efficient, and rapid generation of possible futures on the basis of open & collaborative foresight (Schatzmann et al., 2013). Von der Gracht et al. (2014) present increasing value of foresight support systems (FSS) by reviewing recent trend of incorporating ICT within foresight activities (von der Gracht, 2014). Instead of involving only the "nearby diversity", they suggest incorporating voices from a variety of perspectives, such as psychology, sociology, economy, mathematics, and even graphic artists, to capture creativity and unexpected answers for complex challenges. Tarko and Aligica (2011) further develops the original term, the virtual think tank from Herman Kahn, and suggest the prospect of ICT to extend a range of knowledge aggregation and its application in speculating various aspects of the future (Tarko & Aligica, 2011). Pereira et al. (2007) highlights the various data sources in foresight activities and denote them as *foresight knowledge*.

Recent evidence has been stressing the need for a database concerning not just technologies but also their associated society and environment, we have not yet seen. When refined with proper analytics techniques, such a database is expected to offer a more comprehensive, flexible, open, and powerful foresight process (von der Graht, 2015). This research utilizes the data source primarily for expanding the range of collective intelligence regarding non-technical aspects of future technologies, such as societal, environmental, economic, and political implications.

2.1.2 Wikipedia Data

Wikipedia is a massive online repository of collective intelligence, which allows almost anyone to become contributors to information development (Milne et al., 2007; Mihalcea & Csomai, 2007). In fact, Wikipedia is today the largest, fastest growing encyclopedia in the world (Wang & Domeniconi, 2008; Milne et al., 2007). Wikipedia's distinguishing features are as follows: (1) extensive topic coverage; (2) up-to-dated context; (3) domain-specific description; and (4) rich semantics (Strube & Ponzetto, 2006; Joorabchi et al., 2015). Apparently, Wikipedia holds millions of articles covering subjects in all areas of human knowledge and is continuously updated with new information (Joorabchi et al., 2015). It has been widely accepted as an alternative data source for WordNet and has received significant attention from the computer science research community in recent years. For instance, Wikipedia was utilized to solve various natural language processing tasks, such as semantic relatedness (Milne et al., 2007), text classification and clustering (Wang & Domeniconi, 2008; Hu et al., 2009), text identification (Hassan et al., 2012; Joorabchi et al., 2015), and topic modeling (Ciglan & Norvag, 2010; Allahyari & Kochut, 2016). However, the database of this kind carries with them various well-known limitations. For instance, the accuracy of Wikipedia information has always been a subject of controversy due to its collaborative and crowdsourced nature (Callahan & Herring, 2011). Moreover, the connecting relations of Wikipedia articles do not always explicitly equate "related" or "hierarchical" concepts since they are prone to various types of noises, such as self-loops, direct-loops, and indirect-loops (Joorabchi et al., 2015).

Despite its limited range of applicability, Wikipedia data is a prominent example of *collective intelligence* (Schatzmann, 2013), and recently has it been mentioned in numerous futures research. Gheorghiu et al. (2008) mention that Wikipedia-like systems yield relatively reliable information in a future-related participatory approach. Moreover, Wikipedia is considered a great tool for achieving an open and collective approach in technology foresight (Cagnin & Konnola, 2014). Smith and Saritas (2011) discuss that Wikipedia is the main platform for an extensive collaboration and a crowd sourcing, which could promote more analytical foresight activities. Schatzmann (2013) asserts that it is one useful database capable of gathering individual intelligence and creating a bigger depiction. As shown, Wikipedia data is one untouched data source potentially useful in a data-driven approach of foresight activities. It not only holds a case-specific information and a well-coordinated knowledge structure but also encapsulates a wide range of knowledge bases, including both retrospective and prospective information. This research utilizes this data source primarily for idea generation process of new and emerging technologies.

2.1.3 Scientific Publication Data

Scientific publication database is a widely accessible, electronic information resource, offering a wide variety of professionalized knowledge bases, handling scientific, technical, societal, or political subjects. On single topic of interest, various opinions are suggested by scholars from many different disciplinary fields. No one can tell which opinion is superior over the other; therefore, it is necessary to extensively monitor the concepts that are primarily discussed and pointed out by those intellectuals. The systematic analysis of scientific publication data has been frequently adopted in futures research. The preliminary attempts were performed quite intuitively and qualitatively, like literature review and bibliometric approach. For instance, forecasting approaches of environmental scanning or casual layered analysis use workshop formats for gathering futuristic ideas and insights based on various data sources, like literature and news articles. This is quite a manual but systematic process of analyzing different types of data. A serious weakness with traditional means of gathering and analyzing such information, however, is that it is somewhat time-consuming and labor-intensive.

Along with such qualitative methods, quantitative approaches like text analysis and network analysis have started to gain high interest in the scientific community. The notion of mining meaningful and useful information from an abundant piles scientific disciplines and journals was first pioneered by Swanson (1986), and it was introduced by the name of literature-based discovery (LBD). Behind the basic premise that scientific publication data holds multiple isl*ands of knowledge*, new valuable information is found through identifying semantic relationships between existing concepts (Andronis et al., 2011). New concepts or knowledge were quantitatively identified from various other methodologies, such as measure of relatedness (Klavans & Boyack, 2006) and citation network (Kajikawa et al., 2007; Ittipanuvat et al., 2014).

In recent years, a more diverse set of text mining techniques was incorporated in analyzing science and technology information databases. The most widely accepted tool is called tech mining, and its applications are very wide-ranging from general technology development and management, including new product development, technology roadmapping, and technology development tracking, to various futures research, including technology opportunity forecasting, technology growth modeling, and future research anticipation (Porter & Cunningham, 2005; Islam & Miyazaki, 2010; Shibata et al., 2008; Garechana et al., 2012). As presented, scientific publication database has been actively utilized as one of the main sources of futures-related studies, and this signifies the importance such a database may play in the future-oriented analysis. In this research, scientific publication database is used for deriving technical and algorithm-related solutions, which could be used as the means of reacting or answering to the future problems.

2.2 Methodologies

2.2.1 Latent Semantic Analysis (LSA)

Latent semantic analysis (LSA), also known as latent semantic indexing (LSI) in the context of information retrieval, is a well-known mathematical and statistical method for extracting and representing textual materials into a semantic structure (Dumais et al., 1988). Numerous studies have applied LSA on several promising application areas. LSA is originally proposed as the application for information retrieval (IR). Dumais et al. (1988) and Deerwester et al. (1990) propose a new method of applying Singular Value Decomposition (SVD) for automatically indexing and retrieving relevant documents on the basis of terms offered by queries. The study demonstrated that performance with LSA is modestly superior to that with the system proposed from standard IR methods. Moreover, Berry et al. (1995) compares LSA with conventional keyword vector models for IR on information science test collections and reports that LSA performs best in terms of showing average precision results especially when not much of the shared words exist on the queries and relevant documents. Lochbaum and Streeter (1989) compared the standard vector-space model with LSA and find that combination of two techniques produces a superior performance on retrieval accuracy. In addition, it is also known for accurately reflecting human knowledge. Several studies have actually proven that words or passages derived from LSA are capable of simulating a variety of human cognitive phenomena: developmental acquisition of recognition vocabulary, word-categorization, sentence-word semantic priming, discourse comprehension, and judgments of essay quality (Landauer et al., 1998). LSA is applied to various tasks such as cross-language information retrieval (CL-LSA)

(Littman et al., 1998), essay grading (Rehder et al., 1998; Wild et al., 2005), information filtering (Dumais, 1994), source-code clustering (Maletic & Marcus, 2000), detecting plagiarism (Britt et al., 2004), and topic detection (Sidorova et al., 2008).

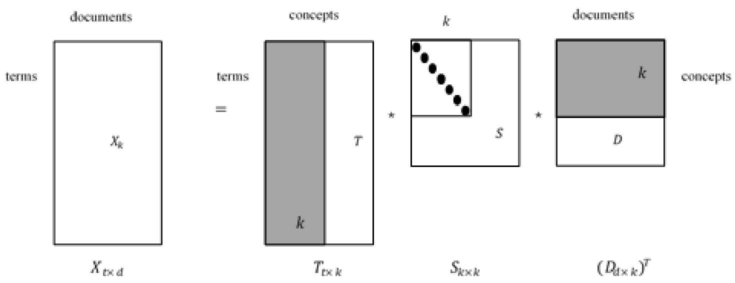


Figure 2-1 Decomposition and truncation of TDM using LSA

In details, LSA applies an algebraic technique closely related to eigenvector decomposition and factor analysis, called singular value decomposition (SVD) (Dumais et al., 1988; Forsythe, Moler, & Malcolm, 1977), as shown in Figure 2-1. LSA is performed in three steps. First, both documents and terms are represented in a vector space, or term-document matrix, which represents a co-occurrence matrix of terms and documents. Second, a term frequency-inverse document frequency (*tf-idf*) weighting is applied to the matrix in order to put more importance on the terms that characterize the content of the document but in a discriminating way (Coussement & Van den Poel, 2008). Third, SVD is used to decompose the term-document matrix (TDM), $X_{t\times d}$, into the product of three other matrices: $T_{t\times m}$, a column-orthogonal matrix with m representing dimensionality; $S_{m\times m}$, a diagonal matrix with singular values arranged in decreasing order; and $D_{d\times m}$, a transpose of the column-orthogonal matrix (Lochbaum & Streeter, 1989), where t denotes the number of terms and d denotes the number of documents. The matrices are then truncated into an arbitrary number of dimensions, k, to eliminate some of the noise existing in the original matrix and thus extracting latent semantic relationship in the collection. Figure 2-1 depicts the process of decomposition and truncation.

Table 2-1 Characteristics and limitations of text analysis techniques

Models	**Characteristics**	**Limitations**
Latent Semantic Analysis (LSA)	· Reduces dimensionality of *tf-idf* using Singular Value Decomposition · Considers synonyms of words · Allows the occurrence of words in other topics · Captures unique semantic structures	· Not built upon robust probability theory · Difficult to determine the number of topics · Cannot interpret loading values as probabilities · Hard to label a topic in certain cases based on the constituting words
Probabilistic Latent Semantic Analysis (pLSA)	· Improves LSA in a probabilistic sense by using a generative model · Captures general themes or trends in documents · Partially handles polysemy	· No probabilistic model at the level of documents
Latent Dirichelet Allocation (LDA)	· Offers topics in which words have probability values · Provides full generative model with multinomial distribution for words in topics and Dirichlet distribution over topics · Handles long-length documents	· Difficult to interpret the meaning of topics · Incapable to model relations among topics (structures of documents)
Correlated Topic Model (CTM)	· Considers relations among topics using logistic normal distribution · Allows the occurrence of words in other topics	· Requires complex computations · Contains a number of general terms in topics

Other than a discriminative model like LSA, there exist numerous generative models including probabilistic latent semantic analysis (pLSA), latent dirichlet allocation (LDA), and correlated topic model (CTM). These text analysis techniques have been developed to overcome the limitations of previous methods. However, the superiority of one methodology over another cannot be assured since the performance tends to differ depending on numerous factors, including the size or the structure of document collection. Kakkonen et al. (2006), for instance, applied LSA and pLSA in essay grading system and proves that LSA outperforms pLSA in assessing the quality of essays. The result was quite interesting in that the result was opposite to the results obtained by Hofmann (2001), and the authors noted that it was mainly due to the size of the collected documents. The characteristics and limitations are very different from one technique from another. The summarization of their characteristics, adopted from Lee et al. (2010), is presented in Table 2-1.

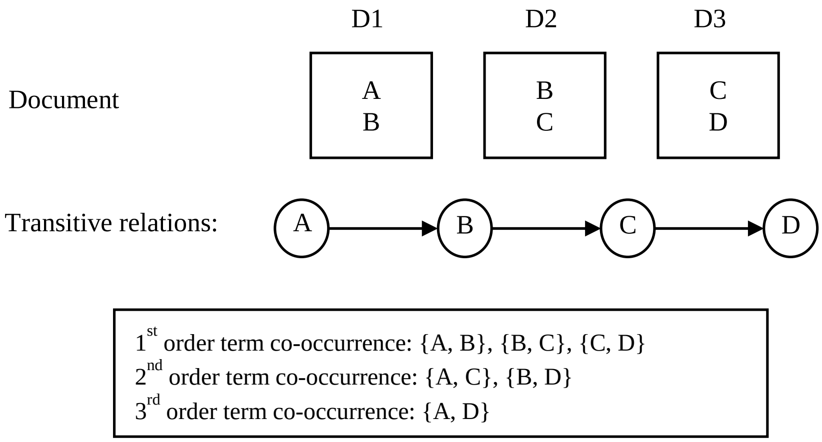


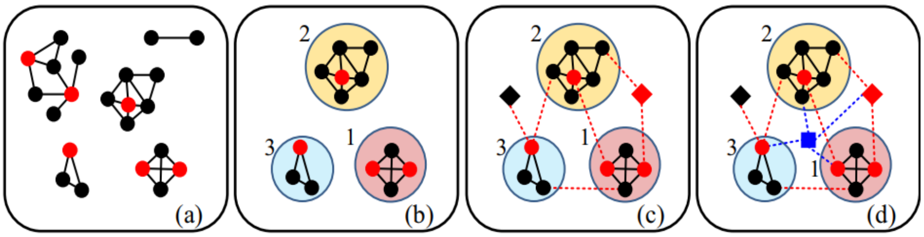
Figure 2-2 Tracing order of co-occurrence

In terms of topic generation, LSA has one distinctive feature: generation of latent topics. It is capable of capturing unique and hidden topics based on high order structures. The illustration of tracing orders of co-occurrence was implemented from Kontostathis and Pottenger (2006). As shown in Figure 2-2, the order of co-occurrence is n+1, where n is the number of hops required to connect the terms. Whilst generative models like pLSA, LDA and CTM focus mainly on shared 1st order structure {B, C}, LSA focuses primarily on higher order structures {A, C}, {B, D}, and even {A, D}. The terms, which have never co-occurred in the same document, may be estimated as closely related and appear near each other in the k-dimensional space. It seizes hidden and distinctive topics by considering indirect semantic relations, while other generative models put weights on general and apparent topics.

This dissertation chose to apply LSA since the primary aim was to construct future storylines, which require a great amount of details and unexpected components. Multiple topics are extracted from unconventional data sources, including *future-oriented web data* and scientific publication data. Furthermore, LSA is considered one suitable methodology since it is capable of retrieving conceptual topics from a pile of data in terms not only of documents but also of constituting terms (Blake & Ayyagari, 2012). Extracted documents may support the interpretation of the topics.

2.2.2 IdeaGraph

Before providing details of IdeaGraph technique, a text mining algorithm of KeyGraph must be discussed. KeyGraph is a graph-based text mining method capable of extracting meaningful but hidden information from a pile of knowledge. It has been particularly recognized in the field of Chance Discovery (CD), where it was extended the boundary from Knowledge Discovery in Texts (KDT) and focused on discovering rare but important events or situations through human-computer interaction (Wang et al., 2013). KeyGraph solved the bottleneck of human cognition by actually mining the significant knowledge and further visualizing through constructing scenario graph. This whole process was based on co-occurrence of keywords and composed of 3 main phases: 1) extracting foundations; 2) extracting columns; 3) extracting roofs. After a thorough preprocessing, the first phase constructs a keyword network made up of nodes representing terms and links representing co-occurrence. Here, the nodes are the high frequency terms and links are based on the equation below. In the second phase, specific keywords that tie and hold clusters in whole are identified based on two auxiliary functions: based and neighbors. Using these two functions, the key items are extracted based on the conditional probability-based equation, as in equation 4. In the third phase, the strength of the links between each high key term and high frequency terms within the clusters is measured. However, the methodology had few limitations: 1) does not detect the relations among low-frequency terms; 2) fails to capture low-frequency and key items inside clusters; 3) ignores direct relations between clusters; 4) does not reveal the relations among chance items. KeyGraph has been successfully applied in discovering emerging topics in the web (Matsumura et al., 2002), analyzing financial trends (Chiu et al., 2008), and triggering creative ideas (Wang et al., 2012).



(a) General clusters (b) Cognitive clusters with ranking (c) Valuable links (d) Key items

Figure 2-3 Overall process of IdeaGraph technique

To overcome such shortfalls of KeyGraph, a new graph-based technique of IdeaGraph was introduced (Wang et al., 2013). IdeaGraph is also a graph-based algorithm with the purpose of discovering latent and significant information from a pile textual data. IdeaGraph can be considered as a modified or an improved version of KeyGraph. The technique has been evolved into diverse techniques, including IdeaGraph plus, improved IdeaGraph, and timing-IdeaGraph (Zhang et al., 2014; Wang et al., 2016). The noticeable differences between these two methodologies are a more systemized process and a different measurement for capturing latent information. IdeaGraph is composed of four steps, as illustrated in Figure 2-2. First, *general clusters* are generated by conditional probability-based equation that measures relations between two terms, as stated in [2-1]. Here, I indicates a term in the equation, and the links are created and the clusters are determined based on a certain threshold value r in the graph G.

$$\mathrm{R}(I_i, I_j) = P(I_i|I_j) + P(I_j|I_i) \qquad [2\text{-}1]$$

Second, cognitive clusters are obtained by quantifying the derived *general clusters*. In order to become a cognitive one, a cluster must not only hold rich information but the size must be small enough for smooth human interpretation. Two indicators of information and information density are employed to quantify *general clusters*, as shown in [2-2] and [2-3]. As seen, information is the sum of all the links in a *general cluster*, and information density is information divided by the number of terms in the cluster. C denotes the cluster and N_e indicates the number of items of the cluster C. Then as in equation [2-4], the harmonic average of these two indicators will be the final measurement for determining the cognitive cluster.

$$\mathrm{Info(C)} = \sum_{I_i, I_j \in C} R(I_i, I_j) \quad [2\text{-}2]$$

$$\mathrm{InfoDen(C)} = \mathrm{Info(C)}/N_e \quad [2\text{-}3]$$

$$\mathrm{ClusterVal(C)} = 2\mathrm{Info(C)}/(N_e + 1) \quad [2\text{-}4]$$

Third, valuable links are captured by measuring the relations between each item and each cognitive cluster. The indicator of this step is shown in [2-5]. Forth, key items are derived through identifying terms that have strong relationship with all other cognitive clusters and newly added terms in the previous step. [2-6] will clarify this procedure. Along with IdeaGraph, there has been various variations from KeyGraph, including Polaris (Okazaki & Ohsawa, 2003), data crystallization (Ohsawa, 2005), and Tangled String (Ohsawa & Hayashi, 2015).

$$\mathrm{PR}(I_i, CC) = \sum_{c_k \in CC} R(I_i, c_k) \quad [2\text{-}5]$$

$$\mathrm{Key(I)} = \sum_{i=0}^{N_c} PR(I, CC_i) + \sum_{\notin CC, I_k \in G} R(\mathrm{I}, I_k) \quad [2\text{-}6]$$

IdeaGraph is shown to be an excellent tool for obtaining meaningful but hidden information. Most of low-frequency terms are generally interpreted as some sort of noises, which must be removed from the beginning of the analysis. However, this paper asserts that there exist certain noise-looking terms with a great importance. This study uses this technique in scenario building process, specifically for the purpose of capturing capture voices of few and generating fruitful, detailed, and creative context of future depictions.

2.2.3 Morphological Analysis (MA)

MA was first introduced by Zwicky, a Swiss astronomer, and it has been since widely accepted in the field of idea generation. The fundamental premise of MA is that all the problems of metaphysics and philosophy can be solved by systematically combining a very small number of basic principles (Wissema, 1976). The idea was proposed by Ramon Lull in the 13th century and first expanded into the fields of solving interdisciplinary problems, prophecy or forecasting, and creative speculations. Recently, studies have been proposing a more objective way of performing MA. They have incorporated patent data and WordNet data in identifying new technology, service, and business opportunities.

MA is about dividing a subject into multiple parts and re-arranging each feature of those parts into numerous different wholes. The result is then organized into morphological matrix, or morphological box, which includes various factors for generating different forms of a subject. The analysis is composed of four major steps. First, the subject is selected and its major functions, or so-called dimensions, which could in-whole represent the subject are identified. For instance, in the research conducted by Jones (1976), a textile wet-processing system can be decomposed into textile additive, fibre substrate, process medium, additive fiber system, etc. Second, all the sub-functions, so-called values or shapes, in which each dimension can manifest itself are extensively identified. The shapes for the second dimension, fibre substrate, are protein, cellulose, nylon, and polyester. Third, the morphological matrix is constructed by labeling the dimensions from left to right and the shapes from top to bottom. The final morphological matrix for textile wet-processing system is illustrated in Figure 2-1. Finally, all the combinations of shapes per dimension are examined. This way, new and different forms of a subject that have yet been realized can be thoroughly identified after removing the existing solutions. The total number of possible configurations in above example is 256, calculated by multiplying the number of shapes for each dimension.

Table 2-2 Morphological matrix example for textile wet-processing

Textile additive	Fibre substrate	Process medium	Additive fibre mechanism
dystuff	protein	aqueous	substantive
O.B.A.	cellulose	aqueous solution	diffusion
finish	nylon	organic solvent	precipitation
antistat	polyester	air gas vacuum	reactive

As described, MA is capable of systematically decomposing a subject and then mixing and matching the parts to expand the possibilities for new opportunities of the subject. Such a process is expected to increase the possibility of producing creative ideas. Based on the assertion that new technologies are not yet fully formed and developed, this research attempts to incorporate this logic into the data-driven technology foresight approach for generating new creative ideas of emerging technologies.

Chapter 3

Foresight for Impact Identification

The first step of proactive management process is to identify and conceptualize potential problems associated with emerging technologies. Unintended and unforeseen societal impacts are the main focus of this research. A deep and ex-ante understanding of those ramifications are vital for fostering a better public acceptance of such unfamiliar technologies. This research question has been dealt in the field of technology assessment; however, conventional approaches suffer from two significant problems: (1) what are the ways of incorporating the voices of the general public in the foresight process? and (2) how could we explore societal future of an emerging technology, while minimizing the involvement of human biases and subjectivity?

This chapter proposes a data-driven foresight framework of identifying societal impacts arising from the emergence of new technologies, while coping with the aforementioned limitations. First, *future-oriented web data* is incorporated for the purposes of embracing as much diverse players as possible and containing not only technical but socioeconomic contents. Second, a text mining technique of LSA is applied to make the process more systematic and data-driven. Such an attempt has greatly reduced human intervention and resulted a more comprehensive and future-oriented impact identification.

3.1 Introduction

Prompting the development of emerging technologies may be worth the ticket to long-term competitive advantage, especially for innovative enterprises or entrepreneurs (Hung & Chu, 2006). This not only leads to lucrative upside potentials for the company itself, but also delivers societal, economic, political, and environmental benefits to societies worldwide (Köhler & Som, 2014; Seear et al., 2009). However, alongside the hopes of such wide-ranging advantages, new technologies unleash unintended and unforeseen societal impacts and raise concerns regarding their integration into society (Andersen et al., 2004). Even though these technologies are technically prepared and on the brink of commercialization, a great number of them are still awaiting mainstream acceptance, mainly caused from a deep skepticism regarding unexpected detrimental effects critically altering the existing standards of our society. 3D printing technology is one of the most prominent examples in such a case. 3D printing technology is capable of creating objects out of thin air though laying down successive layers of materials, and its uses range from practical objects for everyday use to commercial products used in industrial applications. Moreover, technology observers speculate that it holds promise for numerous industries, such as fashion, lighting, computer, telecommunication, health care, etc. Yet, we cannot ignore the potential negative consequences (Pîrjan & Petroşanu, 2013). On the dark side, for instance, the digitization of labor could amplify unemployment rate, and the application in bio-printing may disrupt our moral, ethical, and legal standards. Due to the narrow focus on the added values and benefits of new technology, firms often neglect its resulting unseen impacts (Halaweh, 2013). Hence, for those hoping to foster sustainable technological innovation strategies and out-behave the competition in their industries, an adoption of a well-developed preliminary assessment to deal with such unforeseen challenges is of utmost importance (Köhler & Som, 2014).

We must, therefore, figure out how to redress such a disquiet. When addressing the ramifications of new technologies, various methodologies have been utilized particularly in the field of technology assessment (TA), such as impact analysis, Delphi analysis, risk assessment, and scenario analysis (Chen et al., 1981; Tran & Daim, 2008). Specifically, impact analysis literature has dealt with technologies' consequences in terms of social, political, and economic context. (Ballard & Hall, 1984; Palm & Hansson, 2006; Wright, 2011). Commonly used in project assessments, social impact assessment (SIA) is a type of TA that focus particularly on social impacts and their systematic assessment regarding technologies and projects (Becker & Sanders, 2006; Franks & Cohen, 2012; Tuominen & Ahlqvist, 2010; Vanclay, 2002). However, conventional methods have three-fold limitations: a sheer lack of consideration of emerging technology's nature, an underestimation of public's insights, and a sole reliance on participatory approaches. Emerging technologies consist of unprecedented technological features with socially unaccepted functions, thereby contains complex and intangible characteristics within their nature. In other words, the social implications are difficult to understand due to complex inter-relationship among uncertainty associated actors and intangible consequences that cannot be seen ex-ante. To deal with social consequences of such ill-defined technologies, one has no other choices but to employ people's imagination, vision, foresight, creativity, and intuition to visualize potential impacts and the resulting world we have not yet seen (Day et al., 2004). The key of such foresight is on the diversity of knowledge source, which can be gained from large-participation (Hiltunen, 2008; Eaton et al., 2014). Conventional foresight processes have been performed primarily with a handful of experts; however, experts face various cognitive limitations when foreseeing the future (Fischhoff et al., 1982; Pang, 2010; Schatzmann et al., 2013; Schoemaker, 2004) and the reliability of their judgments are quite questionable (Han & Shin, 2014); whereas publics have recently played relatively critical role in shaping the future technology and society (Day et al., 2004; Pang, 2010). Incorporating the needs and opinions of diverse types of users, such as end users and product buyers, is in fact essential for increasing technology's widespread adoption (Shluzas & Leifer, 2014). Moreover, most studies have relied on qualitative or participatory approaches, which normally involve time-consuming and labor intensive workshops (Carlsen et al., 2014; Frewer, 1999). It may take weeks or even months to collect opinions from scientists, engineers, and other industry officials with complementary domains of specialization.

By dismissing conventional methods that obtain direct thoughts from prospective stakeholders, this paper seeks to overlook rich possibilities in gaining insights into *future-oriented voices of both experts and general publics from the Web*. To this mean, future-oriented documents are gathered from a wide-ranging of players, both potential users and experts from various fields. This future-oriented collective intelligence provides futuristic and comprehensive insights regarding intangible and complex social responses for emerging technologies. Furthermore, we propose a quantitative method for identifying and understanding the holistic overview of societal impacts of emerging technologies. Since future-oriented documents are not only a large amount but composed of unstructured natural language, a text mining (TM) is applied to handle the documents as an analyzable and structured set of keywords. In particular, the paper applies latent semantic analysis (LSA) technique, capable of extracting meaningful semantic patterns, thus generating topics on contextual similarity (Blake & Ayyagari, 2012). Compared to conventional participatory approaches, such a quantitative approach can be more efficient and systematic. As a result, the proposed methodology generates future scenarios, which indicate emerging technologies' early warning signs of potential social impacts and their specific consequences to the society.

To be clear, the objective of our proposed approach is not to suggest the best way to overlook future social impacts. Nor are we suggesting that this is the right answer for imagining what could happen in the near future. However, borrowing future-oriented perspectives from various players may take us one step closer to gaining better insights into unfolding the uncertain future. To further delineate fundamental basis of our approach, Section 3.2. illustrates social impact of emerging technology by suggesting its distinctive nature and its related literature coping with such impacts. Section 3.3 describes the overview and detailed procedure of LSA for extracting meaning contextual patterns. After proposing our methodology in Section 3.4, we provide a practical case study using an emerging technology of unmanned aerial vehicle, or drone, in Section 3.5. Lastly, Section 3.6 offers conceptualization of drone technology's social impacts, comparative analysis to conventional approach, and its implication to theory, practice and policy.

3.2 Emerging Technology and its Social Impacts

3.2.1 Distinctive Nature of Emerging Technology

Looking back to the history of technology innovation, societal and epistemological consequences, whether intended or unintended, seems perhaps inevitable. Technological developments in matured stage, or established technology, constitute systems of relatively well-defined actors, relations, and institutions, thereby resulting social issues that are tangible and identifiable (Köhler & Som, 2014; Van Merkerk & Smits, 2008). Detailed information about potential hazards, such as probability and magnitude of exposure, can be reasonably predicted and quantified through an analysis of empirical knowledge or historical information (Köhler & Som, 2014; Porter & Roper, 1991).

The very nature of emerging technology itself seems quite different from that of established technology (Halaweh, 2013; Köhler & Som, 2014). Technologies in their embryonic stages are not yet fully developed nor indeed extensively diffused. They hold a critically high level of uncertainty due to its unsettled nature, whose future is unknown, unpredictable, or unstable (Eaton et al., 2014; Stahl, 2011). Due to lagging state of technological development and an insufficient amount of knowledge regarding technological standards and specifications, emerging technologies hold intangible social implications (Köhler & Som, 2014). Furthermore, emerging technology in societal and economic realities cannot be thoroughly understood since it holds complex inter-relationship with diverse actors, such as existing technologies or social norms (Köhler & Som, 2014). Due to their inherent complexities, risks migrate unexpectedly and amplify with one another to surprise us with unforeseen systemic effects (Alcock & Busby, 2006; Assmuth et al., 2010). To deal with social consequences of such ill-defined technologies, one has no other choices but to employ people's imagination, vision, foresight, creativity, and intuition and visualize the resulting world we have not yet seen (Day et al., 2004).

Technology foresight involves relatively more qualitative means for monitoring the future, compared to technological forecasting (Cuhls, 2003). Instead of arriving at accurate predictions of the future, foresight results multiple interpretations about the future by providing glimpses of events regardless of the degree of probability (Canongia et al., 2004). The foresight activity becomes severely challenging when emerging technologies come into play since they are involved with countless components, such as policy, market, supply network, and infratechnologies (Featherston et al., 2016). However, the activity with social components can be even more frustrating due to its inherent challenges: "wicked problem" and "black swan". Normally, the problems closely associated with societal and political dimensions of unsettled technologies are characterized by "wicked" nature (Ehrenfeld, 2008). Not only cannot the problems be clearly defined but they also don't hold definite solution principles in the sense of definitive and objective answers (Köhler & Som, 2014; Navarro et al., 2008; Rittel & Webber, 1973; Pang, 2010). In fact, since "wicked problems" defy traditional linear problem-solving approaches, multiple alternative perspectives must be accommodated and focus on comprehensive possibility rather than the accurate (Navarro et al., 2008). Additionally, foresight becomes even more problematic with a phenomenon, so-called "black swan" (Taleb, 2010). Social issues are exceptionally disruptive and difficult for humans to think about seriously (Pang, 2010). Though it is impossible to find objective solutions for these problems, they cannot be ignored due to large, unforeseen impact (Reeves, 2012; Pang, 2010). Previous studies suggest to surrender the assumption that expertise and professional competence will yield better decisions related to such issues.

In response, this research highlights the importance of information diversity and thereby broadens the range of input by capturing not just technical knowledge from experts but creative opinions from publics. It is expected to gain synergy effects between accuracy and originality in foresight ability. Numerous studies have suggested suitability of adopting both perspectives when foreseeing social impacts. As proposed by the study of Hiltunen (2008), the words from futurists and ordinary people are considered two most appreciated sources for foretelling the societal changes in the future. Moreover, Eaton et al. (2014) mention that publics may provide knowledge of non-technical nature, social considerations such as social and ecological impacts, social acceptability, and other social and ethical concerns that are pervaded in the realm of our society instead of science. This is further supported by Pang's Futures 2.0 (2010), which suggested psychological limitation within experts and ordinary people. Experts, for instance, are considered to be the ideal figures who are capable of accurately envisioning the future based on their field of expertise; however, they are incapable of thinking outside the box due to various cognitive limitations, such as overconfidence, illusion of control, and information distortion, especially when it comes to forecasting large, complex, and indeterminate events (Fischhoff et al., 1982; Schatzmann et al., 2013; Schoemaker, 2004; Pang, 2010). At the same time, the general public has become more innovative and thus have started to play a critical role in shaping the future of our society; therefore, they may be more eloquent about their needs or problems of future application situations (Day et al., 2004; Pang, 2010). Yet, they tend to think poorly due to lack of expertise. Based on this information, we envision social impacts through in-depth analysis of multi-faceted future-oriented voices from experts and publics, but in a more efficient and effective way.

3.2.2 Technology Assessment

Numerous approaches in the field of technology assessment (TA) are incorporated for the examination of technologies' adverse impacts (Tran & Daim, 2008). One of the preliminary works on the systematic assessment of technology impact is undertaken by Coates (1974). The author extensively reviews a wide variety of techniques, such as Delphi, cross-impact analysis, morphological analysis, and extrapolation, and lays the groundwork by introducing a term called, comprehensive impact assessment. Ballard and Hall (1984) integrate the concepts of impact assessments and apply into the case of the Western Energy Study. According to the study, the value in addressing the uncertainties depends in large part on management devices, interdisciplinary, external review, and participatory research. Based on these preliminary attempts, the field was TA was established as an early warning system to detect and control negative effects exerted by technological changes and developments (van den Ende et al., 1998; Centron & Connor, 1972).

Numerous concepts have been derived from TA; however, it was primarily evolved in two directions: (1) how to conduct for a specific purpose and (2) how to systematically expand the pool of intelligence. First, the initial form of TA aimed at analyzing specific implications of emerging technologies, including ethics, privacy, society, or environment. Tran and Daim (2008) propose a large collection of analytical models for TA, ranging from macro-system dynamic, environmental, medical, and energy to social impact models. Moreover, variant studies are carried out focusing on specific dimensions of technology impacts, such as health, environment, ethics, and privacy. Some of the research shed light particularly on the ethical implications of technologies. Palm and Hansson (2006) propose a new form of TA termed, ethical technology assessment (eTA), and identify nine crucial ethical aspects of new technology via check-list approach. Wright (2011) suggest a new framework for ethical impact assessment (EIA) for identifying key social values and ethical issues involved within information technology. Privacy Impact Assessment (PIA) was came into general usage to evaluate the potential effects on privacy of a new technology or project (Clarke, 2009). A considerable amount of literature in the field of nanotechnology examine early warning signs for environmental, health and safety (EHS) uncertainties (Healy et al., 2008; Ostertag & Hüsing, 2008; Stefaniak et al., 2013). Through exploring the perception of a number of experts using Delphi survey, Ribeiro and Quintanilla (2015) address the complex and context-specific nature of biofuel technology and identify specific social impacts that could possibly emerge in the future (Ribeiro & Quintanilla, 2015). In addition, several studies look into the way people respond to uncertainties of new technologies and emphasize the role of ethical concerns within the public acceptance (Frewer, 1999; Pidgeon et al., 2011).

Second, the other form of TA sought to expand the pool of actors in the system. Both Constructive TA (CTA) and Innovative Technology Assessment (ITA) broadened the decision making process by involving more aspects and more societal actors at an early stage of technology development, and thereby developing common perceptions of a specific problem (van den Ende et al., 1998; Palm & Hansson, 2006; Schot & Rip, 1997). There are other similar concepts, including Interactive Technology Assessment (iTA) and Participatory Technology Assessment (pTA).

Among the other methodologies, the field of social impact assessment (SIA) must be highlighted in this research. SIA is one of the detailed notions of TA, and it focuses on the systematic identification and estimation of likely consequences of potential social issues regarding socio-economic components of society and environment. It serves the purpose of minimizing public resistance to certain technology or project, thereby reducing potential disruptions to increase market success. However, prior to the study conducted by Vanclay (2002), SIA related research have proposed a variety of social impact variables, which were not in themselves impacts, but rather represent the measurable outcomes of social change processes. Here, social impacts are a social and cultural consequence to human populations that alter the ways in which people live, work, play relate to one another, including the changes to individuals' norms, values, and beliefs that guide and rationalize their cognition of themselves and their society (Burdge & Vanclay, 1995). Considering the impacts suggested by previous SIA approaches, the author re-conceptualizes social impacts and divides into seven categories as summarized in Table 3-1. Quite often, however, previous studies fail to take intrinsic nature of emerging technologies into account and deal only as if this is the case of existing and established technologies. In spite of complex and intangible nature of new technology's consequences, they clearly have not considered using future-oriented knowledge or incorporating the widest possible range of players, including general publics. The remedy of such an issue is further discussed in next section.

Table 3-1 Summarization of social impact variables and concepts (adopted from Vanclay (2002))

Social impact concepts	Social impact variables
Health and social well-being impact	personal loss mental health and subjective well-being autonomy stigmatization or deviance labeling annoyance dissatisfaction experience of moral outrage
Quality of the living environment impacts	perceived quality of the living environment (work, home or neighborhood) actual quality of the living environment disruption to daily living practices leisure and recreation opportunities aesthetic quality environmental amenity value adequacy of physical infrastructure adequacy of social infrastructure perception of personal safety and fear of crime actual personal safety and hazard exposure actual crime and violence
Economic and material well-being impacts	workload standard of living occupational status/ prestige and type of employment level of unemployment in the community loss of employment options
Cultural impacts	change in cultural values cultural affrontage cultural integrity loss of natural and cultural heritage
Family and community impacts	alterations in family structure changes to sexual relations obligations to ancestors changed demographic structure of the community
Institutional, legal, political, equity impacts	integrity of government and government agencies violation of human rights participation in decision making access to legal procedures and to legal advice
Gender relations impacts	women’s physical integrity gendered division of household labour equity of educational achievement between girls and boys political emancipation of women

3.3 LSA for Constructing Scenarios

LSA text mining technique is applied as an alternative way of constructing scenarios in two primary motives. First, it is capable of retrieving multiple conceptual topics based on the documents with similar contexts and thus resulting topical groupings in terms of not only documents but also constituting terms (Blake & Ayyagari, 2012). Since the terms are intertwined with and mutually interdependent with multiple episodes (Landauer et al., 1998), each term cluster can be seen as the new abstraction of a certain topic. Since we incorporate the documents of future concerns, the *term clusters* may represent summarizations of people's future interests or fears. Here, such a summary is denoted as a “scenario”. Second, compared to other grouping approaches, LSI offers effective term-grouping since it captures indirect similarities among the terms based on their indexing. This consequently leads to the grouping of new term classifications (Blake & Ayyagari, 2012). For instance, Dumais et al. (1988) demonstrate the high performance of LSA by tackling the limitations arising in the analysis of natural language: synonymy and polysemy. In case of synonymy, there are cases where people use different words to refer same object or concept. However, the vector space representation hardly captures the relationship between synonymous terms due to separately denoted dimensions. Conversely, the same word may hold multiple meanings, and this corresponds to the concept of polysemy. These two issues thus cause underestimation and overestimation of similarity measures, in respective. LSA is based on a distinctive assumption that there exist some underlying or latent semantic structures in word usage data, which may be obscured by the variability in word choice (Furnas et al., 1988). Therefore, the similarity estimates applied in LSA are not just based on contiguity frequencies or co-occurrence counts but depend on more powerful statistical analysis that accurately infer much deeper relations, or hidden patterns, within textual data (Landauer et al., 1998).

This whole procedure of mining texts and forming into multiple storylines is more like constructing scenarios. The methodological inspiration for this particular research is, in fact, scenario planning. When attempting to envision future impacts, we use futures analysis of scenario thinking about the development of new technologies, and thereby proactively reduce uncertainties and risks (Webster, 1999). Scenario generation is a widely accepted method for envisioning the future, capable of anticipating many possibilities rather than proposing a single detailed story. Its significance must be highlighted particularly in the state where the emergence of new technologies causes complex twists and turns to the society, and there is in need of deepening and broadening the possible alternatives of future impacts. In details, scenario generation captures complex relationships between technology and various environments, such as industry, market, society, regulation and provides the formulation of concrete strategies to take a further action (Robinson, 2009). It is commonly adopted as a supportive tool for interactive workshops in national foresight programs but quite less recognized in TA literature (Carlsen et al., 2010).

3.4 Research Framework

A schematic overview of our methodology is depicted in Figure 3-1. The first step is data collection where information regarding future concerns of certain emerging technology is gathered. The second step is the development of scenarios by analyzing the data sets. Textual data are transformed into structured data format via linguistic preprocessing, and the terms are clustered into a number of corpora based on LSA. Based on the coherency check of grouped terms, each corpus is assessed and defined as a possible scenario.

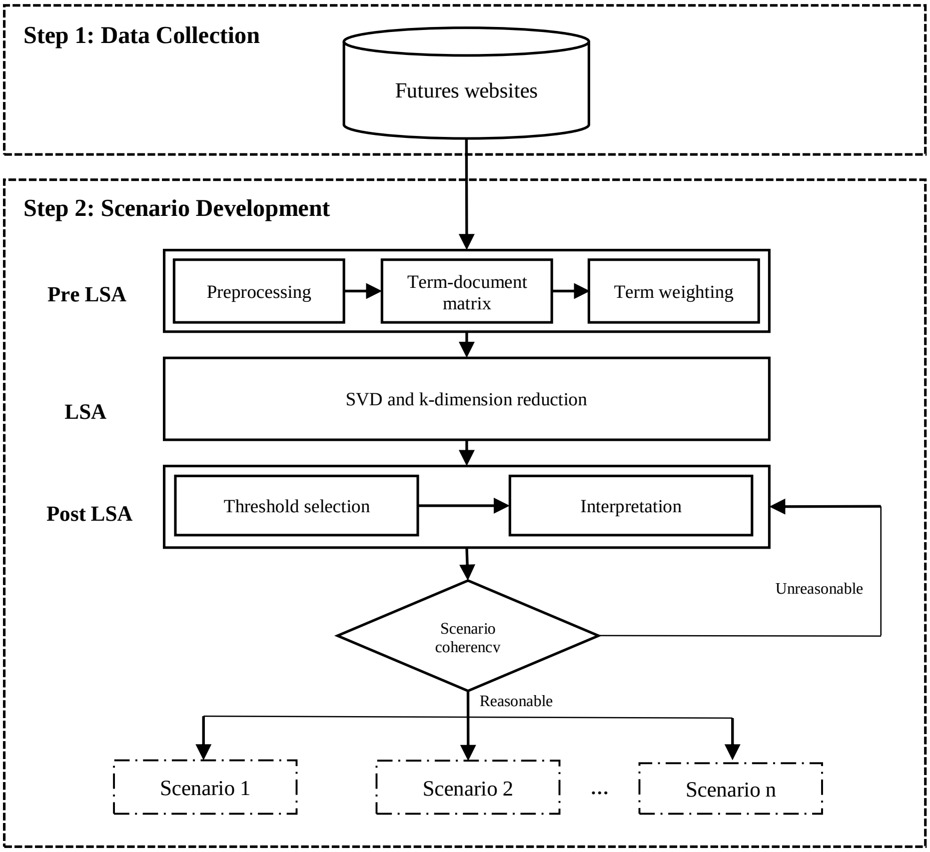


Figure 3-1 Overview of proposed methodology

The whole process seems similar to general LSA-based topic generation approach as the study of Sidorova et al. (2008); however, it is partly modified to suit the research question. Specifically, the changes were mainly in preparation and interpretation phases. In the preparation phase, proper websites are first chosen for the data source and it is carefully conducted based on the website's major technological topics and future orientation of websites. Moreover, proper terms are then selected for actual input in LSA. The terms are filtered using two criteria: frequency filtering and context-related filtering. In the interpretation phase, the scenario development must be conducted in a more systematic way. The terms are divided into technical and social terms and serve different purposes, in that technical terms are the cause and social terms are the effect. In addition, technical experts must be involved when interpreting since technology's specific function, and features must be taken into account when developing scenarios. The details are illustrated in further sections.

3.4.1 Step 1: Data Collection

For the first step, we must choose right websites for the data source. Since the quality of LSA result is highly depended on what input is being fed, we must carefully decide which websites to use for the data crawling process. The data apparently should include potential impacts of emerging technology that have not yet occurred; hence we must take two factors of major technological topics and future orientation into consideration for proper data selection. First, the website must handle the target emerging technology, and the involved players, including experts and public, should actively discuss the target. Second, future-oriented context is necessary, in that the opinions should be mainly about the future development of that technology and its potential implication to our society.

Future-oriented information can be collected from futures websites as a dataset containing future depictions of certain emerging technology. Data sources of futures websites should be chosen after deliberate consideration. The prominent examples of future-oriented websites are Future Timeline, MIT Technology Review, World Future Society, Wired, and io9, as shown in Table 3-2. These websites offer an immeasurable amount of information regarding new technologies and their impact on our lives and society. Future Timeline provides the encyclopedic overview of future topics based on perspectives of futures experts, while relevant public opinions are also expressed within the forum. MIT Technology review offers magazine type futuristic news regarding technological trends and predictions. The general public may share their ideas through leaving comments below the articles. World Future Society encourages innovators and creative thinkers to share ideas, mostly futures experts, on what the future will be like in terms of society and technology. Wired extensively covers current and future trends in technology and their future role in business, science, politics, entertainment, and communications. Finally, io9 covers futurist news and blogs, and they may be commented by publics who have interests in the future. It is seen that above five websites provide future-oriented information based on the collaboration of futurists from a wide variety of fields and the general public.

Table 3-2 Summary of future-oriented websites

Website	Areas of focus	Involved players	URL
Future Timeline	AI & robotics, home & leisure, society & demographics, space, etc.	Experts & Public	http://www.futuretimeline.net
MIT Technology Review	Biomedicine, computing, energy, materials, robotics, etc.	Experts & Public	http://www.technologyreview.com
World Future Society	Social, economic, technology, science, etc.	Experts only	http://www.wfs.org
Wired	Design, science, security, entertainment, design, etc.	Experts only	http://wired.com
io9	Science fiction, futurism, science, technology, etc.	Experts & Public	http://io9.com

Furthermore, accurately locating a specific piece of information among thousands of web pages is also an essential and challenging issue. This research is aimed at examining the concerns and worries regarding the impact of emerging technologies. As described in Figure 3-2, It requires four steps in selecting a proper form of search queries. First, all different synonyms of target technologies must be identified using other databases, like Wikipedia, WordNet or any machine readable thesaurus. We suggest redirect links in Wikipedia data since we are, here, dealing with newly emerged technologies. Wikipedia offers extensive topic coverage, up-to-dated context, domain-specific description and rich semantics, and redirect link of Wikipedia provides synonyms of target concept. For instance, drone technology includes a redirect link to UAV and multirotor. Second, proper negative-tone keywords for describing problematic situations are selected. The words like “concern”, “worry”, “disappointment”, “issue”, or “problem” can be included and lexical databases, like dictionary, thesaurus, or WordNet, can be utilized for a more extensive identification process. Third, a candidate list for search queries is constructed by considering all possible combinations of target technology keywords and negative-tone keywords. Finally, the most suitable search queries are selected by empirically testing in target future-oriented websites. Suitability of the queries are decided based on the contents of websites or the characteristics of target technology. The search queries retrieving a sufficient number of pertinent information are selected as final search queries for the research.

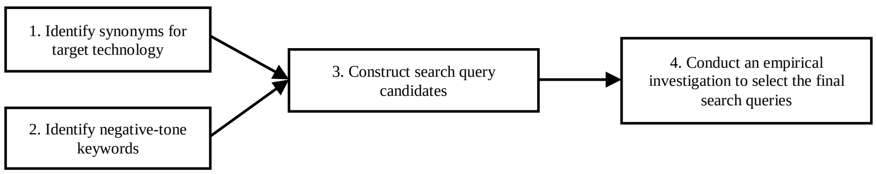


Figure 3-2 Process of selecting search queries

3.4.2 Step 2: Scenario Development

After collecting relevant data sources, we construct potential impact scenarios of emerging technologies. Partly adopted from the study by Evangelopoulos et al. (2012), the general procedure of LSA is shown in Figure 3-3 (Evangelopoulos et al., 2012). Scenario building process consists of following three subsections:

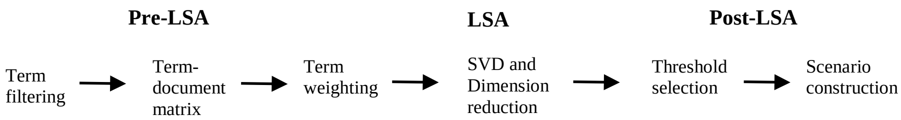


Figure 3-3 Process of scenario development

(1) *Pre-LSA: preprocessing future-oriented web data*. Raw data are well preprocessed in order to gain effectiveness and efficiency of further text analysis.

(2) *LSA: applying Latent Semantic Analysis*. LSA is implemented to compute both direct and indirect semantic similarities of the selected terms and to group with one another based on their similarity values.

(3) *Post LSA: interpreting and constructing scenarios.* Using the TS matrix, high loading terms are observed to extract meaningful factors and interpret them as future scenarios.

3.4.2.1 Pre-LSA: Preprocessing Future-oriented Web Data

The second step of our research framework is data preprocessing. Crawled data from futures websites are in a raw unstructured format consisting of natural language. In addition, one key thing to note is that future-oriented documents in their raw format are not quite fully optimized for LSA. While LSA gives the best analytic performance once each document represents a single topic, certain documents may consist of multiple topics. Therefore, those documents must be segmented into multiple text files of single topics.

Prior to text analysis, all documents must be transformed into numeric vectors, structured and computer-understandable form. Text pre-processing process is composed of three steps. First, raw text cleaning phase is performed through linguistic preprocessing methods including tokenization, stemming, part-of-speech tagging, and stop-word removal. The most challenging and critical process is term filtering, the selection of proper terms for actual input in LSA. When analyzing thousands of text documents, over 10,000 terms are extracted without complete pre-processing process. Term selection is, in fact, remains an open issue (Evangelopoulos et al., 2012). But this paper reduces the term dimensionality in two criteria: frequency filtering and context-related filtering. Only the terms that yield certain threshold frequency values and the ones closely related to target technology are selected automatically and manually, in respective. Specifically, certain term must occur with reasonable number in the collected database, and it should either correspond to detailed technological or social information.

Second, once all the unnecessary terms are eliminated from the data set, our document set must be converted into vectors of term frequency (TF) values and expressed as a term-document matrix. By way of illustration, if *m* bag-of-words are extracted from n future-oriented documents, the database is represented as a m-by-n term-document matrix, where each cell entry contains the frequency in which a keyword occurs in a document.

In the final step, term vector weighting phase is performed where the raw frequency values are transformed into weighted term vectors. Two of the most widely applied transformations are the product of the term frequency and the inverse document frequency, known as *tf-idf*, and the log-entropy. However, log-entropy provides factors that may be biased toward high-frequency terms (Evangelopoulos et al., 2012). This study thus considered *tf-idf* transformation as a more suitable term weighting method based on the characteristics of *future-oriented web data*. In detail, the *tf* correspond to the frequency of occurrence of a term in the future-oriented document. Since it is assumed that the term with more frequency represents a more characterizing feature of the document content, *tf* is used to indicate term importance of content representation (Coussement & Van den Poel, 2008; Luhn, 1957). But in order to adjust importance depending on the varying length of documents, *tf* is acquired by taking a logarithmic transformation of the original term frequency value:

$$tf_{ij} = 1 + \log_2 n_{ij} \qquad [3\text{-}1]$$

,where n_{ij} is the term frequency of term i in document j. Moreover, the idf corresponds to the inverse value of document frequency. The weight of a term is determined assuming that more rarely occurring term throughout the collection of documents, in which the term occurred, shows more discriminating feature. The logarithm of the *idf* value was taken in similar manner:

$$idf_i = \log_2 \frac{c}{df_i} \qquad [3\text{-}2]$$

,where n is the total number of documents in the entire database and df_i is the document frequency where term i is included. Consequently, the final weight of term i is in future-oriented document j is represented by

$$w_{ij} = tf_{ij} idf_i \qquad [3\text{-}3]$$

$$w_{ij} = (1 + \log_2 n_{ij})(\log_2 \frac{c}{df_i}) \qquad [3\text{-}4]$$

The resulted high-dimensional, weighted matrix is utilized as an input for the further analysis of LSA.

3.4.2.2 LSA: Applying Latent Semantic Analysis

The third and perhaps the most essential step of this study is topic detection procedure, where the dimension of weighted term-document matrix is reduced through LSA. As thoroughly described in section 2.4, the input matrix $X_{t\times d}$ is then decomposed into three matrices: term eigenvectors (T), diagonal matrix of singular values (S), and document eigenvectors (D). The decomposition is applied as below,

$$X_{t\times d} = T_{t\times m} S_{m\times m} D_{d\times m} \qquad [3\text{-}5]$$

SVD projects matrix X onto a k-dimensional space where n is far less than the value k. The intensity of dimension reduction of SVD is quite an unanswered question. There is no right answer for selecting the number of topics, or k; however, k must be high enough to prevent from losing important information and to capture underlying topics but, at the same time, low enough to remove unnecessary information, known as noise. Consequently, SVD reconstructs the high dimensional matrix into an orthonormal, semantic, and latent subspace through clustering similar terms into certain topics. Each derived topic is described in detail based on constituting key terms since it holds high discriminatory power compared to other topics in the reduced feature space (Coussement & Van den Poel, 2008).

Based on orthogonal characteristic, LSA is capable of allocating a term in multiple topics capturing latent topics. First, a possibility of cross-loading is inevitable when factors are closely related and they are extracted from a field whose subareas pull from a common language (Sidorova, 2008). This may cause certain terms in a topic to simultaneously appear in other topics. Such a phenomenon seems natural in describing future eventualities with details since same terms could be used in a wide variety of contexts. Second, terms in a certain topic have little relations with the terms in other topics but have strong relations with one another within that topic (Lee et al., 2010). This is based on the premise that certain terms appear in contextually similar documents to establish relationships with one another. In other words, terms, which never co-occur in the same document may be estimated as closely related and thereby appear near each other in the k-dimensional space. Consequently, LSA is applied as a soft clustering technique to set values of indirectly related terms or documents as fractional membership of certain cluster (Manning et al., 2008), thereby capturing latent and hidden clusters. In this particular research, each derived topic is considered as a future scenario that represents summarization of various voices of emerging technologies' unforeseen effects and potential concerns.

3.4.2.3 Post-LSA: Constructing Scenarios

Constructing scenarios involve topic interpretation of $T_{t\times k}S_{k\times k}$ matrix, the component matrices after the process of reducing into k dimensions. As such, this refers to the term loading values on the common principal components of terms and documents (Blake & Ayyagari, 2012; Evangelopoulos et al., 2012). Lists of high-loading terms are compiled and examined to produce topic labels, which may be interpreted as document summarizations. Here, we refer each topic as a future scenario. One key thing for representing each factor as a coherent scenario is selecting the right term loading threshold. Just as choosing the proper k value, there are no definitive methods to determine these thresholds; however, according to the study from Evangelopoulos et al. (2012), domain experts and researchers must manually and empirically derive the right loading threshold values based on the coherency and reasons of each scenario.

Each factor may incorporate different threshold value as long as the constituting terms have high loading values to ensure the scenario is contextually concrete. Consequently, this procedure may be similar to what is commonly done in conventional factor analysis (Sidorova et al., 2008). Each topic is defined based on its high-loading terms. Three different parts of speech, including noun, verb, adjective, are present, and they are either technical or social terms. Such two types are the ingredients for systematic scenario development and serve different purposes. Technical terms are normally the factors that are causing the issue, whereas the social terms are the factors that are affected by the situation. For instance, the term "transmission" is the technical element that corresponds to main cause of the scenario and the term "privacy" is the social element that indicates the effect to our society. Lastly, $S_{k\times k}$ matrix demonstrates the significant information for measuring the relative importance of each scenario. Therefore, scenarios could be weighted, or prioritized, based on the importance values of contextual topics.

The final output of subsection is the generation of possible scenarios, the negative influences of emerging technology to human population's norms, values, and perceptions. Here, technology experts must be involved in the interpretation phase. The final scenarios extracted from the analysis illustrate various eventualities based on the groups of terms. In order to fully understand the context, the interpreter must be well equipped with technical knowledge, especially regarding technology's specific function and features. There is definitely in need of professional support when configuring the terms into sentences and thereby forming a scenario. Once independently examining primarily with constituting terms in each factor, several experts must discuss resulting interpretations and topic labels with one another and validate with corresponding documents in order to reconcile different viewpoints and reconstruct into coherent and reasonable scenarios.

3.5 Illustrative Case Study: Drone Technology

The proposed model is demonstrated with an empirical experiment using concerns and threats regarding unmanned aerial vehicle (UAV), commonly known as drones. First introduced for military combat operation, drone technology is categorized as emerging and promising technology that possesses high prospects for a remarkable proliferation at some future point. In fact, drones are expected to provide a wider range of applications in three sectors: private, public, and recreational. In private sector, it may inspect diverse infrastructures, support wireless communication, monitor natural resource, provide digital mapping, or even manage wildlife environment. In addition, there is a speculation that certain conglomerates like Amazon has a future plan on delivering packages using drone technology. In public sector, it may take aerial photos at traffic collisions or crime scenes, gather security intelligence, and support environmental research. Finally, as drone technology is becoming more advanced and increasingly cheaper, even general public are attracted to this latest technology and interested in capturing cool video footage from high above. Although drone technology is right on the brink of commercialization, such unproven technology raises important questions and concerns about social and other environmental consequences. Therefore, drone can be considered as a suitable sample technology for this particular research, which requires a deep understanding of undesirable ramifications in order to foster a better public acceptance.

Table 3-3 Summary of data collection

Website	Query	Documents collected
Future Timeline	(“drone” or “uav”) AND (“issue” or “concern” or “problem”)	183
MIT Technology Review		994
World Future Society		678
Wired		965
io9		416
		3,236

Future-oriented documents representing experts’ and general publics’ concerns of drone technology are retrieved from five future-oriented websites using our web crawling program. The search queries of “drone or uav” and “issue, concern, or problem” are used to collect data from desired future-oriented websites. Table 3-3 briefly shows the details of data collection result.

A total of 3236 documents are retrieved from five future-oriented websites, and total 3763 terms were extracted. However, as presented in Table 3-3, the numbers of the collected documents are quite unbalanced. This is due to the distinct characteristics of the websites and involved players within them. For instance, relatively greater number of documents were collected from MIT Technology Review, io9, Future Timeline, and Wired since both experts and public are allowed to post opinions in Technology Review, io9, Future Timeline; and science and technology related information is abundant in Wired compared to other experts-only websites.

Furthermore, certain terms are first extracted from the database. To acquire clearer results after the topic detection process, a number of terms are excluded based on frequency and context filtering. Specifically, those with 20 or less frequency values and with meaningless contextual information, such as terms unrelated to society or drone technology, are removed for further process. In result, total of 323 keywords are deliberately chosen for further analysis. After thoroughly executing preprocessing, term weighting is conducted on the original term-document matrix, and the *tf-idf* matrix, $X_{t \times d}$, is constructed. SVD is performed to decompose such matrix into three different matrices, $T_{t \times m}$, $S_{m \times m}$, and $D_{d \times m}$.

In order to extract as topics of future scenarios, the matrix of TS is observed, and the maximum number of contextual patterns is empirically determined through varying the values of k. Initially conducting with 5 contextual patterns, we consecutively performed LSA until the constituting terms and documents with high loading values no longer best represent further topics. In result, first 15 contextual patterns demonstrated logically consistent and plausible scenarios; however, the 16th or greater patterns seemed quite counterintuitive. Moreover, it was difficult to label the topics greater than 15th since documents with highest loading values were relatively unrelated with determined topics. Based on the related terms and documents, scenario details are interpreted and labeled individually by three experts of data and futures research. When the results were compared, the interpretation came out to be 93% similar and the labeling was compromised by thorough discussion. An example of applying cut-off values used for the scenario development.

Since the ingredients for generating the scenarios are composed of three different parts of speech, including noun, verb, and adjective, morpho-syntactic features of term candidates were concatenated to be developed as coherent storylines (such as “noun – noun”, “noun - verb”, “noun – adjective – verb”, etc.). When iteratively interpreting each term corpus, around twenty terms were found to be suitable enough to comprehend overall plot of the scenario describing one certain topic. Whilst all twenty terms may provide the contextual details, the five highest loading terms are considered the key terms that define the main theme, or primary structure, of the scenario. For instance, terms like “insect”, “zoom”, “bother”, “privacy”, and “danger” constitute the first topic with top five highest loading values. When they are used as the stem of words for developing a short plot of a future concern, the main plot of “drones with the size of an insect may considered dangerous since they are capable of invading privacy and bother general public” can be constructed. In addition to the main plot, the terms like “park”, “camera”, and “privacy” provide the details of the places, means, and conditions. We then evaluated documents with high loading values to confirm the validity of the scenario. In a similar manner, total fourteen topics were interpreted. The names are assigned to fourteen scenarios, and their twenty highest loading terms are enumerated in Table 3-4.

Table 3-4 Results of scenarios and their top 20 highest loading terms

Number	Scenario	Top 20 Highest loading terms
#1	Biomimetic Drones	insect, zoom, bother, privacy, dangerous, breaking, park, animal, nano, movement, data, armed, avoid, wifi, technology, biomimetic, sound, location, safety, scare, camera
#2	Corrupted Journalism	news, apartment, arrest, annoy, detect, film, public, hotel, privacy, place, photo, violation, radar, moral, home, upset, habits, journalism, celebrities, accident
#3	Hacked Enforcement Drones	enforcement, live, intercept, surveillance, drugs, police, criminal, video, intercept, infrastructure, risk, national security, illegal, transmission, device, hacker, policy, activity, abuse, location
#4	Vehicle Accidents	safety, accident, aircraft, crash, training, injure, property, regulation, avoid, kill, emergency, sky, visual, fall, condition, instruction, hit, hazard, life, environment
#5	Task Interference	interfere, enforcement, detect, operation, noise, transmission, network, tracking, sound, hazard, alert, detection, mission, emergency, disrupt, task, poacher, hide, service, infrastructure
#6	Terrorism	armed, attack, terror, kill, weapon, threat, moral, bomb, security, surveillance, criminal, stress, spy, prison, arrest, fight, illegal, law, culture, chemical
#7	Prank Accident	kids, fun, scare, animal, shock, bad, tiny, hit, young, life, drop, fall, live, frightening, value, prank, terror, personal, wilderness, problem
#8	Transmission Interception	transmission, personal, interception, network, wifi, device, connect, send, record, information, credit card, moral, password, privacy, crime, location, tracking, ethical, criminal, permission
#9	Park Wilderness	noise, park, wilderness, sound, stress, visitor, monitor, pollution, film, intercept, activity, follow, calm, video, material, environment, spy, disturb, experience, privacy
#10	Wildlife Disruption	disrupt, wildlife, alert, attack, nesting, unauthorized, animal, stress, noise, pregnant, fear, loss, forest, young, interfere, habit, condition, health, monitor, scare
#11	Polarized Economy	property, economy, new, farm, value, photography, stress, unemployment, delivery, middle class, job, automation, policy, taxi, polarized, requirement, deliver, item, truck, loss
#12	Illegal Delivery	drugs, cocaine, heroin, truck, smuggling, service, prison, safety, delivery, data, network, security, issue, mobile, money, regulation, illegal, marijuana, work, border
#13	Bad Weather	weather, accident, detect, snow, water, wind, location, bad, crash, property, life, safety, exposure, issue, question, violation, heavy, threat, value, pedestrian
#14	Pollution	noise, pollution, aircraft, sound, environment, annoy, home, out, apartment, visual, stress, issue, shadow, bird, surveillance, delivery, task, forest, annoy, loss, public

The scenarios are constructed on the basis of term results from LSA. However, a question inevitably arises concerning coherent narratives of scenarios through concatenating the words with similar meanings. In fact, similar to the findings from Wang et al. (2012), the terms with similar meanings here do not imply the concept of synonyms, but rather found to be contextual similarities based on the frequency of co-occurrence within documents (Wang et al., 2012). In other words, two terms gain high similarity values in a certain group if both terms are occurring in multiple contexts describing similar topics. By way of illustration, terms "bomb", "terror", and "security" are grouped in the same cluster. The term "bomb" is one of the widely used means for terrorism-related activities, and "terror" and "security" is resulting condition and value that could be threatened from such misconduct. Even though these two terms are not considered synonyms, there exist contextual similarities, or hidden patterns, among them. In light of these grouped terms, fourteen future scenarios are interpreted. Each of the following scenario summaries was developed by three experts of data and futures research, including the author of this research, based on the constituting terms. Related documents were further used to validate the derived label and contexts. Total fourteen scenarios are thoroughly illustrated as follows.

Scenario 1: Biomimetic drones. Drones are becoming smaller and better. In terms of size, they are starting to look and act just like animals or insects that we may not even realize if we are being discreetly watched. In fact, biomimetic drones are not only mimicking the features of birds but also becoming exquisitely small that they may be scaled down to match smallest known insect, often referred as nano drones or insect drones. Moreover, they may be equipped with significantly powerful technology, such as zooming technology and sound minimizing technology. Therefore, the emergence of biomimetic drones has plenty of potential for a malicious behavior of invading privacy and public's safety. In terms of a more comprehensive concept, those drones may cause psychological disturbance to the public, such as stress from bothering and fear of being harmed.

Scenario 2: Corrupted Journalism. As drones are gradually being applied in a wider range of fields, people are concerned about its employment in the contexts of newsgathering and storytelling for journalistic purposes. With the aerial vantages of drones, journalists may detect, take photos, or even film the accident scenes or habits of target individuals, especially celebrities. This becomes even more problematic depending on where it is happening, such as hotel and apartment, and what it is taking, such as private moments and crime-related contents. Such an improper, or morally wrong, use of drone journalism can be quite annoying and upsetting in public's perspective.

Scenario 3: Hacked Enforcement Drones. Drones in the hands of criminals is obviously a serious concern since hackers are capable of jamming the communication links and intercepting secure data flow from police-surveillance drones and thereby analyzing police activities, locations, and procedures. The information being hacked and gained from various types of enforcement drones will be useful for people committing illegal acts. This can surely jeopardize national security. Therefore, drones owned by police authorities must be aware of such vulnerabilities against cyber and security issues, which could consequently lead to a more complicated act of mischief.

Scenario 4: Vehicle Accidents. The proliferation of drones in the sky may pose a significant threat to flight safety. Whilst people operating hobby drones often attempt to exceed the altitude limits, there are possibilities of breaking into flight paths of aircraft and helicopter and thus leading to catastrophic mid-air collisions. Even flying them at the right altitude range causes hazardous situations since it may collide with aircrafts during take-off and landing. It could hit an engine or a fuselage of the aircrafts. In addition, drones may crash into their surroundings or each other. Therefore, there is a need of proper training program or a specific set of flight regulations considering drones to prevent from various drone-induced accidents.

Scenario 5: Task Interference. Drones are expected to be employed for various types of operations in both private and public sectors, such as wildlife management, infrastructure inspection, search and rescue operation, and environmental research. However, as the number of hobby and recreational drones increase, a conflict between those two separate sectors may arise. In future situations where law enforcement drones are dealing with emergency problems, wildlife management drones are tracking poachers, and commercial drones are performing services, somebody's photographing drones may distract operation personnel by disrupting transmission network, alerting the criminals with its propeller noise, and blocking their desired paths, in respective.

Scenario 6: Terrorism. Due to their original and primary purpose of existence, even hobby drones potentially pose a new and unique threat in terms of terrorism threat and assassination. Potentially used by typical terrorist groups or even individuals and entities with bad intentions, drones can be strapped to inert explosives and act as flying bombs. Moreover, drones can be modified to mount attacks with chemical and biological weapons. Without being present at the crime scene, all one has to do is kill switches. Here, the term "culture" acquired high loading values in this group, and this could be interpreted in two different meanings. First, "culture" and terrorism attack are indirectly related based on specific cultural background. Second, this grouping can represent terrorism in terms of cultural artifact destruction as a way of conducting terrorist attacks.

Scenario 7: Prank. Considering the terms highly related to kids and their pranks using drones, there is a growing concern over children's misuse of drones. Such drone involving pranks may seem like petty crime or juvenile prank; however, this could clearly become one kind of security infringements. This, in fact, will provide no less entertainment for certain people since it could come as a terrible shock or frightening display. Once drone involving pranks go beyond a certain degree of seriousness, lives are at stake.

Scenario 8: Transmission Interception. Through flying overhead without being noticed, spy drones owned by certain individuals or the government may send out tricked Wi-Fi network to route people's outbound calls instead of commercial cell towers. Once mobile devices access, either intentionally or automatically, this sham network, all of the personal records, including credit card information and intellectual property, are transmitted to the spy drones. This new type of cybercriminal sabotage may be a serious issue for general publics since they feel unsafe regarding privacy and identity theft.

Scenario 9: Park Wilderness. People tend to appreciate and enjoy being deep in the moments of wilderness, providing an inexplicable sense of serenity. The most widely known place of such an experience is the park. However, buzzing noise from hobby drones attempting to film aerial footage of the nature or climbers will definitely irritate and stress out other visitors by disturbing their private experiences. Based on the terms like “spy” and “follow”, privacy can be threatened as well by being monitored and followed by spy drones.

Scenario 10: Wildlife Disruption. Drones are widely used for observing wildlife or environment monitoring for research activities. Even though drones open up the possibilities for collecting data in a greater distance and with less risk, wild animals may ire toward growing numbers of those drones. Their unique buzzing sound may negatively affect animals' physiological health by increasing the level of stress. This is particularly serious with those that are pregnant or raising young. There exist some possibilities that animals may either get sensitive and be alerted to attack the drones or get scared.

Scenario 11: Polarized Economy. One of the greatest challenges posed from general robotics is that robots and advanced automation may fill in the vast majority of our jobs. The same problem will arise regarding the emergence of drones. Various industries, such as transportation, delivery, photography, and farming, are closely related to the use of drones. Thus, people are concerned with mass unemployment circumstances, especially for middle classes, and economic collapse. This will increase public's economic insecurity and potentially lead to a polarized economy. However, the term "new" brings silver lining around drone technology, since automation may free up to do various new works.

Scenario 12: Illegal Delivery. An autonomous drone delivery service may become a serious issue when the packages containing illegal products, such as drugs, money, or weapons, are flown across the borders. Moreover, whilst law enforcements are capable of tracking the smuggling activities, it would be difficult for them to find out who is responsible for such misconduct. The terms contained various types of illicit drugs ranging from "cocaine" and "heroine" to "marijuana". This issue of illegal delivery causes severe consequences not only for public health and safety but also for national security.

Scenario 13: Bad Weather Condition. Potential users of drones are concerned about how drones could deal with adverse weather conditions during regular flights or delivery services. Drones must be designed to be resistant to heavy snow or rain, storms, fog, or strong winds. Furthermore, a detailed question about dealing with temperature or humidity sensitive packages, particularly in the case of delivery service drones. If not handled appropriately, an event of bad weather may cause ineffective communication and increase the risk of losing control and crashing. Obviously, the possible crash incidents could lead to fatal injury or damaged property.

Scenario 14: Pollution. Once drones dominate the skies in densely populated areas, future consequence on our current environment will be enormous. The public will be stressed out and annoyed by the small aircrafts hovering over wilderness areas, such as home and apartments, whilst generating distressing buzz sounds and shadows. It clearly causes both noise and visual pollution. Even though such drones are busy flying just to satisfy our needs, the moment when drones invade the skies could be the turning point of drones being everyday objects to being the annoying ones.

3.6 Discussion

3.6.1 Categorization of Social Impacts

For a deeper understanding of the derived scenarios of drones, we attempted to elaborate them into broader-level categories. When exploring social impacts in term-level, we observed several terms in multiple scenarios indicating the direct impacts to the individuals, such as feelings, values, or states in a certain circumstance. They even received high loading values within the scenarios. For example, the condition-related term “stress” co-occurred in scenarios of park wilderness, terrorism, wildlife disruption, polarized economy, and pollution. Even though the subjects of getting stress are different, certain scenarios result the change of involved people's psychological state by increasing the level of stress. Moreover, the value-related term, “privacy” occurred in scenarios including, corrupted journalism, biomimetic drones, transmission interception, and park wilderness. This indicates that one of the individual's fundamental human rights, privacy, is at risk from those situations. Rest of the condition- and value-related terms were also observed throughout the scenarios.

This study primarily aims at exploring emerging technology's unforeseen impacts to the society based on the scenarios. Typically, social impacts indicate social consequences to human populations that alter the ways in which people live, work, play, and relate to one another (Vanclay, 2002). However, such terms indicating direct impacts to the individuals provide very specific impacts and thereby results quite skewed social implications in terms of the depth of social consequences. From the given information, we can figure out only the small portion of what are truly out there. By way of illustration, a condition of ‘stress’ caused by park wilderness scenario is clearly a well-defined social impact, yet we cannot spot an increased level of ‘stress’ in a broad sense indicates, and consequently leads to, a threat of ‘mental health and subjective well-being’.

Therefore, for the purpose of conceptualizing social impacts in broader levels, we incorporate conventional concepts suggested by previous SIA literature. Including a variety of social impact investigations and categorizations, the study conducted by Vanclay (2002) is primarily adopted. The author's generic list of social impacts is most suitable regarding our research objective since it explores weaknesses of conventional social impact categorizations and proposes relatively more concrete social impact conceptualization on a basis of previous social impact variables. In fact, the main reason we incorporate conventional social concepts is that we are not capable of broadening the social impacts. According to Vanclay (2002), numerous studies have attempted to explore the full range of social impacts; however, most of them involved charlatan consultants, those with little training in the social sciences, and just applied so-called ‘expert judgments’ to specify likely social impacts (Vanclay, 2002). This paper is no exception. Since we neither are experts nor novices on social sciences, we must refer to studies dealing with fundamental nature of the social impact.

Whilst attempting to relate our scenarios with conventional concepts, we realized that the descriptions of social impacts from existing SIA studies are contextually composed of individual feelings and essential human value related terms. Therefore, both condition- and value-related terms demonstrated the potential of being essential linking factors. To this end, we perform detailed analysis of such terms co-occurring within the scenarios and other SIA-related literature. This provides more comprehensive concepts of social impact, including meso- and macro-level, and offers a better understanding of the overall structure of pertinent social impacts. The entire lists of co-occurring condition- and value-related terms and corresponding meso- and macro-concepts of social impacts are described in Table 3-5.

Table 3-5 List of linking terms and social impact concepts

Types of impacts	Linking terms	Macro-level concept	Meso-level concept
Value-related terms	economy	Economic and material well-being	Economic prosperity
	property		Property value
	ethical	Health and social well-being	Experience of moral outrage
	moral		
	privacy		Privacy
	national security	Institutional impact	Integrity of government
	security		
	environment	Quality of living environment	Environmental amenity value
	safety		Perception of personal safety
Condition-related terms	annoy	Health and social well-being	Annoyance
	bother		
	disturb		
	upset		
	scare	Quality of living environment	Perception of personal safety
	terror		
	frightening		Mental health and subjective well-being
	shock		
	stress		

Table 3-6 provides overall future social impacts regarding drone technology. If drone technology proliferates, our society must be well-prepared to handle four macro social impacts: quality of living environment, health and social well-being, institutional impact, and economic and material well-being. Specific dimensions of involved social impacts can be observed from second to the furthest right column in Table 3-6. In details, if the scenario of Biomimetic Drones becomes reality, people will consequently be in states of scare and bother, and their essential values of safety and privacy will be threatened. Such a circumstance will lead to meso social impacts of negative impact on personal safety, annoyance, and privacy which are the main constituents of the quality of living environment and health and social well-being. This describes only the small portion of entire social impacts from drone technology proliferation. Rest of the details can be interpreted using Table 3-6. Yet, there were several exceptions when utilizing micro-level terms to conceptualize social impacts. First, some of the social impacts are classified differently depending on the technology. The scenarios containing the term "privacy", for instance, were classified in health and social well-being category by considering the features of information technology. Second, the context of the scenario must be carefully considered due to different subjects. The terms like "fear" or "scare", for instance, occurred in both Biomimetic Drone and Wildlife Disruption scenarios; however, the subject is animals for the latter, instead of general public.

Table 3-6 Comprehensive overview of social impacts from drone

Macro-level social impact	Meso-level social impact	Micro-level social impact (Linking terms)	Scenarios
Quality of living environment	Perception of personal safety	safety	#1 Biomimetic Drones #4 Vehicle Accidents #12 Illegal Delivery #13 Bad Weather Condition
		scare/terror	#1 Biomimetic Drones #6 Terrorism #7 Prank
	Environmental Amenity value	environment	#4 Vehicle Accident #9 Park wilderness #14 Pollution
	Adequacy of social infrastructure	national security	#6 Terrorism
Health and social well-being	Annoyance	annoy/disturb/bother/upset	#1 Biomimetic Drones #2 Corrupted Journalism #9 Park Wilderness #14 Pollution
	Privacy	privacy	#1 Biomimetic Drones #2 Corrupted Journalism #8 Transmission Interception #9 Park Wilderness
	Mental health and subjective well-being	stress	#6 Terrorism #9 Park Wilderness #10 Wildlife Disruption #11 Polarized Economy #14 Pollution
		shock/frightening	#7 Prank

	Experience of moral outrage	moral/ethical	#2 Corrupted Journalism #6 Terrorism #8 Transmission
Institutional impact	Integrity of government	security/national security	#3 Hacked Enforcement Drones #6 Terrorism #12 Illegal Delivery
Economic and material well-being	Property value	property	#4 Vehicle Accidents #11 Polarized Economy #13 Bad Weather Condition
	Economic prosperity and resilience	economy	#11 Polarized Economy

Specifically, in terms of intangibility, we were able to foresee fourteen detailed impact-related situations that have not yet happened by generating scenarios on a basis of future-oriented perspectives. In terms of complexity, each scenario was constructed by taking the unexpected involvement of players into account. For instance, the Wildlife Disruption scenario demonstrates that drone technology may affect various animals’ physiological health and thus harmfully impact animal environments. The Polarized Economy scenario predicted public's economic insecurity by presenting very specific industries that may be affected by drone's emergence, including transportation, delivery, photography, and farming. Moreover, some of the scenarios, such as Prank and Park Wilderness, revealed creative and heartfelt concerns from public's perspectives. The Prank scenario discusses the potential hazardous use of drones regarding children's misuse of drones leading to a serious juvenile prank. The Park Wilderness describes a way of triggering one of the modernity's serious issues: disturbance of private experiences. Such scenarios may hardly be derived from experts’ viewpoints, and they were, in fact, extracted mainly from the publics’ insights when the topic's related documents were reviewed.

3.6.2 Comparative Analysis

To further strengthen internal validity and generalizability of the proposed result, a comparative analysis was conducted with a conventional technology assessment study. We have conducted an empirical investigation to identify any scientific research or futures reports with the most pertinent purpose. The result from the study presented by Rao et al. (2016) was chosen for the comparison, since it had aimed at serving an exact same purpose but with a very different approach. The paper has used discourse analysis and explored the impacts that commercial drones have had on various societal aspects. It further offers recommendations for various stakeholders, such as practitioners, policy makers, and researchers. There exist two major differences compared with our proposed approach. First, we have utilized relatively prospective perspectives for exploring societal issues; whereas, the mentioned study has used more retrospective data sources of academic and non-academic databases. Second, our proposed approach is based on a text mining technique, but the study uses a discourse analysis, which qualitatively investigate spontaneous discourse of articles, blogs, books, reports, or any text produced by an individual or an organization. The social impact-related results are presented in the correspondence with our proposed result, as follows.

The first societal impact is associated with safety, and this has been caused from airworthiness of drones and unsafe practices. The study has suggested possible situations with electric poles and airports. The second impact is about security, which could result from jamming/spoofing and malicious usage. Third, the value of privacy can be degraded from blurring of private and public spaces and invading private properties. Other specific scenarios include airspace interference, ownership of data, and commercial liability; however, no specific societal values affected from such eventualities were not further mentioned.

The presented study has supported their assertions by providing pertinent historical data; whereas, our proposed approach has as evidently lacked some reliability within the data sources and the results. Despite its limitation, the proposed data-driven approach holds four distinctive strengths: detailed storylines, future-oriented contents, and extensive societal impacts. First of all, the result of this research offered more detailed contents for describing certain situations. Hacked Enforcement Drones scenario, for instance, illustrates similar problematic situation with spoofing-related situation in the presented study. The proposed methodology has illustrated with more details by mentioned the malicious use in police-surveillance and how it could jeopardize not only public but also national security. Second, the result of our approach has provided a more diverse types of problematic situations. The privacy invading issues, for example, were described with four different specific scenarios. Third, the result is relatively more futuristic. Biomimetic Drones scenario, for instance, offered a quite forward-looking situation, where the drones are becoming smaller and smarter and causing more diverse problems. Finally, societal impacts can be captured in a more systematic and comprehensive manner. Every problematic situation is involved with diverse societal problems, and it is quite difficult and challenging to relate with single degraded value. Moreover, societal impacts are intertwined in a very complex manner. It is essential to analyze and conceptualize them in a more systematic and comprehensive manner, as presented in the results derived from our data-driven approach.

3.6.3 Implication for Theory, Practice, and Policy

Our proposed approach provides an extensive understanding of possible scenarios regarding drone technology, and this thus has a novel implication for theory, practice, and policy. First, the application of LSA text mining technique in exploring and assessing the future of technology contributes to the foundation of futures research, such as scenario planning, technology foresight, technology assessment. Traditionally, most scholars have been focusing only on qualitative approaches. They gathered reliable information directly from a handful number of stakeholders by organizing workshops and were able to produce coherent outputs. However, such a participatory method can be considered as a laborious process since it requires weeks or months and no one can assure the meaningful result. This paper, in response, approached in a more quantitative manner to expand the pool of collective intelligence and thereby propose an alternative or a supporting method in futures research. The proposed process can be seen as a systematic overlook of futuristic perspectives on the impact of emerging technologies, and this is expected to take us one step closer to gaining better insights into unfolding the future.

Second, the end result of this study can be served as an essential tool for those seeking for wide-ranging uses of new and emerging technology, such as potential end users or tech-companies. People craves for new innovative technology; however, the most crucial factor that pushes away from wide usage is the distrust of technology's unfamiliarity. Despite of countless benefits from the use, public disguisedly holds various concerns and even fears regarding the emergence of new technologies. In such a circumstance, it is now tech-companies' turn to preemptively identify unforeseen impacts and relieve public skepticism from the unfamiliar. This may stimulate better user acceptance and thereby achieve business success.

But even more fundamentally, proper regulatory governance system must be in place for new technologies to be introduced to our society in the first place. Even though there exist numerous innovators who are at the forefront in commercializing emerging technologies, no technology could be widely diffused unless the government paves the way with right policy system (Ruan et al., 2014). Nonetheless, how are they expected to make good policies when the possible outcomes are unknown? Here, technology foresight activity is considered as a critical success factor in decision makings of policy design due to its capability of actually providing supplementary information for actionable recommendations (Hassanzadeh et al., 2015). Through the proposed approach, the scenarios and the final categorization of social impacts offer the general overview of potential outcomes and their specific impact on our society. It is hoped that such an ex-ante examination supports policymakers to design more balanced regulations between tech companies and potential consumers.

3.7 Conclusion

The development of new technologies raises, whether intended or unintended, a wide variety of societal and epistemological consequences during their integration into society. However, an ex-ante examination of prospective impacts is quite difficult to perform due to distinctive nature of emerging technologies: the social implication is not seen or known unless the technology is brought to a particular context (Halaweh, 2013). This paper thus proposes a novel method of envisaging future society once new technology proliferates and conceptualizing pertinent societal impacts. To this end, we take full advantage of ICT into our foresight activities. A great number of players, such as futures experts and general publics, are complaining early about potential unforeseen impacts of emerging technologies in numerous futures websites. We employ such *future-oriented web data* and mine frequently occurring but meaningful contextual patterns using LSA text mining technique. When assessing coherency of each terms clusters, some of the terms indicated individual's conditions or moral values, and they happened to co-occur in conventional SIA literature. Therefore, we further utilize condition- and value-related terms as the key linking factors and directly relate to broader dimensions of social impacts: meso- and macro- level social impacts. The effectiveness of our approach was demonstrated with a case of drone technology. Our case study illustrates that integration of drone technology into our lives may entail some societal impacts in terms of quality of living environment, health and social well-being, institutional issue, and economic and material well-being.

Taken together, the proposed process and the result have a novel implication for theory, practice, and policy. First, the methodology was based on LSA text mining technique and *future-oriented web data*. This may contribute to the foundation of futures research since such a quantitative foresight activity could be served as an alternative way of understanding the future depictions of emerging technologies. Second, the result can be an essential tool for those wishing for emerging technology to be widely spread in the near future, such as potential consumers and tech-companies. In the perspective of end users, pre-emptive identification of unforeseen social impacts may alleviate public skepticism and distrust of unfamiliar technologies. Likewise, such a better user acceptance and reactions could promote private sectors to achieve product or service success (Halaweh, 2013; Hernández-Ortega, 2011). Finally, from policy making perspective, this information may support policy-related decision makers to promote better regulatory policies of emerging technologies and thereby proliferate them in a sustainable and socially friendly way.

As a final remark, we would like to note the importance of intelligent interpretation on the part of the researchers prior to and post analytic process (Evangelopoulos et al., 2012). This perhaps can be considered as inherent limitations exhibited in proposed approach, or in LSA itself. The scenario generation and impact conceptualization process cannot be done fully automatically. There is in need of human intelligence and manual procedure, particularly when useful terms are filtered out prior to LSA or when derived scenarios are interpreted based on constituting terms. Moreover, due to lack of expertise, there are possibilities of missing out some of the essential value- or condition-related terms, which are the key ingredients for conceptualizing social impacts. This may have resulted less-detailed and less-comprehensive social impact concepts. To sum, this research provides a slight possibility of envisioning future society and developing scenarios in a more quantitative way. However, we have only scratched the surface of foresight and social impact assessment development, and significant improvement remains to be seen for the future work.

Chapter 4

Foresight for Impact Analysis

The second step of proactive management process is to analyze and specify future impacts of emerging technologies. Potential issues are identified in the previous step; however, we have not yet observed the details or the causes of those problems mainly due to methodological limitation. An in-depth investigation of those specifics is essential for understanding their nature, thereby supporting decision makers more aware of, and responsive to, various unexpected ramifications. Such a research question has been dealt in the field of scenario building, which is a widely acclaimed activity in technology foresight. It is about systematically constructing storylines about possible complex and uncertain futures (Peterson et al., 2003). However, one of the main limitations with conventional scenario building approaches is determining the right balance between creativity and plausibility when constructing scenarios.

This chapter presents a data-driven foresight approach to cope with this specific issue in scenario building. In details, *future-oriented web data* is applied to gather futuristic insights of various experts and the public. LSA text mining technique is used to capture generalized and compromised view of the future; IdeaGraph algorithm is further conducted to supplement creative and unexpected perspectives. The proposed methodology has incorporated a wide range of opinions, including both the many and the few, for the purpose of understanding the detailed future impacts of emerging technologies.

4.1 Introduction

New and emerging technologies exist for many of today's ills that people find unpleasant. Yet, every good there is a dark side. Google search engine, for instance, has made knowledge searchable by allowing us to retrieve any piece of data on the World Wide Web (WWW); however, the company is raising serious concerns regarding the protection of information or civil liberties and is further emphasizing their responsibility in new technology development. As such, socio-technical features of emerging technologies have been the core and essential subject of much debate in both in academic research and markets initiatives (Rotolo et al., 2015).

To manage the problems in a proactive and preventive manner, companies must be able foresee every potential disruption that could occur from the use of emerging technologies. However, such an issue must be handled somewhat differently than any typical forecasting issues based on two different perspectives. First, emerging technologies are associated with very distinctive features compared to established or existing technologies: *uncertainty* and *complexity*. The envisioning process must carefully take these characteristics into account and extensively discover uncertain factors and their complex relationships involved in the future of emerging technologies. Second, the principle of continuity in a technology's social and economic features is no longer applied in the envisagement of socio-economic future. Since it is the most fundamental premise that makes technological forecasting, like generic extrapolation and projection, possible, there is in need of a more prospective approach, rather than a retrospective one.

Technology foresight is one available solution in such a circumstance. It highlights the collection of wide ranging visions in identifying future technology developments and their interactions with society for a more desirable future (Barre et al., 2008). Evolved as parallel and sometimes overlapping with the field of technology foresight, numerous different notions were also derived, including technology assessment and impact assessment. These studies systematically examine unintended and unforeseen effects on society that may occur when a technology is introduced, extended or modified (Coates, 1974). In methodological perspective, they are all on the basis of one futures methodology: scenario building. Scenario building process is generally used to create a set of alternative futures while being aware of the uncertainties (Kamprath & Mietzner, 2015). This technique is exceptionally useful in times of uncertainty and complexity (Amer et al., 2013).

However, conventional method of this analysis has two major limitations: (1) a difficulty in ensuring both plausibility and unexpectedness within the scenarios and (2) a high reliance on participatory way of building scenarios. First, determining the right balance between these two conflicting features is essential; however, there is no concrete way of properly realizing them. Second, scenarios are normally generated using Delphi, interviews, workshops and experts' discussion. Although it may ensure reliability of the data source or high plausibility, both information gathering and actual scenario building process require much time and effort.

In response, this article proposes a *data-driven scenario building process*, designed to better cope with *uncertainty* and *complexity* of emerging technologies. In other words, both unknown factors and influenced agents associated with the future of emerging technologies are extensively explored, and their inter-relationship is analyzed to systematically interpret the scenarios derived from text analysis. Our ultimate objective is to propose an alternative or supplementary way of building well-balanced storylines equipped with a much of plausibility but at the same time a fair amount of unexpectedness, but with the purpose of envisioning unforeseen societal ramifications of emerging technologies. This proposed approach, thus, constructs the main and general plot of scenarios by grouping high frequency terms and further supplements latent and unexpected details by identifying low frequency terms closely related to the general plots. LSA and IdeaGraph text mining is applied to develop the general plots and the latent details. The incorporated data source is *future-oriented web data*, which encapsulates a wide variety of future challenges of emerging technologies.

4.2 Uncertainty and Complexity

What sort of technologies are we currently handling? Numerous studies have attempted to explain the distinctive characteristics of emerging technologies compared to that of established or existing technologies (Halaweh, 2013; Köhler & Som, 2014; Kwon et al., 2017; Rotolo et al., 2015). For instance, Halaweh (2013) conceptualized the characteristics of emerging technologies into uncertainty, network effect, costs, unobvious impact, limited creator, and lacking investigation. Rotolo et al. (2015) conducted a more extensive empirical research and identified five major attributes, including radical novelty, relatively fast growth, coherence, prominent impact, and uncertainty and ambiguity. Both academia and industries are constantly craving for a clear conceptualization of the future of emerging technologies. Even though all these features make emerging technologies deemed worthy of an in-depth investigation, we argue for the necessity of foresight to reflect upon a more fundamental factors: *uncertainty* and *complexity*.

Emerging technologies are associated with unknown, unpredictable, and volatile factors interacting in unanticipated ways (Day & Schoemaker, 2004). In details, the uncertainty of emerging technology poses a threat of not knowing what comes in the near future and not being able to proactively determine the success of a technological path (Meijer et al., 2007). This is normally caused by a lack of common knowledge and agreement about what a new technology will be relevant in the future (Halaweh, 2003). Since uncertainties exist in different domains, including technology, resource, supplier, consumer, politics, and environment, it is important to identify various sources which the actor is uncertain about (Dickson & Weaver, 1997; Meijer et al., 2007). Furthermore, there exist a complex network effect, or a cause-effect chains, which are not pre-given but evolves over time when new and advanced technologies are applied in societal and economic realities (Köhler & Som, 2014; Rotolo et al., 2015). They hold highly non-linear, dynamic and complex relations with a wide variety of uncertain structures.

Generally, the conceptualization and assessment-related studies of emerging technologies, such as measuring novelty or forecasting technical growth of a new technology, can only be clearly determined if we first understand a wide-ranging set of factors, actors, players, or stakeholders and their inter-relationships with one another. The company that is about to launch a new technology product, for instance, may assess various prospects of new artifacts, such as novelty, customer expectation, or potential implication. The methods of market research, competitor analysis, or scenario analysis are normally conducted in response. However, they require ex-ante investigations of acquiring information regarding potential markets, stakeholders, or customer types and their inter-relationship with one another. Much of the pertinent literature highlight that a successful anticipation requires the understanding of the dynamics of promising, but uncertain, constituents that shape technological futures (Borup et al., 2006; Stilgoe et al., 2013). Such uncertain components and their complex relationship are the key factors we attempt to focus when describing the thoughts towards the future of emerging technologies.

4.3 Data-driven Foresight Process

There are no statistics on the unpredictable future of societal processes, which are stemming from social, economic, and cultural dynamics (Durance & Godet, 2010; Sollie, 2007). When dealing with such a problem, individual judgments and perspectives are often the most reliable source to implement (Durance & Godet, 2010). The analysis of multiple prospects and visions are of paramount importance in evaluating the potential impacts (Rotolo et al., 2015). This section discusses the literature review regarding previous attempts of exploring social impacts of emerging technologies.

Whilst these diverse concepts were derived to study societal impacts of emerging technologies, a whole new disciplinary field of technology foresight was introduced in the late 1990s for the purpose of foreseeing future implications of new technological innovation upon society. Rooted in futures studies and technology forecasting, technology foresight is also partly derived from TA on the basis of a common set of goals (Miles, 2010). It is defined as ‘a systematic process to look into the long-term future of science, technology, economy, environment and society’ (Reger, 2001). Technology foresight is conducted in a collective and participatory manner, and it has supported government, industry and other organizations to better shape the future of our society (Saritas & Nugroho, 2012). It utilizes and networks disparate sources of futuristic knowledge to envision a wide variety of possible depictions. Although the activities include both qualitative and quantitative means, most of the basic foresight exercises depend on participatory tools: surveys, expert discussion, interview, workshops and panels.

Recent evidences have been stressing the need of flexible, open, and powerful foresight system that incorporates diverse mining techniques, including text mining, data mining, semantic web, and artificial intelligence (von der Graht, 2015). Such a methodological shift seems meaningful; however, a tremendous amount of effort is still needed to properly incorporate these mining techniques. Such a gap could not ever be closed when scenarios are built solely based on participatory approaches. The challenge is two-fold: quantity and quality of data. First, due to the unavailability of futuristic information regarding societal context, the process is relied on the opinions of a handful number of experts. However, the amount of the acquired data from such a process may not be sufficient enough to apply mining techniques and derive meaningful information. Second, a heavy reliance on the experts may result subjective and self-fulfilling information retrieval. Such a biased and limited scope of perception may result obvious and generalized future depictions, rather than unanticipated and surprising depictions.

This research, thus, proposes a data-driven foresight process to realize a more 'flexible, open, and powerful' system. It involves *future-oriented web data* and two text analysis techniques of LSA and IdeaGraph. In terms of data sources, *future-oriented web data* encapsulates the perspectives and perceptions of people from different expertise and backgrounds, such as engineering, sociology, psychology, or economy, regarding the potential implications of new and emerging technologies. In terms of text analysis, the mining technique is capable of not only summarizing a large number of opinions into several generalized concepts but also capturing latent or hidden details based on quantitative measures. Such a data-driven approach could play a significant role in promoting a shift from just ordinary scanning and systematic information retrieval to actual data analysis, interpretation, and implementation in foresight exercises (Keller & von der Gracht, 2014).

4.4 Scenario Building Beyond the Obvious

Scenario building is a widely acclaimed activity in TF exercise, and it is one suited method to cope with the abovementioned features of emerging technologies. Scenarios do not present the most accurate and probable state of a single outcome (Peterson et al., 2003); they are not prognoses or predictions (Kamprath & Mietzner, 2015). They are rather alternative dynamic stories that incorporate a variety of detailed ingredients of uncertain futures. In comparison of traditional predictions, forecasts, and projections, it is a powerful approach in managing accelerated change, greater complexity, and genuine uncertainty based on systematic and unconstrained thought process (Ringland, 1998; Gausemeier et al., 1998; Lindgren & Bandhold, 2003). Other than scenario building, there exist similar scenario-related studies like scenario development and scenario planning. Scenario planning generate future storylines and further suggest strategic planning implications. A more comprehensive course of study is scenario analysis. Scenario building is considered the most fundamental step to all scenario-related studies. The unique strength of using scenarios is that such methodology is capable of envision not only technical but socio-economic implications of emerging technologies. Scenarios are referred as the main methodological inspiration for bringing in the societal context into envisioning process and revealing potential trend break, discontinuities or emergence phenomena (Carlsen et al., 2014).

Numerous methodologies have been proposed in futures research to better cope with extraordinary uncertainties and surprises; however, most were subject to participatory approaches. Based on the review studies of Amer (2013), scenario planning has been developed into three schools of major approaches: (1) intuitive logics, (2) probabilistic modified trends (PMT), and (3) la prospective. First, the approaches of intuitive logics do not incorporate any mathematical algorithm. They are considered as qualitative techniques, which are based on generic tools like brainstorming and STEEP analysis. Second, PMT is a quantitative method that involves the probabilistic modification of extrapolated trends. However, the approaches are still heavily relying on expert judgments when constructing initial database, such as future trends, occurring probabilities, or event inter-dependencies. Third, la prospective is a combination of qualitative and quantitative tools that puts its purpose on developing normative and idealistic future images serving as a guiding visions to policy makers.

There exist numerous studies attempting to explore socio-economic ramifications of emerging technologies using scenarios. Wardak et al. (2008) use scenario analysis approach to identify potential environmental risks of nanotechnology. The study gathers publicly available data and uses expert elicitation to obtain scores on likelihood and severity of those risks. Stemerding et al. (2010) suggest techno-ethical scenarios to demonstrate a systematic way of exploring moral concerns of genetic susceptibility screening technology. And most recently, Carlsen et al. (2014) identify ethical threats posed from future domestic robots through organizing iterative participatory workshops. In result, three detailed scenarios are demonstrated to depict society's responses to new technologies. As seen, conventional studies had heavily relied on participatory approaches. The process is evidently time-consuming and labor-intensive; however, the most crucial problem arising from this approach is inevitability of cognitive biases. When opinions are openly shared with one another, perceptions tend to lean towards a certain way, resulting either excessively obvious and general scenarios or excessively absurd and unreasonable scenarios. There is, thus, in need of a more systematic and forward-looking way to overcome the drawbacks of previous approaches.

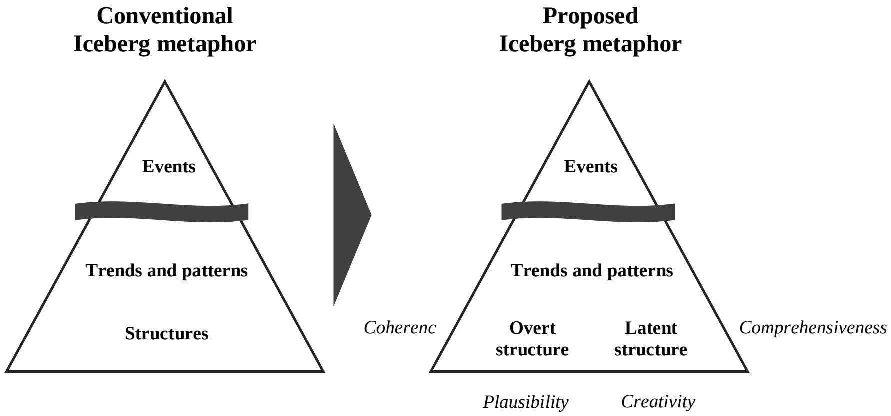


Figure 4-1 Illustration of the proposed Iceberg metaphor

To this end, this research attempts to modify the existing structure of scenario thinking and building process through adopting a data-driven foresight approach. There exists a common underlying principle within scenario-related studies: the iceberg metaphor. As shown in the left iceberg of Figure 4-1, what we can eventually observe are the events and actions composed of unknown structures, including societal, technical, political, environmental, or economic factors, and their causal patterns (Chermack et al., 2001; Lindgren & Bandhold, 2003). It is of vital importance to adopt network-based thinking for identifying unknown structures and their latent relationships at lower levels that drive future systems (Gausemeier et al., 1998; Markmann et al., 2013).

The proposed scenario building procedure is on the basis of iceberg metaphor but with a more specific objective of "envisioning future beyond the obvious". Numerous studies have suggested multiple evaluation criteria that ensure the scenarios to be formed in an adequate basis for making proactive decisions. Many have highlighted plausibility, consistency/coherency, creativity/novelty, relevance/pertinence and some have underlined other supplementary features, including comprehensiveness, importance, transparency, and completeness (Amer et al., 2013; Druance & Godet, 2010; Coates, 2000). Whilst all these features are essential, this article specifically aims at constructing well-balanced scenarios with maximized creativity and comprehensiveness while maintaining plausibility and coherency, as illustrated in Figure 4-1. Determining the right balance between creativity and plausibility is quite a difficult, but essential, task in scenario development. The majority (85%) of content pitfalls of scenario generation is driven by a lack of creativity; on the other hand, the remainder are the results of excessive amount of creativity of the participants in the process (MacKay & McKiernan, 2010).

The proposed methodology follows a basic story-building process. A storyline, in general, must be structured with a main plot and various details and descriptions, in order to become a whole narrated event. In this same manner, a general frame or plot is first developed based on overt structures, composed of the opinions of the many (high-frequency terms) and subsequently supplemented based on latent structures, composed of the opinions of the few (low frequency terms). Rather than focusing on just general structures, both overt and latent structures are sequentially derived to satisfy the abovementioned features, as shown in the right iceberg of Figure 4-1. As mentioned earlier, if we follow the crowd we probably will not be able to acquire creative and unexpected ideas; if we focus too much on the few, we probably will end up with absurd and meaningless thoughts. Such a step-by-step procedure may not only best balance two contradicting features, plausibility and creativity, but stimulate coherency and comprehensiveness in scenarios by capturing the voices of the many and the few.

4.4.1 Capturing Plausibility using LSA

Scenarios are utilized to answer 'what if…?' questions and to consider contingency, what is likely, what is plausible, and what is possible (Stilgoe et al., 2013). The feature of plausibility is the most fundamental and essential prerequisite, highlighted by numerous literature (Chermack et al., 2001; Bradfield et al. 2005). A recent study by Amer et al. (2013) reported that out of 14 out of 15 prominent scenario planning articles identified plausibility as vital criteria for high quality scenarios. This feature is normally ensured by following two primary conditions using morphological analysis (Eriksson & Ritchey, 2002; Amer et al., 2013): *logical relativeness* and *compromised perspective*. First, influence factors or driving forces, which are the components of constructing scenarios, are assessed from one another based on their consistency, and only their logically possible configurations are chosen to become conceivable scenarios. Second, the analysis is a qualitative process, involving a number of iterations are essential to reconcile different viewpoints. This process of plausibility negotiation eventually is another form of reaching consensus on certain problem definition (von Schomberg, 2011).

In order to capture plausibility in a more quantitative manner, the proposed methodology adopts topic detection technique using LSA. It is a method of reconciliation considering that there exist a number of different voices that need to be observed in the future-oriented database. The utilization of LSA satisfy both of the abovementioned criteria: logical relativeness and compromised perspective. First, the documents are imbued with the terms in a logical structure or perhaps in a contextual structure. LSA is capable of modelling latent semantic relations between words and contexts (Buljan & Nemet, 2014), and we assert that factor configurations based on logical relativeness is rationalized by the terms, which are considered to be the constituents of scenarios, being grouped based on their contextual relativeness. Second, LSA minimizes the impact of noises by retaining only the elements with higher information content (Buljan & Nemet, 2014). In order for a term to contain high information, term frequency must be highly relative to the number of documents containing the term. This process inevitably results the scenarios to be composed of relatively high-frequency terms. These frequently occurring opinions are considered as the compromised view of stakeholders, and this study utilizes them as overt structures that depict generalized and apparent glimpse of the near future.

4.4.2 Capturing Creativity using IdeaGraph

Creativity has been frequently acknowledged as a significant factor in scenario development by numerous literature (Amer et al., 2013; Van der Heijden, 1996; Durance and Godet 2010). It may not seem like a fundamental prerequisite; however, we assert that creativity must be emphasized in this particular research, where unforeseen impacts are explored as diverse and detailed as possible. Furthermore, creativity is vital in anticipating risks (von der Gracht & Stillings, 2013), since creative ideas and thoughts yield various surprises and unexpectedness that the majority cannot imagine. According to the study of MacKay and McKiernan (2010), only the scenarios that follow the nature of 'duality', the embracement of both novelty and utility, can be interpreted as creative depictions. The feature of creativity is thus captured through following these two primary conditions: novelty and utility. Nonetheless, it is still challenging to differentiate creative ideas from outliers and noises since they exist only in a small number. Extreme positions and unique arguments, for instance in risk analysis, are highly valued for early risk identification; however, they are often considered as outlier opinions and neglected in the statistical analysis of surveys (Markmann et al., 2013). Creativity is closely related to the concept of individualism, as in the isolated and disconnected from the majority (MacKay & McKiernan, 2010).

In response, a quantitative approach is utilized to analyze a vast amount of future-oriented voice. Creativity is captured by adopting the logical and methodological basis of IdeaGraph technique in the proposed methodology. IdeaGraph has been widely adopted in the field of Knowledge Discovery to detect rare but significant events or situations by mining texts and visualizing in a graph-based map (Wang et al., 2013). The latent information derived from such a quantitative and systematic process is considered as the creative factors, and they ensure both of the abovementioned features: *novelty* and *utility*. First, novelty is measured based on whether the factor is new and relatively unusual. The proposed methodology identifies the factors, or the terms, which occur infrequently. In other words, assuming that original and brilliant minds exist in a small number, the terms with low-frequency values are selected for the candidates of *latent structures*. Second, utility is measured based on whether the factor is valuable to become the additional details for *overt structures*. The methodology determines only the useful factors using their contextual relations to the overt structures comprising each scenario plot. These infrequent but important opinions are considered as the brilliant and creative view of stakeholders, and this study utilizes them as latent structures that supplement unexpected and surprising details of the generalized future depictions.

4.5 Research Framework

The proposed methodology is aimed at developing a well-balanced storylines equipped with a much of plausibility but at the same time a fair amount of unexpectedness. Figure 4-2 illustrates the overall process composed of three main stages: pre-analysis, text-analysis, and post-analysis. Pre- and post-analysis stages are data preparation prior to actual analysis and analytic interpretation of the resulting scenarios derived from the analysis, in respective. The scenarios are constructed based on a consecutive development of overt structures and latent structures.

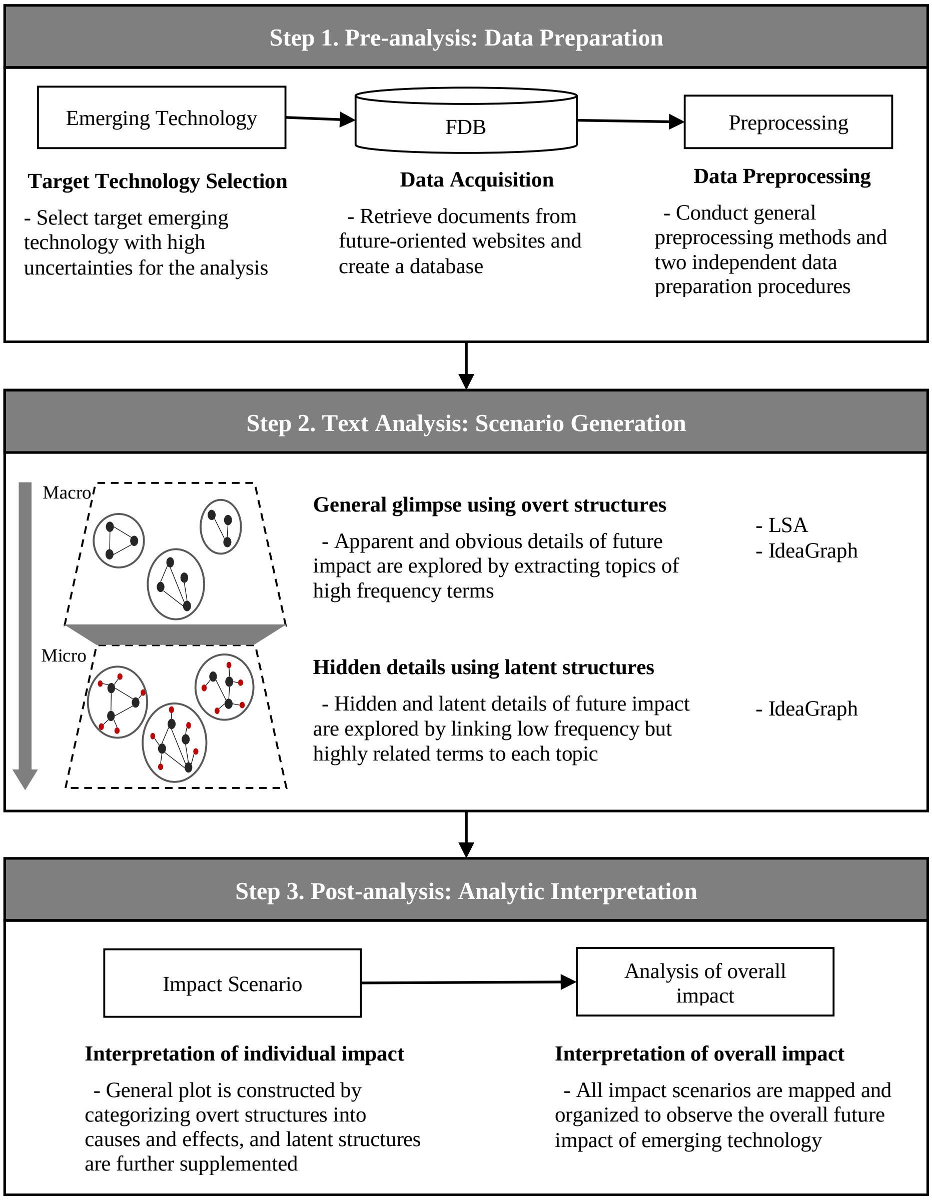


Figure 4-2 Overall framework of the proposed methodology

4.5.1 Step 1. Pre-Analysis: Data Preparation

4.5.1.1 Target Technology Selection

A target emerging technology is first chosen for the analysis. Any innovative technologies or emerging capabilities on the verge of making a difference to society are suitable. In fact, the more application uncertainties and diverse unknown unknowns, the more unique and surprising results are expected.

4.5.1.2 Data Acquisition

The textual database is constructed based on future-oriented websites discussing both attractive prospects and undesirable future of the target emerging technology, including futures report, future technology news article, or futures forum. Depending on the types of data source, the result of the analytical exploration will be greatly varied. The websites mainly composed of players with high expertise, including futures report or technology news articles, will provide quite general but reliable information. On the other hand, the websites largely composed of player just with high interest in future society, including futures forum, will offer mostly unstructured but interesting information. The database can be first created based on various research purposes; however, this article incorporates both types of websites and creates the database involving as much diverse perspectives as possible. To retrieve concerns and worries regarding the impact of emerging technologies, the process of selecting search queries, as proposed in section 3.4.1, is utilized.

4.5.1.3 Data Preprocessing

Data preprocessing step is essential prior to textual analysis. General preprocessing methods, including term extraction, stop word removal, tokenization, and part-of-speech tagging, are first applied. Then, two different preprocessing procedure is required for Stage 2. The first step is associated with

LSA, which groups the terms through analyzing the semantics of documents. The terms must be, thus, structured into a term-document form. Stop word removal, the term filtering process, must be deliberately conducted for developing overt structure scenarios based on two criteria: frequency value and context pertinence. Since the objective of Step 1 is acquiring generally accepted opinions, only the terms above certain pre-determined threshold must be selected for further analysis. Furthermore, the terms with high relevance on the target technology are properly filtered out. The information of the selected terms and documents are then converted into vectors of term-document matrix, the matrix (m-by-n) is represented based on frequency in which a term (m) occurs in a document (n). Finally, term vector weighting process is performed in order to transform raw values into weighted values. The product of the term frequency and the inverse document frequency (*tf-idf*) is applied to highlight the terms that better characterize the document content. The steps incorporated by IdeaGraph technique are associated with sentence-based contextual relations among the terms. Determining relations in a sentence-level is selected since it yields a more accurate interdependencies among the terms, when compared with that in a document-level. The texts are, therefore, preprocessed into basket datasets.

4.5.2 Step 2. Text Analysis: Scenario Building

4.5.2.1 General Glimpse using Overt Structures

Two text mining techniques, including LSA topic detection technique and the basic premise of IdeaGraph, are applied to generate topics with overt structures. The objective of Stage 2.1. is to capture generalized and apparent glimpse of the future through extracting relatively frequently occurring opinions and analyzing their relations. The resulting network is utilized to observe the main plot or core frame of impact scenario.

In details, LSA is applied to extract a set of frequently mentioned opinions/voices, the overt structures, by grouping high frequency terms comprising the texts based on their latent semantic/contextual patterns. In details, singular value decomposition is used to decompose the weighted matrix ($X_{t \times d}$); however, instead of projecting all the unnecessary details, certain k value is selected not only to maximize the inclusion of significant information but to minimize the unnecessary information (Kwon et al., 2017). Once it is decomposed into T, S, and D, the topics derived in TS matrix are thoroughly interpreted. In result, each term will be assigned with certain loading value, which characterizes the importance to the corresponding topic. Different threshold values are determined for each topic to ensure the inclusion of only the contextually significant and consistent terms. This step is similar to what is done in conventional factor analysis (Sidorova et al., 2008). The terms are grouped based on their contextual similarity values, and each topic is defined based on its high-loading terms. However, an experts' discussion is required to settle different interpretations of each general plot.

At a glance, it is difficult to instantly grasp the main idea of a scenario by solely examining generalized and apparent information. A co-occurrence measure, which is on the basis of IdeaGraph technique, is applied to examine the interdependencies of overt structures. This yields a clearer understanding of the general contents. The relation of overt structures is calculated using *overt relation measure* $OR(I_i, I_j)$. The relations of two terms of overt structures I_i and I are measured based on their co-occurrence relationship within each sentence.

$$OR(I_i, I_j) = \sum \min\left(|I_i|_s, |I_j|_s \right) \qquad [4\text{-}1]$$

$|I_i|_s$ indicates the number of times a term I_i occurs in a sentence s, and relation OR is computed by the summation of values of all involved sentences. After ranking by their values, only top N_R term-pairs are represented as links. It ultimately offers a network-based visualization output to support understanding detailed relationship of high frequency terms, reducing subjectivity in building descriptive contents of impact scenarios.

4.5.2.2 Hidden Details using Latent Structures

The network of overt structures is then supplemented with latent structures in order to be formed into an impact scenario, as depicted in Figure 4-3. Latent structures are low frequency but significant and meaningful terms considering the entire contents of basic overt-structures, and they are derived using IdeaGraph technique. They are captured to supplement hidden and detailed descriptions of the overt structures based on the terms' co-occurrence relationships.

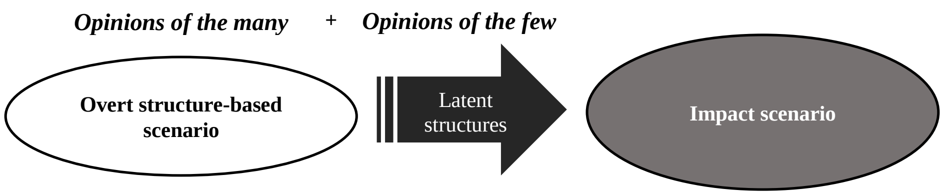


Figure 4-3 Process of building impact scenarios

Latent structures are extracted based on *latent relation measure*. The relations of two terms I_i and I_j are derived from the transfiguration of the concept of conditional probability, as shown in [4-2].

$$LR\left(I_i, I_j\right) = \frac{P(I_i \cap I_j)}{P(I_i)P(I_j)} \qquad [4\text{-}2]$$

$P(I)$ corresponds to the probability of a given sentence containing term I_i, and $P\left(I_i \cap I_j\right)$ indicates the probability of a given sentence containing both terms I_i and I_j. *Latent relation measure* LR is capable of identifying latent opportunities regardless of the low absolute frequency. That is, term I_i and I_j will have strong relationship as they frequently appear in a same sentence. We, however, are focused on retrieving high relation values of low frequency terms, and the simple adoption of joint probability, $P\left(I_i \cap I_j\right)$, of two terms cannot properly satisfy our needs. This research locates $P(I_i)$ and $P\left(I_j\right)$ in the denominator in order to increase the chance of low frequency terms to attain high LR values and to prevent LR from being dominated by high frequency terms. The multiplication of those two probabilities, rather than the addition or $P\left(I_i \cup I_j\right)$, minimizes the influence of high frequency values.

The conventional measure of IdeaGraph was modified based on two criteria: (1) maximizing the relation value when a low frequency term is involved and (2) minimizing the relation value when a high frequency is involved. As described in Table 4-1, second and third row correspond to the former condition and the last row corresponds to the latter condition. What is interesting in this data of the first condition is that the proposed measure obtained the highest amount of change, the percent an amount changes from the original joint probability, when compared to conventional measure. In addition, it was seen that the value of alternative measure did not change much when probability of low frequency term, $P(I_i)$, was decreased. Only proposed and conventional measures satisfied the first condition. To observe how the measures handled high frequency terms, we have increased the probability of latent structure into 0.9. The amount of change for the proposed measure was much lower compared with conventional measure. In result, we have adopted the measure that maximized the value with a low frequency term and minimized the value with a high frequency term, and the relations of all terms extracted from the preprocessing step are measured.

Table 4-1 Example presenting rationale for modifying conventional measure

latent structure $P(I_i)$	**overt structure** $P(I_j)$	**relation** $P(I_i \cap I_j)$	**conventional measure** $P(I_i \vert I_j) + P(I_j \vert I_i)$	*amount of change*	**proposed measure** $P(I_i \cap I_j) / P(I_i)P(I_j)$	*amount of change*	**alternative measure** $P(I_i \cap I_j) / P(I_i \cup I_j)$	*amount of change*
0.05	0.7	**0.03**	0.64	20.43	**0.86**	27.57	0.04	0.39
0.1	0.7	**0.05**	0.57	10.43	**0.71**	13.29	0.07	0.33
0.9	0.7	**0.5**	1.27	1.54	**0.79**	0.59	0.45	-0.09

Next step is aimed at filtering out only the significant ones, the terms assigned as latent components in impact scenarios. The relative importance is useful in this circumstance. Each term and the terms comprising each overt structure scenario are measured based on the *latent structure measure*, as shown in [4-3].

$$LS(I, OS_i) = \sum_{I \notin OS_i, I_{OS} \in OS_i} LR(I, I_{OS}) \qquad [4\text{-}3]$$

LS corresponds to the summation of LR values between term I outside overt structure scenario OS_i and every term I_{OS} within overt structure scenario OS_i. All the values of term-scenario pairs are sorted for each scenario, and only the top N_{LS} pairs, which are derived from a predefined threshold value, are selected as latent structures. When supplementing into a network, only the top N_{LR} pairs extracted from another threshold value are linked from one latent structure to other overt structures.

4.5.3 Step 3. Post-Analysis: Analytical Interpretation

4.5.3.1 Individual Impact Scenario

Each cluster of term network, composed of overt and latent structures, is considered as an individual impact scenario. In this sub-stage, each scenario is interpreted in narrative storylines through analyzing the terms and their relations. The interpretation is performed in the same manner as any general story is developed. First, the general plot is discussed by categorizing the involved overt structures into *causes* and *effects*. Normally, impacts are associated with the *influencing factors* that causes such an event and *influenced agents* that are affected by such an event. By discriminating these two concepts, the storylines become more logical and coherent. Second, unexpected and disruptive details of each overt structure are further described using a variety of latent structures.

4.5.3.2 Overall Latent Impacts

Once the narratives of each impact scenario are extensively interpreted, we categorize them into specific types of effects. For instance, there exist a wide variety of forms of impacts: environmental, political, technological, social, economic, etc. This step is aimed at offering an overview of potential impacts by summarizing the various types of implications that could result from the advent of the target emerging technology. An experts' discussion is vital to gather and reconcile different viewpoints regarding the types of impacts.

4.6 Illustrative Case Study: 3D Printing Technology

As an illustrative case study, the article applies the proposed methodology in additive manufacturing technology, or also widely known as 3D printing technology. It is the process of joining materials in layers to make three-dimensional objects. The market has expanded drastically in recent years, with the participation of not only industries but independent creators, hobbyists and early adopters, and it is expected to continuously yield a great amount of technical economic and social effects (Arcos-Novillo & Guemes-Castorena, 2017). Despite its great progress, this just has been the tip of an iceberg of what is could further provide to our society. (Arcos-Novillo & Guemes-Castorena, 2017; Wohlers & Caffrey, 2013). There still exist a variety of uncertainties regarding its application, growth, and, implication. Additive manufacturing is a new and emerging technology that on the verge of revolutionizing our society, and a more systematic and ex-ante exploration is required to stimulate a better social acceptance.

The database is constructed through retrieving documents from future-oriented websites. The websites are selected based on the pertinence to the target technology and the involvement of both experts and the public. Numerous search queries were empirically investigated, and the queries of ("3d printing” or “3d printer” or “additive manufacturing” or “rapid prototyping”) AND (“issue”, “concern”, “worry”, “problem) were utilized for the data retrieval. Total 2,228 documents were collected, and total 16,520 terms were extracted after general preprocessing process. Prior to extracting overt structures, stop-word removal was deliberately conducted to remove as much noise as possible. Not only generic stop words with little informational content, including *of*, *it*, *of*, and *on*, were removed using various text-mining package, a separate session of experts’ discussion was additionally involved, by focusing on the previously mentioned criteria of frequency value and context pertinence. In result, the final input terms were reduced into 413. The term-document matrix was, thus, constructed in the form of 413-by-2,228 and further transformed into weighted matrix by applying *tf-idf*.

Table 4-2 List of overt structure-based scenarios

Number	Overt structure-based scenario	Top 15 highest loading terms
#1	Counterfeiting money	card, patient, counterfeit, power, credit, blueprint, illegal, digital, atm, money, copyright, economic, cash, currency, scanner
#2	Unhealthy air emission	environmental, environment, air, emission, pollution, scale, toxic, particle, green, consumption, carnage, chemicals, harmful, manufacture, citizen
#3	Printed food	food, eat, safe, kitchen, people, bacteria, business, health, policy, cash, government, safety, clean, cells, retail
#4	Taxation problem	tax, income, cash, legal, consumption, carnage, currency, retail, government, economic, legislation, harm, old, manufacturing, financial
#5	Recycling plastic	plastic, medical, biogradable, manufacturing, recycled, policy, economic, sector, traditional, security, environmental, metal, implication, health, policymaker
#6	IP problems	copyright, intellectual, property, rights, business, manufacturer, trademark, legal, industry, market, licensing, owner, policy, brand, economic

#7	Bioprinting issues	cells, tissue, organ, bioprinting, medicine, ethics, patient, medical, muscle, fda, pharmaceutical, biomedical, health, function, lung
#8	Energy hog	policy, heat, energy, clean, consumption, security, power, excessive, green, risk, loss, economic, reduce, environmental, ethical
#9	Printed firearms	gun, legal, firearms, weapon, rights, property, security, illegal, legislation, license, undetectable, act, policy, crime, ban
#10	Printed drugs	drug, fda, filament, pharmaceutical, legal, heated, medicine, patient, safe, chemical, blueprint, counterfeit, property, bacteria, compound

After the data preparation stage, LSA is performed. LSA was conducted via MATLAB software. The value of k was selected based on several trial-and-error processes, varied from 10 to 20. 10 factors offered a fair amount of meaningful topics, and the value over 15 generated incoherent and uninterpretable clusters. Considering the trade-off between quantity and quality, we have concluded by that 13 fully satisfied both of the features. A thorough discussion of five field experts was conducted regarding the reconciliation of diverse interpretations of general plots. In result, three topics were excluded to be considered as scenarios due to redundancy, and total 10 topics were considered as meaningful plots, as shown in Table 4-1. Relatively universal voices were extracted from the analysis, where the frequency values were all higher than 40. These resulting terms, composing each topic, are the ingredients for constructing general scenario plots. To better understand their relationship and interpret them to form the plots, *OR* was calculated on each term cluster. The thresholds of *OR* scores varied from 20 to 40 depending on the coherency of individual network structure. The network-based outputs of all ten overt structure-based scenarios are shown in Figure 4-4.

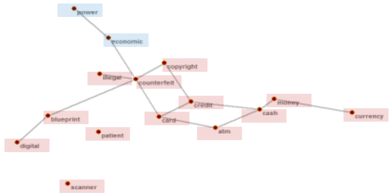


Figure 4.4(a) Scenario 1: Counterfeited money

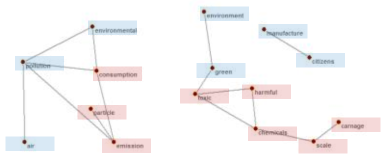


Figure 4.4(b) Scenario 2: Unhealthy air emission

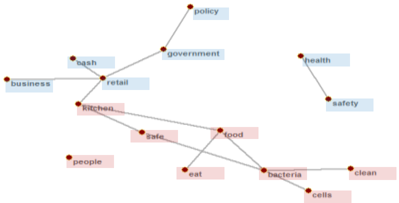


Figure 4.4(c) Scenario 3: Printed food

Figure 4.4(d) Scenario 4: Taxation problem

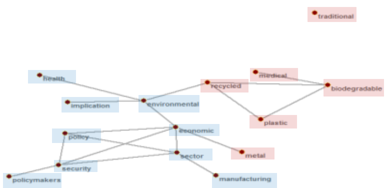


Figure 4.4(e) Scenario 5: Recycling plastic

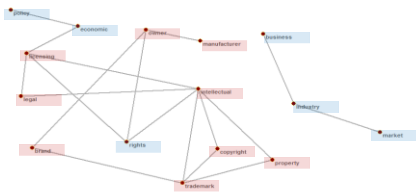


Figure 4.4(f) Scenario 6: IP problem

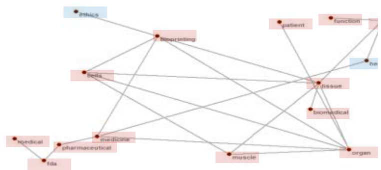


Figure 4.4(g) Scenario 7: Bioprinting issues

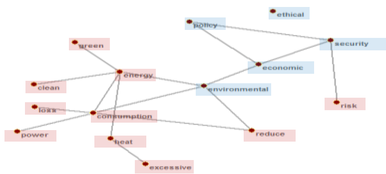


Figure 4.4(h) Scenario 8: Energy hog

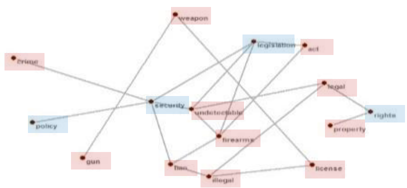


Figure 4.4(i) Scenario 9: Printed firearms

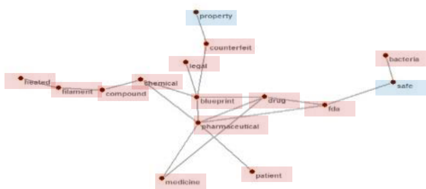
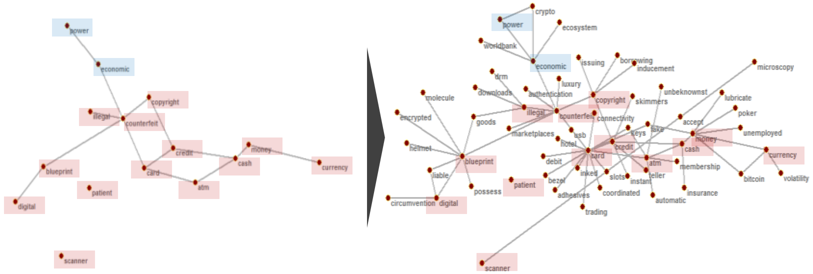


Figure 4.4(j) Scenario 10: Printed drugs

Figure 4-4 Graph-based output of overt structures-based scenarios

The succeeding step is supplementing hidden, latent details of the overt structures. *LR* was first calculated on the entire set of extracted terms to identify latent relations, and *LS* was further measured to filter out the significant latent structures. The threshold values of *LR* were fixed ranging from 500 to 900, and those of *LS* were varied from 10000 to 14000. The above threshold scores were determined based on the quantity and quality of the resulting terms. When the value was too high, it was difficult to spot any valuable latent structures; once the value was set too low, it included various noise-looking concepts, which hindered from a proper scenario interpretation. The most suitable threshold values were assigned based on a number of trial-and-error processes. Once the candidates for latent structures were obtained, a thorough discussion was done to decide top pairs for the inclusion in impact scenarios. A wide variety of latent structures were extracted to be included in impact scenarios, as listed in Appendix A, and they were further connected into the network structures. To avoid confusion and further develop more coherent and interpretable scenarios, only one or two links were generated per latent structure. The networks of all ten impact scenarios are illustrated in Figure 4-5.

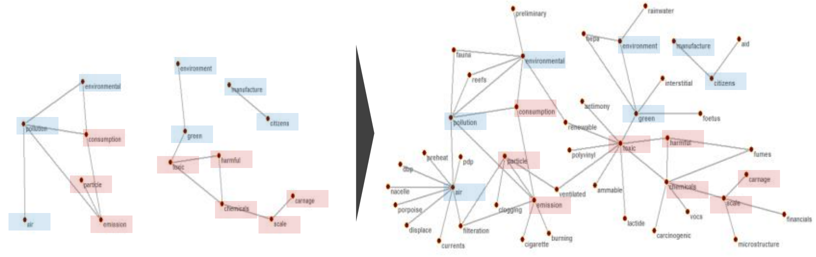
Economic Impact



Overt structure-based scenario　　　　Impact scenario

Figure 4-5(a) Scenario 1: Counterfeited

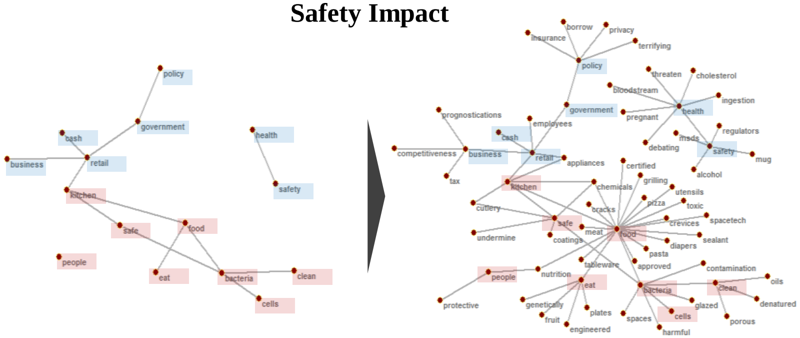
Environmental Impact



Overt structure-based scenario　　　　Impact scenario

Figure 4-5(b) Scenario 2: Unhealthy air emission

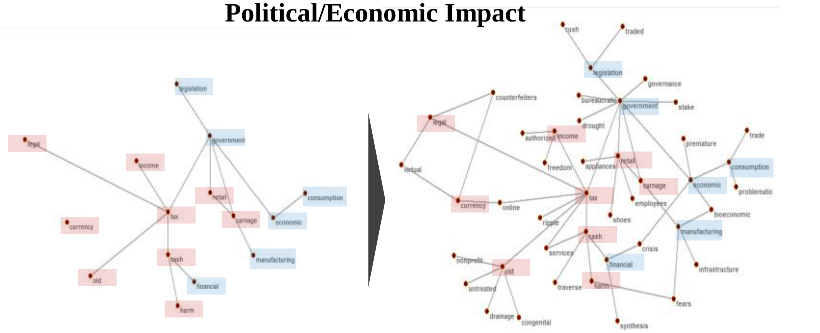
Safety Impact



Overt structure-based scenario　　　　Impact scenario

Figure 4-5(c) Scenario 3: Printed food

Political/Economic Impact



Overt structure-based scenario　　　　Impact scenario

Figure 4-5(d) Scenario 4: Taxation problem

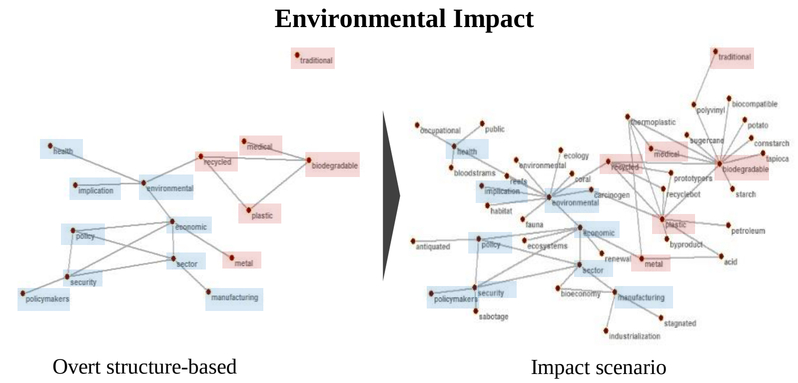


Figure 4-5(e) Scenario 5: Recycling

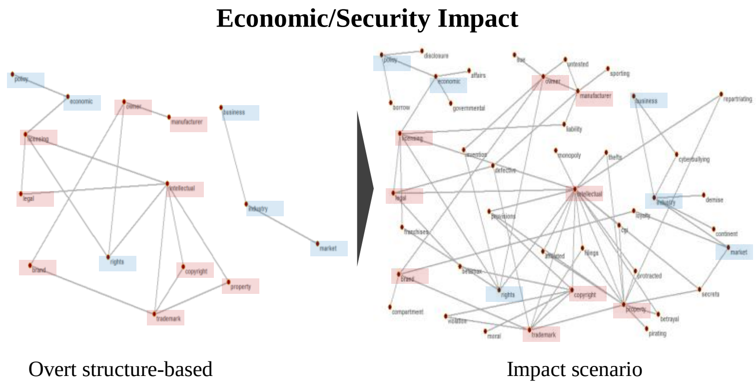


Figure 4-5(f) Scenario 6: IP

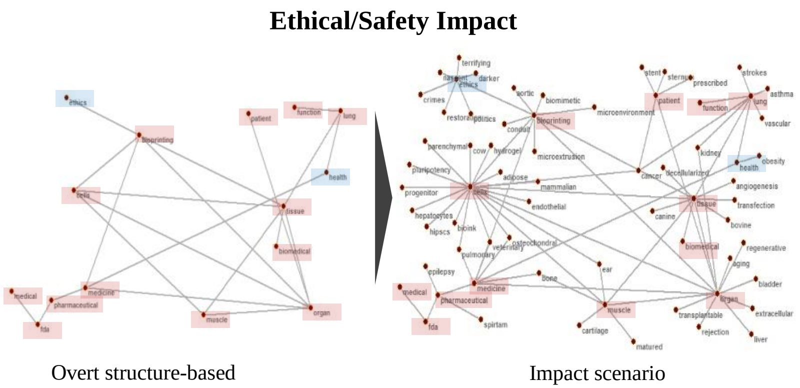


Figure 4-5(g) Scenario 7: Bioprinting

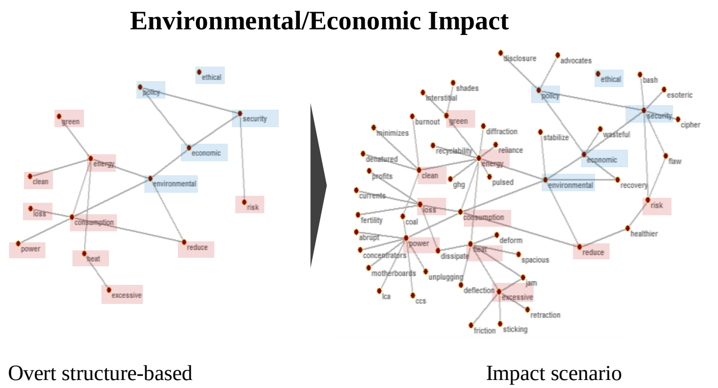
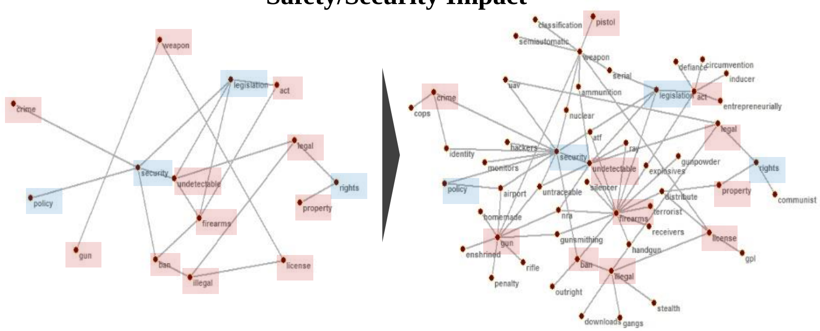


Figure 4-5(h) Scenario 8: Energy

Safety/Security Impact

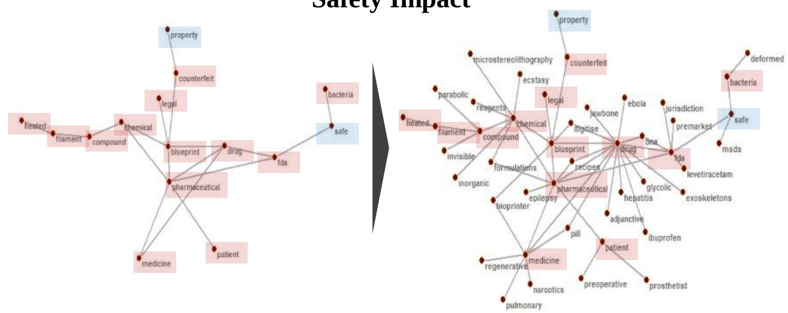


Overt structure-based scenario

Impact scenario

Figure 4-5(i) Scenario 9: Printed

Safety Impact



Overt structure-based scenario

Impact scenario

Figure 4-5(j) Scenario 10: Printed

Figure 4-5 Impact scenario networks using LSA and IdeaGraph

Finally, each individual scenario was interpreted in narrative storylines by analyzing the constituting terms and their relations. A focus group discussion was conducted to classify overt structures into *causes* and *effects*: the ones responsible of such a state and the ones directly or indirectly affected by the state, respectively. As shown in Figure 4-4 and 4-5, *causes* and *effects* are denoted as red and blue box in the resulting keyword networks. Then, latent structures were interpreted based on the linked overt structures. The categorized terms for each impact scenario and their corresponding latent structures are listed in Appendix A. Through such a series of steps, the overt structure-based scenarios have been supplemented with latent structures to form a set of individual impact scenarios. All ten impact scenarios are thoroughly validated based on the associated documents for each topic, and they are utilized to offer an overview of potential impacts caused from additive manufacturing technology. As shown in Figure 4-5, such an emerging technology could result economic, environmental, safety, political, ethical, and security issues.

4.7 Discussion

4.7.1 Scenarios Beyond the Obvious

Since the process of analytic interpretation may seem difficult to follow, this section attempts to demonstrate how overt and latent structures are combined to generate impact scenarios with much of plausibility and creativity. Given the space constraint, only two impact scenarios of Printed Foods (Scenario 3) and Bioprinting (Scenario 7) were chosen for an in-depth illustration.

Scenario 3. Printed Foods

The first step in the analytical interpretation involves the general understanding of a plot. An experts' discussion was conducted to observe the constituting overt structures and their relationships with one another, and it was found that the scenario plot was about the risks associated with 3d printed foods. The overt structures were assorted into *causes* and *effects* according to the logic and plausibility of a storyline. The causes-related terms include "safe", "food", "kitchen", "people", "eat", "bacteria", "cells", and "clean"; the effects-related terms include "government", "policy", "retail", "cash", "business", "health", and "safety". The *causes* can be interpreted as "safety concerns with bacterial growth within printed foods", and the *effects* can be understood as "governments designing new business policies regarding retail industry and regulation system for public health and safety". Each and every word comprising the storyline is a representative fragment of the majority opinion, and the proposed approach attempts to integrate such a disparate information into a whole, coherent storyline. As seen, the overt structure-based scenario seems quite apparent and holds generalized information with a much of plausibility.

The impact scenario becomes more detailed and fruitful when latent structures are further supplemented. In details, there exist two types of latent structures: general details and unexpected elements. First, overt structure of "food" was relate to latent structures of "meat", "pasta", "grilling", "pizza", "tableware", and "sealant". These low frequency concepts yield very specific but quite general details, and their presence indicates that people are worried about certain conditions and types of printed food and kitchenware. Second, overt structure of "safe" was related to latent structures of "undermine", “coatings”, “chemicals”, “crevices”, and “cutlery”; “bacteria” was associated with “cracks”, “spaces”, “porous”, and “contamination”. Such relations of concepts indicate that safety-related issues are fundamentally resulting from micro porous structures, which could develop bacterial growth. Moreover, overt structure of “business” was closely related to “prognostications”, and this suggests that the future events must be envisioned for various companies to set proper business strategies. These low frequency terms offer quite unexpected details, which are only mentioned by a handful number of people in the web. In addition, it is challenging to distinguish the valuable latent structures among the many. The context must be thoroughly understood by analyzing the terms and their relations in the networks or examining the *LR* or *LS* values. The associated documents could also be used to validate the significance.

Scenario 3. Bioprinting

The plot of the scenario was about the risks regarding bioprinting. The causes-related terms include “cells”, “tissue”, “biomedical”, “medicine”, “muscle”, “organ”, “lung”, etc. The *causes* can be interpreted as “challenges of printed organs and medicines used in biomedical and pharmaceutical fields”. On the other hand, the effects-related terms include “ethics”, “patient”, and “health”. This refers to the interpretation of “there exist issues not only of health but of ethics in result, and patients are the major stakeholder in such an issue”. Such analysis has also enabled the investigation of both general and surprising details of impact scenarios. In terms of general details, much of recent trends and scientific findings were identified in abundance. All specific types and conditions of “tissue”, “organs”, “cells”, “medicine”, “lungs”, and “muscles” were found to be in people’s minds. The states and types of “organs”, for instance", include "bladder", "liver", "kidney", etc.; the types of "cells" include "ear", "hydrogel", "parenchymal", "adipose", "endothelial", etc. Other than "organs" and "cells", most of overt structures included such an information. When pertinent articles were analyzed, these structures were mostly the latest technology-related news about scientific findings, which may seem significant but possible to cause future problems.

Furthermore, there were several latent structures that indicated quite an unexpected potential results. For instance, the latent terms like cows", "mammalian", "bovine", and "canine" implied that people were disturbed with utilizing animal cells as an input for printing replacement organs and joints. The latent concepts of "stent", "sternum" and "cancer" from overt structure of "patient" suggested some people are doubtful about the reports that bioprinting-based surgeries could actually better treat patients suffering from stents, sternum or cancer. In addition, the latent structures of "spirtam" and "epilepsy" from "pharmaceutical" implied the past problematic situation where pharmaceutical industry was greatly affected by 3D printing technology.

4.7.2 Comparative Analysis

In order to enhance internal validity and generalizability of the research, the result was compared with that of a previous study. A discussion was conducted to select relevant studies with the same research objective. The research article by Huang et al. (2013) was chosen for the comparative study. The study investigates societal impacts of additive manufacturing technology through qualitatively reviewing scientific articles from diverse fields. The societal impacts are as follows. First, energy consumption and environmental impact was suggested to a serious issue since manufacturing processes of 3d printing required extravagant use of materials and energy resources. Moreover, the input materials had poor biodegradability, and the resulting environmental impact was a subject of great concern. Second, the study suggested various health and occupational hazards, caused from air/water emission, noise, and fluid spills. Possible diseases include dermatitis, asthma, lung fibrosis. As seen, the previous study has offered potential risks with an in-depth technical details, including the types of chemical, solvent, and emissions. This indicates that such an approach may generate much diverse and more latent structures. However, the derived terms seemed too scholarly, and we assert that the results represent only the perspectives of experts. This is mainly due to the sole reliance on scientific publication data or technical reports, which are generally dealt in the fields of environmental impact assessment and life-cycle analysis. They incorporate an abundance of numerical data regarding toxicology and environmental science.

Conventional research suggests very different implications compared to our proposed approach. The fundamental distinction is the interpretation and the focus of social impact. The notion of social impact is quite vague because it may include too wide variety of aspects, such as political, economic, environmental, and moral consequences. The previous studies tend to focus on the aspects with the availability of actual database, and this has caused a serious negligence of societal ramifications. To this end, our proposed methodology is more focused on societal issues that have been discussed not only by experts from diverse fields but also by the general public with high interest in new technologies. The approach has yielded a relatively small number of technical terms, but it rather offered numerous societal details, like “safety”, “cutlery”, “terrifying”, “meat”, and “obesity”. Some may seem too broad, familiar, and sometimes absurd; however, these resulting terms and scenarios are, in fact, what we have sought to explore from the beginning of the research. We were curious about what most general people have in their minds when they envision the future society affected by new technologies. This is indeed considered one of the study's distinctive strength. Furthermore, the data-driven technology foresight approach has provided a visualized output, which may stimulate a better understanding of both overall impacts and specific latent structures.

4.8 Conclusion

This research proposes a *data-driven scenario building process* as an alternative or supplementary way of building well-balanced storylines equipped with a much of plausibility but at the same time a fair amount of unexpectedness. It serves its purpose on envisioning unforeseen societal ramifications of emerging technologies. A wide range of futuristic perspectives was acquired based on future-oriented web documents, and their summaries were extracting using LSA and IdeaGraph text mining technique. In details, impact scenarios were built by following two consecutive steps: generating main plot by grouping high frequency terms and further supplementing latent and unexpected details by adding low frequency terms closely related to the general plots.

The proposed data-driven approach is a preliminary attempt of performing scenario building process in a quantitative manner. An immediate improvement may seem small in scenario planning community; however, our result may open the possibility of incorporating futuristic concern-related voices from the web and applying text mining techniques into exploring the holistic overview of emerging technology's future uncertainties. In practical point of view, this methodology could be used as a vital tool for understanding a wide variety of uncertain ramifications of emerging technologies, thereby promoting a more responsive management. However, the most major limitation of this study is that the results cannot be properly validated. Since we are dealing with the results that have not yet happened to our society, there is no appropriate way of proving our derived result is accurate and correct. To move forward, a more systematic process of validating the quality of resulting scenarios must be developed and further applied.

Chapter 5

Foresight for Plan Development

The third step of proactive management process is to develop detailed plans for preventing future issues associated with emerging technologies. This research focuses particularly on deriving technical guidelines for a more generalized management process. A proper planning is required for a new technology to be technically prepared for difficulties that will inevitably arise once new technologies proliferate. Such an effort has been dealt in the field of responsible development, which is a systematic process of exploring human values at stake when an emerging technology proliferates and suggesting design solutions that could embody pertinent values. Major drawbacks of previous studies are that no clear-cut methodology has been developed to perform responsible development and that there has been a lack of comprehensiveness and diversity in developing future issues and design solutions.

This chapter suggests a data-driven problem solving approach in closing a gap between philosophizing and existing when it comes to responsible development. As conducted in Chapter 3, LSA technique is conducted in *future-oriented web data* to extract multiple hazardous eventualities of emerging technologies. The similar process but only with scientific literature data is proceeded to derive multiple solution concepts. The proposed approach is an alternative and detailed guidance of developing a new technology in a responsible manner, promoting the right impacts to our society.

5.1 Introduction

Considering the great history of technology, no person can deny that new and emerging technologies have resolved numerous challenges of humanity. Meanwhile some minds think in a different point of view; they address the destructive nature of those technologies, including global warming or privacy and security issues. These unprecedented consequences are reconstructing our existing societal, economical, and environmental frames and transforming fundamental conditions of our lives and moral assumptions (Marx, 2010). As such, technology is holding a paradoxical nature: a novel solution for societal challenges but, at the same time, a critical agent that could cause far-reaching chaos. In response to these ramifications, we say the notion of *responsibility* must be incorporated at the earliest technology development, or engineering design, of technology's innovation process. The *responsibility* points out our neglect of technology's unpredictable consequences on individuals, societies, and environments in favor of immediate commercial success and economic growth (Grunwald, 2011; Blok & Lemmens, 2015). In the field of responsible innovation (RI), scholars have asserted that technology must be shaped or designed according to social values, in the hope that unintended and often undesirable side effects would never occur in the first place. However, one criticism of much of the literature on responsible development is that there exists a tendency of separating thoughts from actions (Firat et al., 2008; Grunwald & Achternbosch, 2013). The field is still at its embryonic stage, and the research to date has focused on the vigorous clarification of the notion of responsible innovation rather than the establishment of specific guidelines that could actually be utilized in real life.

The notion of acting on responsibility in technology development has spurred few initiatives, such as Value Sensitive Design (VSD) and Value Conscious Design (Flanagan & Nissenbaum, 2014). By gathering and further analyzing information from field experts through participatory approaches, they have offered systematic problem-solving processes and have led to a renewed interest in the fields of technology assessment and technology foresight. However, there still exist several drawbacks of the existing participatory approaches: inefficiency and ineffectiveness. It is labor intensive and time-consuming to gather experts from a wide ranging fields, and the results are usually derived in a predictable manner with a very limited amount. The proposed methodology incorporates two types of data sources: *future-oriented web data* and literature data. Generated by Web 2.0, future-oriented information encapsulates the opinions of futures experts and the public interested in future technology and its societal change. We use Latent Semantic Analysis (LSA) text mining technique to extract issue-related concepts based on terms' semantic relations. Subsequently, the technique is applied to scientific publication data to capture novel solutions from the visions of eminent scholars from a range of disciplines.

The proposed method envisions a set of eventualities where a newborn technology plays out to be a harmful tool to the society and, in response, suggest essential technical requirements that must be focused in the design process for the right technological preparedness. Once such a process is well achieved, numerous problematic impacts will be prevented ex-ante and thereby promoting only the *right impact* (von Schomborg, 2013; Owen et al., 2012), or the right effect, to the society while protecting certain social values from degradation from the use of emerging technologies. The increase in digitalization is promoting the need for technology managers' skills for data analytics, specifically focusing on commercial, environmental, ethical and societal implications (Cetindamar et al., 2016). The proposed framework is expected to offer responsive and responsible insights to designers and engineers of emerging technologies.

5.2 Theoretical Paradigm Shift

Quite different from established technologies, emerging technologies are completely altering our existing social norms and values. Their distinctive characteristics have gradually affecting and thus changing how people perceive the consequential nature of technologies and how people react and resolve such situations.

5.2.1 Technology-focused vs. Society-focused

Fundamentally speaking, what sorts of impacts could be derived from the advent of emerging technologies and is it even possible to have a control over their development? First, new technologies could result unwanted consequences in two different ways: (1) an impact from technology's inherent features, and (2) an impact dependent upon the social context. In other words, emerging technologies could either naturally become one due to its technical or physical nature or be misleadingly forced to be a harmful tool by a wrongful user. In this context, it is worth noting two opposing philosophical views of technologies. Commonly, the former is expressed more formally as the notion of *technology determinism* (Keller, 2008) and the latter as the notions of *technology instrumentalism*, *the social constructivism,* or *soft determinism* (Surry &

Farquhar, 1997; Grunwald & Achternbosch, 2013; Smith & Marx, 1994).

Notably renowned by Marx, McLuhan, and Toffler, *technology determinism* puts philosophical premise on technology's inevitable and autonomous nature. It was believed that technology drives its own applications and thereby resulting either a social prosperity or a destruction (Smith & Marx, 1994). This was based on technology's predefined implication, which was considered beyond direct human control (Surry & Farquhar, 1997). Since many assumed that the overwhelming power to social change is imputed to technology itself, or to some of its intrinsic attributes, the principle of designable and controllable technology was completely ignored. In result, the only solution was to expect the society to react a certain way and to prepare itself for coming technologies and their unseen impacts (Grunwald & Achternbosch, 2013). Such a notion has also altered the minds of the producers of technologies. Without much of conscious awareness on the consequential effects, engineers and developers focused mainly on augmenting instrumental values, including the advancement in technological functionality and efficiency. This was particularly prevalent due to the intrinsic features of past technologies. Past waves of emerging technologies were characterized by a definite and single purpose. By way of illustration, a mp3 player was developed to serve as a portable music device and a cellular phone as a wireless communication device.

However, as newly introduced technologies increasingly getting equipped with unprecedented features and functions, the range of their uses got expanded and thereby yielded more diversified and unique influences on existing social practices. This have resulted a deep consideration of the interaction between technology and social context. Opposed to *technology determinism*, the notions of *technology instrumentalism, the social constructivism,* or *soft determinism* were brought into the dispute. All three grand narratives share a common principle: the role of social conditions in technology's future implication and human aspiration in shaping technology. *Technology instrumentalism* assumes that the technology is just a tool, and their implications are depended upon the intentions or existing conditions of the society employing the tool (Surry & Farquhar, 1997). Extended to this, *the social constructivism* perceives the technology as the output of societal processes (Grunwald & Achternbosch, 2013) and highlights the value of "shaping technology". Furthermore, *soft determinism* addresses that the origin of power of new technologies is associated in a far more various and complex social, economic, and cultural conditions (Smith & Marx, 1994; Keller, 2008). Based on these notions, social context became a vital factor in technology development. Social sciences research associated with the perspectives of end users, such as human, social and interpersonal factors, were involved in the development process. Responsibility was completely shifted to the society, and people highlighted the human capability of controlling technology's disastrous consequences.

5.2.2 Co-evolution of Technology and Society

Such dichotomy has been dominant in the past; however, we instead argue for a safer position: co-evolution of technology and society. Technology and society are closely inter-related each other and mutually influences one another (Grunwald & Achterbosch, 2013). Societal functions and habits change as new and emerging technologies are being introduced and those technologies are further developed in accordance with the altered societal perceptions and norms (Carlsen et al., 2014). As shown in Figure 5-1, the logic of the co-evolution is the integration of two contradicting paradigms, considering each viewpoint is as valid as any other. We specifically offer the details of the existing paradigms and the proposed paradigm in Table 5-1. As listed, the impacts associated with emerging technologies could result either directly by their inherent technical features or indirectly from their social conditions. The former can be seen as unintentional and inevitable consequences, which are beyond human control. The latter, on the other hand, can be considered as intentional and avoidable implications, which are largely under human control. By way of illustration, unmanned aerial vehicle (UAV) technology, also widely accepted as drone technology, is expected to offer various recreational contents and may greatly impact the existing recreation industry. The economic implication caused from drone technology's intrinsic function of recreation service is hardly prevented or solved ex-ante. However, the impacts resulting from the use in recreational sectors, such as safety issues, could be avoided either socially or technically. Certain regulations, for example, can be set to limit the range of hovering; drone's microprocessor could be pre-programmed to fly above and below certain aerial ranges. These resolutions may prevent from threatening the safety associated with pedestrians and aircrafts.

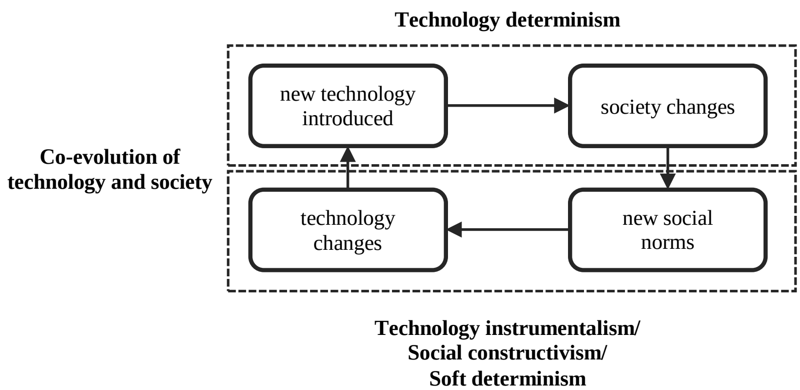


Figure 5-1 Basic logic of co-evolution of technology and society

Co-evolution of technology and society has always been an important theme in science and technology studies; however, the existing accounts have criticized the complexity and uncertainty concerning the management of societal consequences based on Collingridge Dilemma (Collingridge, 1980). The whole problem-solving process was considered a tough, almost impractical, task (Grunwald & Achternbosch, 2013). However, this paper views quite positively since the actions are actually achievable. There exist a wide variety of untouched data sources discussing the issues of emerging technologies and technical solutions that could possibly prevent the issues from occurring. By incorporating a text mining technique in these sources, we could realize the paths from "new technology introduced" to "technology changes" as in depicted Figure 5-1. Assuming a certain emerging technology is emerged, we explore how society could be changed; what sorts of social norms are affected by the transformation; and how the technology must be changed considering these altered contexts.

Furthermore, we specifically focus only on the technical solutions, even though there are different ways of coping social issues, such as maintenance and regulatory directions. There are two primary reasons for such an emphasis. First, science- and engineering-related direction can be universally accepted and effectively utilized without strict constraints. On the other hand, non-technical direction must encompass geographical and infrastructural boundaries in order to be effective. Therefore, the suggestion of technical tips is most suitable for many stakeholders to easily act upon over such unforeseen issues. Second, the merits of cost reduction may be achieved with technical solutions since no technology could evade from potential social environment expenditures. If identification and technical management is achieved ex-ante in the earlier development process, where greater degree of freedom exist in technical modification, future cost could be comparatively decreased.

Table 5-1 Characteristics of existing and the proposed paradigm

	Existing Paradigms		Proposed Paradigm
	Technology Determinism	**Technology Instrumentalism, Social Constructivism, Soft Determinism**	**Co-evolution of Technology and Society**
Source of Impact	technology itself	societal condition	both
Intension of Impact	unintentional	intentional	both
Responsibility	technology	society, users	both
Solution	societal change	shaping of technology	both
Control of Technology Development	beyond human control	under human control	both

5.2.3 Responsible Development

No technology ever has performed an action that are not pre-programmed by human beings (Smith & Marx, 1994). It is just a human product, an output of conscious human choices and activities (van de Poel, 2015). With a right engineering design decision makings, the development of technology and its impact to our lives could be controlled. The notions of responsible development or responsible innovation is worth noted in this context. Based on the definition proposed by von Schomberg (2011), responsible innovation is a transparent, interactive process of sharing the view of the ethical acceptability, sustainability, and societal desirability for the innovation process and its marketable products.

When viewed in a broader sense, the responsible innovation can also be defined as the process of taking care of the future through collective stewardship of science and innovation in the present (Stilgoe et al., 2013). VSD is the most widely used methodology in such fields. Based on the premise of technology being constructed according to social values (Grunwald & Achternbosch, 2013), VSD stresses the notion of people's experiences and contexts and identifies human values that are possibly at stake when new technology proliferates. It then systematically explores different technical designs and requirement that could embody pertinent values. Through following a systematic procedure, social and moral values can be placed at the core of new technology design, thereby promoting a better public acceptance of unfamiliar technologies. Specifically, VSD follows three main steps: conceptual investigation, empirical investigation, and technical design (Cummings, 2006). In conceptual investigation, a discussion is conducted to identify basic human values that could be either supported or diminished by technology emergence. In empirical investigation, quantitative and qualitative exploration is applied to examine specific eventualities of human-technology. In the final step, technical directions that could best embed those values and norms are suggested.

Even though it was first mainly utilized in the domain of information systems and software (Millet et al., 2001; Friedman et al., 2002; Miller et al., 2007), recent studies are proposing its possibility of a more general applicable fields, including the field of emerging and new technology development. VSD's systematic procedure effectively deals with disruptive, complex, integrative properties of technology advancement. For instance, Simons and Verhagen (2008) used VSD in order to support design and policy decisions in healthcare Information and Communication Technology (ICT) innovations. Timmermans et al. (2011) suggested specific features associated with nanopharmacy and promoted stakeholders to act on responsibility. Moreover, Oosterlaken (2015) highlighted various values in developing wind-technologies in order to augment public acceptability. Other applications include medical devices, home technologies, and web browsers (Denning et al., 2013; Denning et al., 2014; Xu et al., 2012).

VSD aims for similar objective with the fields of technology assessment and social impact assessment: deep concerns for unprecedented impacts of new science and technology that could create or destruct human values. Both underscore the interdisciplinary and democratic approaches and assert the integration of various opinions regarding unintended side effects and pertinent solutions (Grunwald, 2011). However, the factors that VSD makes itself special is that VSD suggests detailed technical guidelines in the hope that they would no longer occur in the future, and they are performed in a relatively more systematic way (von de Poel, 2015). As such, this paper attempts to suggest technical directions as well for solving pertinent societal problems, but rather done in a more quantitative manner.

5.3 Methodological Paradigm Shift

5.3.1 Participatory Approach

New and emerging technology is characterized by its inherently distinctive natures: uncertainty and complexity (Maine et al., 2014). Adoption rate at the early stage of technology's lifecycle or promised return from the investment are unclear (Halaweh, 2013). In addition, it is challenging to fully understand the cause and effect chains and network effects among various actors associated with new technologies (Köhler & Som, 2014). Such social uncertainties are considered as one of the key factors to unpredictable changes and the key barriers to innovation success (Hueske et al., 2015). To assess such uncertain and complex future, the studies in the field of technology foresight and TA highlight the power of wisdom of crowds and collective intelligence (Gray & Hovav, 2008; Firat et al., 2008). An integration of various opinions from various players, not just from experts but also from general public, is recognized for gaining greater knowledge about the future that has not yet occurred and thereby promoting a more concrete and proactive decision making. This notion is also well-applied in the fields of TA and VSD through gathering opinions from involved players based on qualitative techniques, such as interviews, workshops, and surveys. The field of responsible innovation also emphasizes the collective thoughts from various lead users in order to mitigate plausible societal risks and ethical issues associated with new technologies (Brem & Bilgram, 2015). One critical question that needs to be highlighted is, however, whether such participatory approach is the best and only way to gather the voices regarding the future of emerging technology.

Far too little attention has been paid to the disadvantages of participatory approaches. First, the entire process is quite labor intensive and time consuming. It is excessively demanding when gathering prominent experts professionalized in the target technology. Since the purpose of VSD is to explore the inter-relationship of technology and society by extracting potential issues and possible solutions, it becomes even more challenging to incorporate a more extensive pool of stakeholders with diverse disciplines. Furthermore, the tasks of collecting desired information and reconciling different opinions normally require weeks or even months. Second, participatory approach, especially in

VSD, seems ineffective in quantity and quality perspective. Primarily due to a small pool of involved participants in the foresight discussion, the results of future issues and potential solutions are not presented only in obvious and predictable manner but also with a very limited amount.

5.3.2 Data-driven Approach

To date such a participatory approach was quite inevitable. Not much is known about emerging technology, especially in a societal perspective; and therefore, not much of information is available until they are widely adopted and used. The most fundamental problem that has been remained unsolved was a dearth of reliable and pertinent data for envisioning and exploring the potential issues and solutions. Thus, the need for an alternative method to capture a more extensive range of collective intelligence must be heightened.

Previous studies have stressed the importance of interactive and collective intelligence that enables the observation and governance of innovative technology's future impacts. This paper, in response, stresses the latent potential of *future-oriented web data* and literature data, as a silver lining. We are currently interested in gathering reliable and suitable knowledge sources for two major processes: problem formulation and solution derivation. These two data sources were selected based on the philosophy stated by Stilgoe et al. (2013). The study explores conceptual and policy background of various approaches and suggests four primary dimensions for responsible innovation: *anticipation, reflexivity, inclusion, and responsiveness*. Figure 5-2 delineates the rationale for using these data sources. *Future-oriented web data* satisfies the notion of *anticipation*, in that it holds a wide variety insights regarding technological and societal future. Scientific literature data satisfies the notion of *responsiveness*, in that it provides a capacity to change technological shape and direction in response to public values. These data sources exhibit both *reflexivity* and *inclusion* since they include a large collection of new voices, capable of prompting reflections of new technologies on the socio-ethical context.

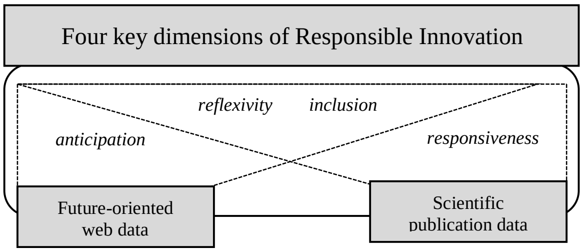


Figure 5-2 Illustration of data sources capturing key dimensions

For deriving future problems, we address the use of *future-oriented web data*. There exist numerous online platforms that share future-oriented perspectives about new and emerging technologies and socio-technical circumstances, which have not yet happened or experienced. By extensively overlooking the concerns of the stakeholders with tacit knowledge of new technologies and expertise in futures studies, we envision the problem-arising circumstances that must be corrected or avoided. For identifying scientific and technical solutions, we analyze literature data and attempt to search for "nuggets" of key reviews, models, or developments that could be served as the solutions for potential issues. Scientific publications provide a wide variety of professionalized knowledge regarding scientific, technical, societal, or political domains. On single topic of interest, various opinions are suggested by scholars from many different disciplinary fields. No one can tell which opinion is superior over the other; therefore, it is necessary to extensively monitor the concepts that are primarily discussed and pointed out by those intellectuals. By perceiving them as target problems and corresponding solutions, implicit knowledge can be inferred from seemingly unrelated concepts since they are seldom contextually linked and form a complementary relationship with one another.

5.4 Rationale for using LSA

How could we then manage to better understand a massive collection of texts and point out only the key concepts or topics? Various methodologies can be accommodated in response, including Latent Semantic Analysis (LSA), Probabilistic Latent Semantic Analysis (pLSA), Latent Dirichlet Allocation (LDA), and Principal Components Analysis (PCA) (De Melo & Siersdorfer, 2007; Lee et al., 2010; Newman et al., 2014). These approaches measure the co-occurrence relationships among the terms and thus generate multiple topics, which describe the main contents of the documents in the database. In our work, we chose to incorporate LSA. Latent Semantic Analysis (LSA) identifies core concepts implicitly presented in a set of documents through dimension reduction of term-document matrix (De Melo & Siersdorfer, 2007). Assuming that each of brilliant thought is expressed in forms of raw texts, we extract underlying topical structure of document corpus through analyzing the relations of contextual usage of terms (Landauer, 1998). Originated in the late 1980s for the purpose of information retrieval and search engine query, LSA has been widely applicable to extensive analytical scope, including essay grading, text categorization, and, most importantly, topic detection (Sidorova et al., 2008; Blake & Ayyagari, 2012; Kwon et al., 2017).

Considering the contexts of *future-oriented web data* and literature data, LSA seems most suitable based on two reasons. First, truncated SVD captures not only significant underlying structure in the association of terms and documents by removing variability in word usage, but at the same time captures latent structures even though the words never co-occur in the same document (Rosario, 2000). The concepts that are difficult to recognize based qualitative investigation of data will be derived in result. Second, LSA offers distinguishing ability to extract not only the relations between the terms and derived concepts but also relations between the documents and the concepts. Since the terms are grouped based on co-occurrence patterns, and they provide details for the interpretation of key concepts. Moreover, grouped documents can be utilized for further interpretation and validation. The high loading literature documents may provide most appropriate and detailed location for solving pertinent problem. In addition, the term clusters could be validated using those relevant documents.

5.5 Research Framework

The overview of research framework is illustrated in Figure 5-3. The first step is envisioning social problems, and the second step is deriving technical solutions and identifying pertinent social values for each social problem. Each step is performed based on LSA text mining technique. The following sections will provide detailed explanations of the sub-steps.

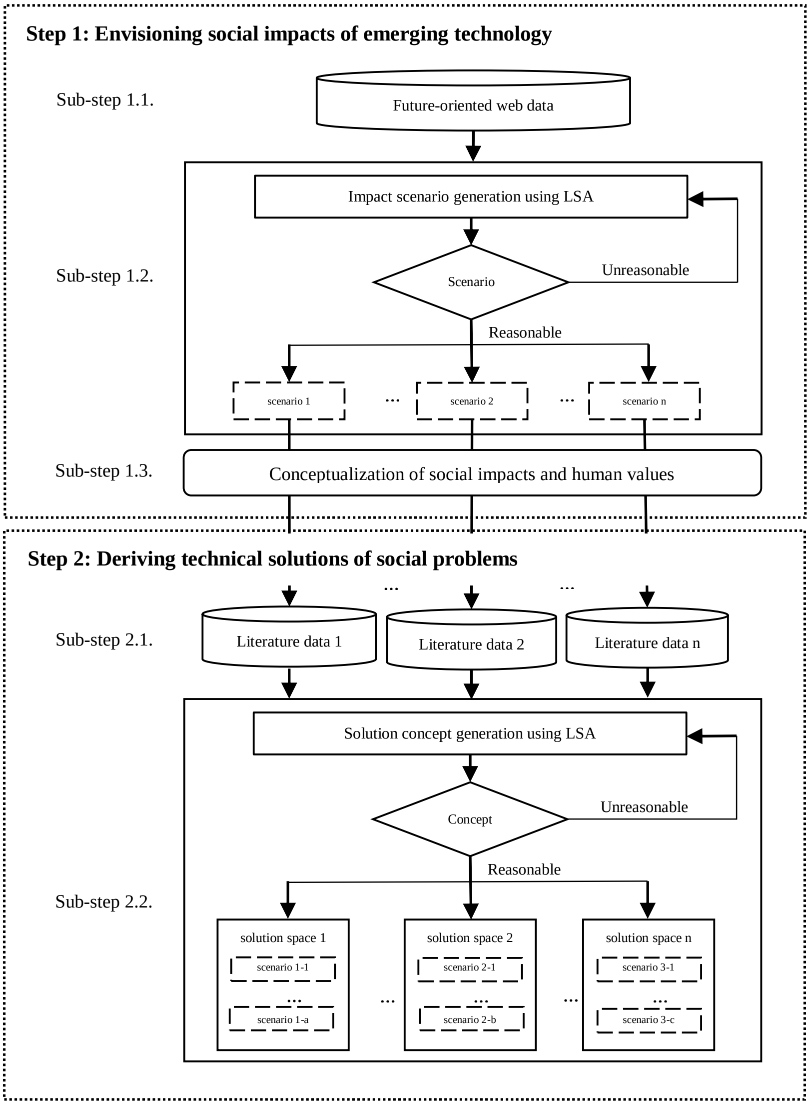


Figure 5-3 Illustration of overall framework

5.5.1 Step 1. Envisioning Social Issues

5.5.1.1 Collection of Future-oriented Web Data

The first sub-step of step 1 is the retrieval of *future-oriented web data*. To obtain proper sets of future-oriented information, the empirical investigation and assessment of technology websites are thoroughly conducted. In details, future-oriented websites can be differentiated from regular technology websites by three features: viewpoint, technological focus, and involved players. First, future-oriented online communities are in varied forms, ranging from web magazines and news articles to public forums, and each contains its own distinctive point of views. Web magazines normally cover both subjective and objective opinions regarding future technology; news articles are mostly dealing with objective information about recent radical discoveries or innovative development; and public forums may hold all sorts of futuristic opinions, mostly constituted of subjective voices. Second, each website has its own technological focus of the contents. Some websites may primary focus on certain future technology, like drones, while some may focus only on future Information and Communication Technology (ICT). Lastly, involved players in those platforms are quite diverse. For instance, some of the platforms are mostly composed of voices of only futures experts; on the other hand, some may include only the small portion of expert's opinion, just to propose certain topic for the discussion, and hold mostly thoughts of general public. Depending on the purpose of the research, the involved websites are carefully selected and the data are extensively gathered to build one whole concrete database. For this particular research, the voices of both experts and general public are extensively gathered in order to acquire not only reliable and expertized descriptions but creative and unexpected future scenarios. To retrieve concerns and worries regarding the impact of emerging technologies, the process of selecting search queries, as proposed in section 3.4.1, is utilized.

5.5.1.2 Construction of Impact Scenarios

For extracting scenarios, LSA text mining technique is applied on the acquired database. The analysis generates multiple concepts or topics through mining meaningful semantic patterns existing in massive amount of opinions (Blake & Ayyagari, 2012). The results are represented by a compound of key terms, and they are basically the summarized opinions illustrating specific eventualities where new technology may harm conventional standards or human values. In methodological perspective, this process is a LSA-based topic detection method, as successfully realized by Sidorova et al. (2008) and Evangelopoulos et al. (2012). On the basis of previous attempts, this paper's scenario building process is divided into three parts: preprocessing (pre-LSA), LSA, and interpretation (post-LSA).

Preprocessing part (pre-LSA) is of utmost importance for almost every text mining process since the overall performance will be heavily depended on how well the preprocessing process filters out the unnecessary ones while highlighting the significant ones. Therefore, all the crawled documents must be first transformed into structured and computer-understandable format. Basic preprocessing techniques, including tokenization, part-of-speech tagging, stemming, and stop-word removal, are performed. Stop-words, those terms with out-of-context or low occurrence frequency, are carefully examined and excluded for the further analysis. The text data is then converted into vector space model; in other words, database is represented as m-by-n term document matrix, where each matrix cell entry includes the frequency of term appearing in a document. The last phase of preprocessing stage is about term vector weighting. Since each term hold its distinctive amount of information, the term's relative importance is measured based on the term frequency and document frequency. Thus, term's raw frequency values are transformed into weighted vectors based on the product of term frequency and inverse document frequency, also known as *tf-idf*. A logarithmic transformation of original term frequency and inverse document frequency is applied in order to put more importance on the terms that distinctively characterize the content of the document (Kwon et al., 2017). The final weight of term i in future-oriented document j is, thus, represented by

$$w_{ij} = tf_{ij} idf_i \quad [5\text{-}1]$$

$$w_{ij} = (1 + \log_2 n_{ij})(\log_2 \frac{c}{df_i}) \quad [5\text{-}2]$$

, where n_{ij} is the term frequency of term i in document j and df_i is the document frequency term i is included. The second part is where terms are clustered with one another to finally form the future scenarios. The weighted term-document matrix $(X_{t\times d})$ is decomposed into three different matrices, including term eigenvectors $(T_{t\times k})$, diagonal matrix of singular values $(S_{k\times k})$, and document eigenvectors $(D_{d\times k})$. At the same time, the matrices are reduced into k-dimensional space. Depending on the number chosen for k, the quality of resulting topic clusters is decided. High k value may allow to capture multiple underlying topics, but it may hold too much noise; therefore, there is in need of selecting the right k value that can satisfy such trade-offs. Here, determining the number of topics is one of the most difficult questions in unsupervised methods like LSA. There is no right or wrong answer for selecting the number of clusters; the only focus here was to keep as much information as possible while minimizing noises. Theoretically, terms appearing in contextually similar documents will be allocated in the same topic cluster, and this will result strong relationship among terms within topics but weak relationship terms from other topics (Lee et al., 2010). The LSA, in results, provides k number of topics, and constituting terms in each topic will be assigned with fractional membership value of certain cluster (Manning et al., 2008). The terms with high loading values thus represent, or even illustrates, the details of certain topic.

The last part is interpreting topics based on the keywords with high loading values. The component matrix of TS is particularly focused for scenario interpretation and construction. It is the term loading values as in common principal components of terms and documents (Blake & Ayyagari, 2012). Similar to the process of conventional factor analysis, the top loading terms are utilized to form the detailed descriptions of each scenario. The involved terms may include nouns, verbs, and adjectives, and they could inform about places, stakeholders, and human values. Each scenario's related terms are sorted by importance and combined with one another to form one big and specific storyline of target technology's future impact. Specifically, this context development process is conducted for each term cluster, independently by technology and futures experts. The ideas are then shared to settle into several concrete storylines based on discussion session and verification session. Discussion session is performed to incorporate, rather than to eliminate, diverse perspectives and opinions. In verification session, documents with high loading values are manually examined to confirm whether representative terms are relevant to the representative documents.

5.5.1.3 Conceptualization of Impact Scenarios

Lastly, we conceptualize social impacts and human values involved in each impact scenario. People pursue diverse goals and purposes, and they express their state of affairs in a variety of ways. This step is about understanding those principles or properties of things that they care about and strive to attain (Flanagan & Nissenbaum, 2014). As performed by Kwon et al. (2017), we first analyze the resulting terms in the scenarios and select specific value-related terms, or "signs", that may characterize social impacts or human values. The terms could be expressed as affective and ethics terms, and they may be associated with a variety of fields, such as personal traits, cultures, politics, and environments. Based on the overall context of each scenario, the selected terms are utilized to associate with previously determined or classified social impacts. By overviewing the problematic eventualities and their resulting impacts to society, we conduct a thorough experts' discussion to conceptualize the human values that must be embedded in emerging technologies.

5.5.2 Step 2. Deriving Technical Solutions

5.5.2.1 Collection of Scientific Publication Data

Next, we locate appropriate literature data that hold pertinent solutions of those issues. This subsequent process is an extended from the previous envisioning step; the results of LSA is used for the information retrieval. As mentioned earlier, LSA was originally started as information retrieval or search engine query purposes. For this research, representative terms with high loading values are used as the inputs for scientific publications gathering process, and it is performed separately for each scenario. The number of datasets will be, therefore, equal to the number of derived scenarios. Instead of full articles, abstracts are collected to construct the datasets. Even though the analysis of full articles may yield a more comprehensive range of information, they offer unnecessary information, such as backgrounds and methodological literature reviews. This may surely affect the performance of topic modeling since the algorithm is highly depended on the coherency and consistency of the contents.

In result, technology-related knowledge from a wide variety of academic domains will be contained in the data sets, ranging from the fields of electrical engineering to material science. When it comes to solving problems, it would be useful for technology developers to consider multiple solution spaces. For instance, certain problem could either be removed or avoided, and the solutions could be involved with using certain types of sensors or changing into different types of materials. We, thus, attempt to gather as wide disciplines of scientific knowledge as possible in order to open up the possibility of capturing various solution-like academic concepts, instead of just relying on one best resolution.

5.5.2.2 Construction of Solution Concepts

Based on multiple set of literature data, the same exact process of sub-step 1.2. is proceeded. LSA is performed for each data set, and the derived concepts from this process are considered as specific knowledge domains, or so called solution concepts. They are not just some imaginary depictions of solutions with a dearth of technological proofs. The concepts are the concise and coherent summaries of multiple or single academic themes, which have been intensively discussed and verified regarding certain problem. In this particular research, they are served as technical solution options, which may be incorporated in the early stage of technology's design process to handle a specific situational problem. The composition of terms of the resulting clusters will be different from the previous step. Technical or scientific terms, rather than social terms, will be the main constituents and the interpretation must be conducted with greater expertise and in-depth experience. Similar discussion and verification processes are further conducted as well. Consequently, a set of solutions concepts are derived for each societal issue, extracted from Step 1.

5.6 Illustrative Case Study: Autonomous Vehicle

In order to illustrate the applicability and effectiveness of our proposed approach, we have conducted a case study with autonomous vehicle technology. Also known as driverless car or self-driving car, autonomous vehicle has attracted world-wide attention for its potential benefits on emerging markets, such as car-sharing and entertainment, and has resulted societal advantages on road safety, travel behavior, and urban development. Such a new land emerging technology is sure at the brink of commercialization; however, the most fundamental problem in diffusing autonomous vehicle is the occurrence of unforeseen hazards and the proper ways of handling them. Numerous studies are being conducted in diverse academic fields, such as robotics, information processing, and even social science, in order to minimize the uncertainties of the technology. In addition to the existing attempts, we attempt to demonstrate an alternative way of conducting such a future-related problem solving process.

To extract societal issues, we investigated various websites that offer the most suitable and sufficient information regarding the ramifications of autonomous vehicle technology. The assessment of websites was conducted based on the proposed criteria, including quantity of articles, main focus of technology types, and features of involved players. Over fifteen websites were considered appropriate, and five websites, including MIT Technology Review, Wired, the guardian, io9, and Business Insider, were selected for the further analysis. Using the queries described in Table 5-2, the data was collected using our Java-based web crawler. Total of 1,491 text documents were collected in result. We have extracted total numbers of 8,440 terms and conducted the basic pre-processing steps and reduced into 4,332 terms.

Table 5-2 Summaries of data collection

Query
(“autonomous” or “self-driving” or “driverless”) AND (“car” or “vehicle”) AND (“issue” or “concern” or “worry” or “problem”)

Next step is the construction of future-oriented database based on the application of LSA text mining technique. The algorithm was realized based on Matlab software. First, the general preprocessing techniques was conducted, and the terms with out-of-context and low occurrence frequency were removed. Total 310 terms were selectively prepared to be utilized in the process of LSA. When performing LSA, it is important to decide the most appropriate number of factors since there exist a trade-off between quantity and quality of contexts. The value of k, the number of factors, were varied by 10, 15, 20, 25, and 30. The results of loading terms and documents were observed for each topic, and we concluded that 20 factors delivered not only interesting but reliable contexts of issue-related scenarios. Specifically, 10 and 15 factors offered only the main storylines, but it seemed lacking in terms of quantity; over 20 factors provided a sufficient number of contexts, but the composing terms were quite incoherent and sometimes incomprehensible. As shown in Table 5-3, the terms were grouped based on their semantic relationship and formed multiple term clusters, representing experts’ and publics’ concerns regarding the unforeseen social consequences of autonomous vehicles. Total 16 impact scenarios were derived based on an in-depth discussion of six experts of data and futures research. Specifically, the groups of terms were first analyzed and interpreted independently. The main focus of term clusters and the details were inferred based on the constituent terms and documents with high loading values. An in-depth discussion was then conducted to decide final scenarios with sufficient coherency and reliability, and only the ones that have been agreed upon by at least three experts were taken into further consideration as scenarios. Four clusters were omitted in result. Two groups contained overlapping contexts with the existing ones and the other two were hardly interpreted due to varying high loading terms and documents. The final interpretation result of those impact scenarios is shown in Table 5-4.

In result, not only the terms were grouped with enough consistency but the documents were also grouped based on the topic's main focus. Scenario 5 (Ethical dilemma), for instance, is named as "moral judgment" when its top 20 highest loading terms were observed, including "robot", "ethical", "moral", "dilemma", and "react". The result of LSA is the ingredients for experts to interpret the potential issues, which were frequently and jointly mentioned from futures experts and the publics. All six experts were able to conclude that Scenario 5 (Ethical dilemma) is mainly about "An ethical or moral dilemma of the machine's decision when accident is about to occur". In addition, the value-related terms like "ethical", "moral", "security", and "accident" suggested that various stakeholders from the web were concerned about the moral, security, and safety issues and were deeply skeptical of today's technological advancement capable of properly "reacting" to such a situation.

Table 5-3 Result of LSA describing impact scenarios

Labels of impact scenarios	Top 20 Highest loading terms	Value-related terms
1. System failure	police, design, accident, software, detect, machine, camera, computer, control, brake, wheel, malfunction, engine, critical, crash, issue, failure, problem, danger	accident, safety
2. Empty spaces	crash, object, cover, highway, empty, position, open, spot, map, problem, sensor, spot, parking, detect, spatial, difficult, measure, types, camera, assume	crash
3. Situational awareness	technology, sensor, police, monitor, react, vehicle, data, pedestrian, collision, spot, insurance, reduce, activity, safety, potholes, obstacles, detect, wildlife, uncovered, danger	safety, danger
4. V2V communication	truck, communication, collision, v2v, crash, map, system, read, turn, poor, sensor, brake, real-time, bad, night, data, capability, technology, constant, reliable	collision, crash
5. Ethical dilemma	robot, machine, ethical, human, moral, legal, computer, dilemma, question, react, data, information, situations, risk, security, principle, decision, rules, complex, accident	ethical, moral, security, accident
6. Event of emergency	driver, system, traffic, emergency, safety, autonomy, insurance, control, vehicle, information, reduce, brake, collision, detect, injury, stop, machine, law, computer, drive	safety, autonomy, collision, injury
7. Bad weather condition	accident, weather, road, poor, life, bad, rain, handle, signal, safety, situation, condition, dangerous, severe, heavy, fog, traffic light, snowy, obstacle, sensor	accident, safety
8. Privacy invasion	security, privacy, hacking, attack, cyber, data, information, terror, collect, protect, standard, communication, threat, issue, responsible, bad, police, cybersecurity, problem, law	security, privacy, threat
9. New training program	electric, standard, test, crash, train, low, market, delay, quality, brake, rule, engine, principle, firm, drug, process, challenge, build, law, legal	rule, law, legal
10. Reshape of job economy	job, robot, machine, life, social, unemployment, night, transport, education, knowledge, bad, left, poor, task, own, low, inequality, read, taxi, right	unemployment, inequality, right
11. Who's responsible	liability, transit, manufacturer, design, legal, infrastructure, principle, responsibility, robot, law, insurance, complex, quality, issue, service, firm, individual, collision, developer	infrastructure, responsibility, collision
12. Local laws	different, between, question, position, law, state, reduce, legal, turn, sign, rule, traffic, taxi, concern, suffer, area, environment, illegal, injury, object	law, legal, environment
13. Fall of gasoline	electric, job, transportation, transit, state, situation, gasoline, adoption, concern, unemployment, market, crash, eliminate, industry, new, vehicle, challenge, suffer, improve, bomb	unemployment, industry
14. Loss of autonomy	standard, right, autonomy, transport, enjoy, highway, park, quality, ride, taxi, safety, own, handle, technology, death, driver, vehicle, control, life, question	right, autonomy, safety
15. Social equity	cost, engine, afford, sign, job, expensive, bad, everyone, market, own, garage, decision, install, infrastructure, buy, low, price, handle, law, liability	afford, infrastructure, law
16. Illegal activities	drug, accident, illegal, weapon, state, lethal, detect, potential, terrorist, bomb, rule, explosive, risk, attack, delivery, reduce, suffer, security, react, high	accident, security

The value-related terms could be also served as important linking terms to conceptualize social impacts. These linking words directly imply several social impacts, like “ethical impact” and “security impact”; however, it is quite difficult to understand a more detailed levels due to their conceptual ambiguity. In response, we refer to the conventional studies from social impact assessment (SIA) or TA and incorporate pre-determined impact classification system. In this case, we adopt the list provided by Vanclay (2002). Beyond the notions of “ethical impact” and “security impact”, we were able to determine more detailed, but broader, impacts like “actual safety and hazard exposure”, “change in moral rules and beliefs”, and “quality of living environment”. Scenario 4 (Loss of autonomy), for instance, was associated with impacts of “changes in an individual’s independence” and “perception of personal safety”.

Table 5-4 Summaries and related social impacts of scenarios

[Topic label] Summaries of impact scenarios	
[System failure] safety issues posed by various accidental system failures on the equipments, including camera, brake, wheel, or engine	[New training program] challenge of setting new regulations systems and training programs necessary for new vehicles networking infrastructures
[Empty spaces] safety concerns regarding the difficulties of autopiloting in environments with open and empty spaces	[Reshape of job economy] fears of mass unemployment and economic collapse in transportation sector
[Situational awareness] issues regarding the failure of detecting and learning unexpected obstacles, including pedestrians, potholes, or animals	[Who's responsible] concerns of who would be responsible in the event of an accident? Manufacturer, developer, driver, or service provider?
[V2V communication] safety concerns regarding the errors in real-time v2v communication systems, which could lead to serious vehicle accidents	[Local laws] chaos in safety regulations due to inconsistent and conflicting mixture of federal, state, and local laws
[Ethical dilemma] ethical or moral dilemma of the machine’s decision, when accident is about to occur	[Fall of gasoline] disruption of oil market due to a widespread adoption of electricity in vehicles

[Event of emergency] deep skepticism regarding the vehicle's response to emergency situations	[Loss of autonomy] reluctance to owning the vehicles due to the reduced degree of autonomy
[Bad weather condition] safety concerns about the ways of managing severe weather conditions, such as fog, rain, and snow	[Social equity] social equity concerns due to the increased costs of private car ownership
[Privacy invasion] privacy risks associated with the collection of and access to personal information, such as location and entertainment data	[Illegal activities] security issue caused from a more diversified criminal activities using autonomous vehicles

Step 2 is about extracting technical solutions of each impact scenario. The article discusses the result of Scenario 3 (Situational awareness) for a more focused and detailed illustration of the proposed methodology. This scenario is about potential safety implication resulting from the difficulties of autopiloting in the environments with open and empty spaces. Google Scholar search engine was utilized for the data collection of literature data, and we have utilized the high loading terms from the result of LSA as the search query of the information retrieval. Once again, the experts conducted a discussion session and a trial and error session for selecting the most pertinent and suitable terms that describe the theme of each corresponding scenario. The terms related to the details, such as “police”, “pedestrian”, and “wildlife”, were less likely to be chosen as the search queries since they may cause the collection of relatively biased information. For the target scenario, the words of “react”, “data”, “collision”, “spot”, “safety”, “obstacles”, and “detect”, along with the queries of target technology, were selected for the queries for the data collection of Step 2. Total 360 abstracts are collected and total 472 terms were extracted from the text mining process. The subsequent step is the construction of solution concepts based on LSA text mining technique. We have filtered out the terms, which are out-of-context and low frequency values, and carefully selected out total 213 terms for the actual

LSA process. Furthermore, the numbers of factors were varied by 10, 15, 20, and 25. The results of grouped terms and documents were observed, the factors over 15 derived quite inconsistent and incoherent term groups and corresponding documents. The factor was consequently set to 15, since it offered relatively new interpretable and meaningful summarizations of the technical solutions.

The final result of derived solutions is shown in Table 5-5. The terms were grouped based on their co-occurrence relationships and the concepts represented succinct summarizations of a wide variety of scientific knowledge. Compared to the result of the previous step, the analysis of literature data demonstrated more coherent concepts based on two reasons: the abstracts are composed of professional and scientific terminology, thereby including much less noise, and the use of proper search queries resulted the construction of organized and consistent database. Out of 15 solution candidates, we have conducted another experts' discussion and deliberately selected 11 term groups, which can be considered as potential solution concepts.

Although each concept may offer a different technical approach from one another, these total of 11 solution concepts are the key technical ideas or themes, which may ultimately increase a specific social value of "safety". The results were labeled based on the top five highest loading terms of each concept. Solution concept 2, for instance, was named "texture assessment" based on the terms of "pavement", "texture", "assessment", "pothole detection", and "reconstruction". According to the analysis, we have concluded the concept was mainly about the algorithms of assessing the conditions of various pavements, such as cracks and opened potholes. All solutions concepts were validated based on the evaluation process of documents with high loading values.

Table 5-5 Result of LSA describing solution concepts

Technical solutions	Top 20 highest loading terms	Summaries
1. Localization (LaneQuest)	lane, lanequest, localization, location, curve, relative location, accuracy, algorithm, estimation, velocity, ambiguous, smartphone, detection system, energy, automatically, evaluation, implementation, deployment, gps, lane boundary	accurately estimating the surrounding environments to detect the vehicle's lane by localization technique
2. Texture assessment	pavement, texture, assessment, pothole detection, reconstruction, region, automated pothole detection, severity, gpr, depth, recognition, accuracy, computation, effectiveness, identification, 3d, intensity, efficiently, detection accuracy, distress	algorithms of assessing the conditions of various pavements, such as cracks and opened potholes
3. Vehicle infrastructure	safety, infrastructure, vii, mobility, highway, transportation, vision, sensor, civil infrastructure, software, system, igvc, deployment, vision system, capacity, productivity, lidar, perception, presence, radar	vehicle infrastructure that emphasizes safety warning system
4. Incident detection system	incident, management, lane, accident, unit, framework, internet, incident detection, productivity, generation, destination, location, device, communication, sensor, database, intensity, relative location, drivers, safety	automatic detection system of various traffic incide
5. Vehicle sensor network (VSN)	safety, sensor, vsns, vii, infrastructure, recognition, deployment, sonar, transportation, road detection, communication, highway, behaviors, solution, ultrasound, obstacle avoidance, avoidance, vehicle sensor, radar, assessment	intelligent vehicle network based on advances in communications
6. Future combat system (FCS)	destination, sensor, fcs, system, path, environment, intensity, device, response, ugv, control, pothole, possibility, operation, capability, radar, safe path, detection accuracy, module, area	robotic system used in combat situation capable of handling mobility hazards like holes and ditch
7. Backscattering	lidar, backscatter response, response, liability, navigation, road, property, surface, sensor, path, pothole, destination, database, capacity, collision, solution, unstructured, intensity, logic	backscatter response to measure various faults on road surfaces
8. Localization (Dejavu)	dejavu, city, localization, automatically, energy, gps, error, management, highway, sensor, tunnel, battery, internet, productivity, efficient, evaluation, navigation, communication, robotics, outdoor	localization approach that leverages road landmarks, including bumps and potholes
9. Advanced driver assistance system (ADAS)	adas, depth, collision, avoidance, direction, transportation, alert, camera, safety, lidar, reconstruction, roadway, assistance, severity, driver assistance system, recognition, awareness, software, scenario, logic	driving system that recognizes upcoming road environments using cameras and sensors
10. Obstacle avoidance technology	avoidance, obstacle avoidance, obstacle detection, laser scanner, ultrasound, sonar, navigation, waypoint, logic, radar, laser, pothole, unstructured, sensor, adaptive, region, evaluation, calibration, illumination, computation	obstacle avoidance system that detects and avoids any road obstacles based on sensor network
11. Catadioptric system	catadioptric, infrastructure, area, system, inspection, depth, sonar, sensor, scenario, assessment, accuracy, accelerometer, relative location, occlusion, awareness, transducer, grid, stereo vision, calibration, optical	perceiving various information around a vehicle by using catadioptric sensors with optical axis

To sum, we show overall procedure of reaching the conclusion of responsible development. As illustrated in Figure 5-4, an emergence of new technology like autonomous vehicle arises unpredictable and dangerous eventualities, like 16 impact scenarios. Every possible case is likely to result various social impacts, ranging from quality of living environment impact to institutional, legal, and political impact, and numerous existing social norms may be subsequently affected by those consequences. This is the part where we must notice specific social values that must be embedded in the design of emerging technology. As shown in the middle of Figure 5-4, numerous values were found to be associated with the impact scenarios. Based on the premise that technology can embody values, we say system design of autonomous vehicle must be typically guided by certain goals, including safety, security, equity, privacy and etc. A particular eventuality like Scenario 3 (Situational awareness), for instance, may result the degradation of "safety". In response, 11 state-of-art scientific knowledge concepts are the preventive guidelines for designers and producers of autonomous vehicle technology, suggested to demonstrate the way to translate the value of safety into the design process.

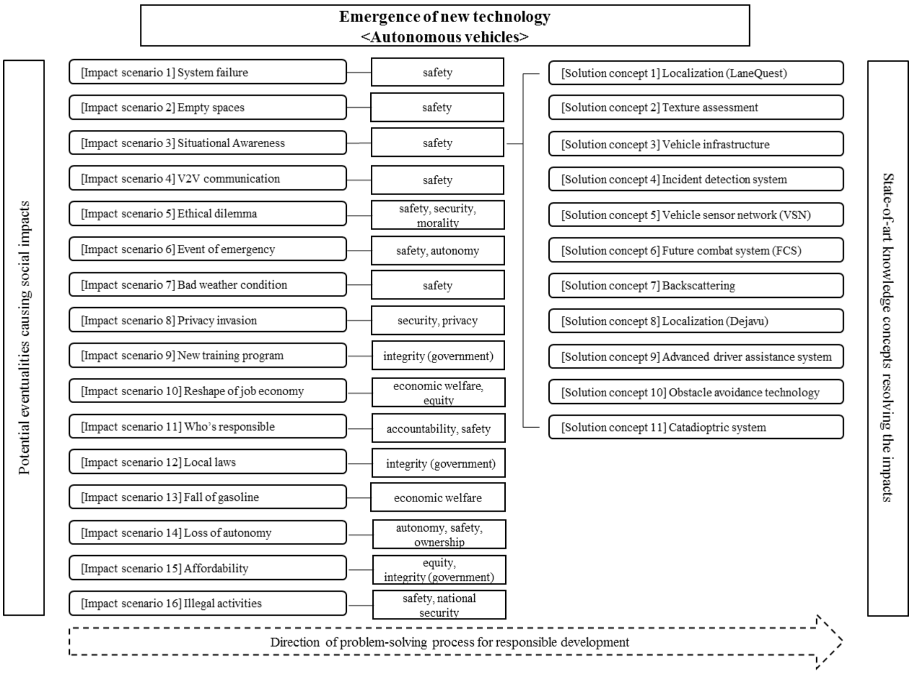


Figure 5-4 Overview of impact and solution results for autonomous vehicle

5.7 Discussion

5.7.1 Comparative Analysis

To further strengthen internal validity and generalizability of the proposed, we have conducted two comparative studies with existing literature. Previous articles with the most similar objectives were selected for the analysis: (1) one with same target technology but more conceptual approach and (2) one with same systematic procedure but different target technology. First, Leenes and Lucivero (2014) has analyzed the assertion of "embedding rules in the positronic brain of robots" and proposed previous attempts and failures to embed legal norm compliance into the design of robots, particularly self-driving vehicles. The study has not offered separate sections of potential issues and corresponding technical solutions; it has rather deliberately discussed four specific means of preventing various issues from occurring: 1) regulating robot design, production through law, 2) regulating user behavior through the robot's design, 3) regulating the effects of robot behavior through law, and 4) regulating robot behavior through code. A wide ranging situational issues were presented. offered a wide ranging situational issues, including "interacting safely with other vehicles and pedestrians", "passing a slow moving vehicles", "correctly stopping at a stop sign", "successfully navigating a jammed intersection", and "understanding verbal instructions or gestures by authorized officials". Furthermore, the study has yielded several technical solutions with details, including "structured driving", "unstructured driving", "distance keeping", "merge planning", and "localization".

The issue-related results presented by the conventional study seemed more practical and reliable considering that the research was conducted based on experts' knowledge and historical data sources. Most of the potential issues were closely related to the problem whether the vehicle could obey current traffic laws and regulations. On the other hand, the proposed approach offered more detailed and future-oriented issues, including a privacy invasion risk associated with the collection of and access to personal information and social equity concerns related to the increased costs of private car ownership. The solution-related results were quite similar to that of the proposed methodology; however, our result suggested more detailed techniques and algorithms, such as LaneQuest, FCS, backscattering, and catadioptric system. Above all, the most distinctive difference between these studies was the information regarding human values. The presented study underlines the significance of a heterogeneous set of requirements, including safety, human health, privacy, justice, and sustainability, but also points out the vagueness of human values and how they could be directly used and translated in the design process. The proposed study, in response, approaches in a quite different manner. We have used semantic relations among the words in textual data and proposes a novel way of realizing such responsible development of emerging technologies based on text analysis.

Second, Denning et al. (2013) has attempted to realize responsible development of smart home technology through conducting a value sensitive design approach. The authors have surveyed potential attacks against in-home technologies and identified human values posed by such hazardous circumstances. Then the study suggests various security goals and device structures to avoid them ex-ante. There exists an array of potential attack scenarios to be considered, like “resource theft”, “blackmail”, “viewing sensors”, and human assets for possible degradation, like “emotional well-being”, “personal data”, and “relationships”. The solution or security goals were also thoroughly suggested. Three major goals were device, digital data, and environment goals, and each included 4 or 5 sub-goals, including device availability, data privacy, and sensor validity.

As presented, conventional approach yields a wide range of challenges and attacks within home technologies and suggests a number of guidance for the security needs. The results are structured by constructing a taxonomy of attack scenarios, and they have extensively covered both usual and unusual consequences. The level of reliability is considerably high, since the result was derived directly from experts. However, several limitations must also be addressed. First, the details were lacking. For instance, the study has presented “viewing private data” for one of the potential attacks to the home ecosystem;

however, various types of data with a large volume are being processed in different system structures. Moreover, there exist so many different routes to invade their privacy. It is, therefore, essential to provide specific situations including causes and effects associated with such an attack. On the contrary, our data-driven foresight approach has not only yielded information regarding causes and effects but also included much of society-related terms, such as unemployment, afford, security, and responsibility.

Second, although the study has provided security goals that pertain to the home technologies, the resulting solutions included too general information. The goal of "data availability" was only described as "defenses should ensure the user's data does not suffer from malicious access condition". Such a guidance is meaningful to the technology developers or designers; however, the description is too vague and abstract. On the other hand, the proposed approach relatively more practical solutions, which have been derived from a wide variety of scientific knowledge. It is capable of acquiring technical directions or guidance used for a purposeful design. Further distinctive strengths are elaborated in further sections.

5.7.2 Major Strengths in Envisioning Social Impacts

The effectiveness of the first step is elaborated with the following arguments. First, we are neither criticizing nor denying fundamental advantages that participatory method could offer, such as data reliability and context plausibility; however, we are herein highlighting the need of capturing efficiency and creativity in performing such an activity. The analysis of future-oriented text data not just reduces the time and effort in observing the overall concerns of emerging technologies but also leads to a whole new approach of developing future scenarios. A scenario is just a storyline, an essay, or a short novel that describes a certain series of future eventualities. It is normally in the form of one or two paragraphs; each paragraph is composed of multiple sentences; a sentence comprises several terms or keywords. The result of LSA offers the group of terms, capable of to be served as the ingredients for constructing the scenarios. Specifically, three different part of speech is involved here: noun, verb, adjective, and such morpho-syntactic features of these candidates are used to form sentences and further into one coherent storyline. A context can be constructed based on the configurations of noun – noun, noun – verb, or noun – adjective – verb. Scenario 15 (Social equity), for instance, is about the technology's affordability. The details of the scenario can be defined as "not everyone can own/afford/buy an autonomous vehicle due to its expensive price" based on the key terms, including "everyone", "own", "afford", "buy", "handle", "price", and "expensive". Through re-combining multiple constituting terms, we were able to catch a glimpse of the main idea of impact scenarios.

Second, we observed insightful details of impact scenarios, including the causes and effects of the unforeseen consequences, when the concerns were overviewed based on LSA. In details, the terms that describe technical features or functions were the main causes of the impact scenario; furthermore, the terms that portray human values, involved players, or places were considered as the effects of the scenario. By way of illustration, Scenario 8 (Privacy invasion) was about privacy invasion issue regarding autonomous vehicles. Several terms indicated functions and features that are likely to cause the problem, including "hacking", "attack", "collect", and "communication". Furthermore, certain terms were related to affected factors, such as "security", "privacy", "terror", "cybersecurity", "police", "law", information". These key terms suggest that autonomous vehicles handle extra communication information beyond the reach of the existing vehicle systems, such as vehicle-to-vehicle (V2V), vehicle-to-infrastructure (V2I), vehicle-to-pedestrians (V2P), and vehicle-to-cloud (V2C). Such a new type of personal information is being threatened, thereby raising concerns on unprecedented security and privacy issues. Related to this, the term "police" must be highlighted in this context. When we analyzed the documents with high loading values, it was found that the public was quite unnerved with law enforcement agencies gleaning such personal information from autonomous vehicles. The details are sometimes obvious or too absurd. However, it was found that general information promoted a better understanding of the main topic of a term cluster, and illogical terms provided unexpected and surprising details.

5.7.3 Major Strengths in Overviewing Solutions

The LSA result of the second step demonstrated not only contextually comprehensive but also meaningful information for resolving potential issues caused from emerging technologies. However, more specifically, two arguments should be heightened to demonstrate its effectiveness. First, the resulting terms offered technology development guidelines or recommendations, specifying potential technical requirements that must be highlighted in the early design process. Solution concept 2, for instance, was associated with technical algorithm of texture assessment. Other than aforementioned representative terms, certain terms like "severity", "gpr", "depth", "accuracy", "intensity", "3d", and "efficiently" suggested very detailed technical information for the purpose of preventing the issue of situational awareness. Specifically, there was in need of a technology or an algorithm capable of accurately and efficiently measuring the severity (intensity, depth, width, etc.) of cracks and holes, and 3d ground-penetrating radar (gpr) was a frequently-used technology in such a circumstance. Following considerations are the highlighted themes frequently dealt in the pertinent research area. The terms were, in fact, the key factors or a more specific technical feature frequently referred by multiple scholars to solve this particular issue. They may be served as technical directions or design recommendations, which must be at the heart of the technology's engineering design.

Second, according to the study from Grunwald (2011), a proper responsible development requires an intense inter- or trans-disciplinary collaboration between engineering, social sciences, and applied ethics. Our proposed approach was quite satisfying in this perspective. The result offered a wide range of academic fields, possibly accommodated in technical solutions for responsible development. Specifically, certain impact scenarios were associated with various academic fields, ranging from engineering to social studies. The Scenario 5 (Ethical dilemma), for instance, is about moral dilemma of autonomous vehicles when faced with an accident. When pertinent abstracts were collected and analyzed, we found multiple solution concepts in a wide variety of domains, including social robotics, information sciences, machine ethics, computational intelligence, and autonomous systems. Some scenarios were exclusively involved in technology-oriented disciplines, ranging from engineering to computer science. The resulting solution concepts from Scenario 3 (Ethical dilemma) suggested several different domains, including mobile computing, civil engineering, avionics, optics, and microelectronic engineering. The result of Scenario 3 (Ethical dilemma) first seems solely depended on robotics domain. However, there exist more diversified fields within the robotics field, and it is worth noting these specific academic disciplines for the purpose of promoting a more effective collaboration for responsible development.

5.8 Conclusion

When it comes to the responsible development of emerging technologies, there has been a prime gap between philosophizing and existing. A guidance or a clear-cut methodology was in need for engineers to develop them with some sort of responsibility. As a remedy, this article offers a systematic and quantitative method of realizing such a notion. The proposed method is a data-driven problem solving approach capable of envisioning potential social problems caused from the emergence of new technologies and further determining technical design focuses that could increment human values. The basic premise of our approach is that an abundant number of stakeholders who stand to be directly impacted by new technologies being introduced and who have tacit knowledge of their application (Šabanović, 2010) should be somehow involved and included in the development of new technologies. Instead of relying on a few numbers of experts, we attempt to incorporate as much diverse knowledge as possible and thereby expanding the pool of collective intelligence in envisioning the impacts and identifying the solutions for them.

Despite the abovementioned effectiveness, the proposed approach shows several drawbacks that hinder emerging technologies to be responsibly developed. First, the method is not capable of suggesting any solutions for unintended impact scenarios, such as impacts of job environment, vehicle law, and gasoline industry. In this case, the result of the proposed approach may only be served as a way to understand the future. No technical solutions can be suggested, other than promoting stakeholders to prepare a responsive strategic plan. Second, the proposed approach failed to capture creative and insightful knowledge of the few. In pre-LSA, hundreds or even thousands of low frequency terms are omitted in order to reduce noise and thereby to increase performance. There is in need of a more sophisticated identification algorithm for discriminating valuable information from the noise. Lastly, the approach still requires some subjective judgments from the experts equipped with a high level of technical and futures knowledge. Significant improvements particularly in pre-processing step and interpretation step are remained for the future work.

Whilst many believed that not much of reliable knowledge sources exist regarding the future depictions of emerging technologies (Grunwald, 2011), we have offered a new data source, so-called *future-oriented web data*, and demonstrated its usefulness through extracting a holistic overview of potential fears and concerns of new technologies. Even more, the text analysis of literature data has demonstrated its effectiveness in acquiring technical directions or guidance used for a purposeful engineering design. Taken together, such a data-driven problem-solving process is expected to guide emerging technologies to yield only the right impacts to our society, by that resolving the issues of rejection and non-acceptance of such unfamiliar technologies.

Chapter 6

Foresight for Technology Ideation

The final step of proactive management process before entering the general technology management scheme is to generate creative and innovative ideas regarding the future structures of emerging technologies. In order to better exploit the potential of a new technology, it is vital to thoroughly understand how technology is going to turn out or look like in the near future. This issue has been dealt in the field of idea generation. This study is normally carried out by combining disparate bodies of existing knowledge. However, two major drawbacks of conventional methods must be addressed: (1) involvement of human subjectivity in idea generation and (2) incorporation of unsuitable data source considering highly complex and uncertain nature of today's matters.

This chapter offers a Wikipedia-based approach to the development of morphological matrix, which could be served as a supporting and supplementary tool in the field of idea generation. The data source, utilized in this field of study, has been expanded to a new boundary. Valuable features of Wikipedia data for idea generation have been investigated, and they were further incorporated to build two models for constructing a new form of morphological matrix. Given its case-specific contents and well-coordinated structure, the utilization of Wikipedia in morphological analysis promoted more fruitful ideas regarding the future structures of emerging technologies.

6.1 Introduction

Innovation has been a central issue in both academia and practice. The general definition of innovation is discussed in numerous research studies. Evangelista (1998) defined innovation as a process from research to invention and then a diffusion of a new technique. A study by Drucker (1985) defined as “the act that endows resources with a new capacity to create wealth”. In recent years, numerous scholars emphasized the importance of generating creative ideas in achieving such an innovation (Girotra et al., 2010; Rietzschel et al., 2014) and further proposed the notion that creative and innovative ideas originate from the right combination of disparate bodies of existing knowledge (Ward, 2004; Dosi, 1982; Geschka, 1983; Bjork & Magnusson, 2009; Schilling & Green, 2011; Nakamura et al., 2015).

Morphological analysis has been particularly employed as a prominent tool for generating new ideas. It is a method that leads to structured inventions by determining all possible alternatives for solving a certain problem (Wissema, 1976; Yoon & Park, 2005; Geum et al., 2016). This technique has two strong advantages for idea generation. First, morphological analysis decomposes a complex system into parts and systematically rearranges combinations to generate ideas. This can be considered a “combinative” characteristic of the technique that sparks an abundant supply of creative ideas. Second, morphological analysis breaks down the target subject and reconstructs them to explore unprecedented structures. This can be considered an “inventive” characteristic of the technique that explores many different novel ideas, thus providing possible solution options for a given context.

Despite its popularity, one innate shortcoming of morphological analysis must be solved: the involvement of human subjectivity. The entire process of decomposing and restructuring into new knowledge can exist in the first place only if the morphological matrix is designed properly and specifically. This preprocessing step is the most fundamental step when considering the notion of GIGO–short for garbage in, garbage out. Previous studies, however, handled the morphology building process in a qualitative manner that depended on experts' opinions. Regardless of their advantages in capturing high reliability and validity, the involvement of human judgment was subject to numerous cognitive biases that could subsequently lead to a decrease of team performance and serious misjudgments (Zec et al., 2015). Furthermore, the conventional participatory process can no longer be applied to recently emerged products or services, which are so complex that a concrete decomposition cannot be achieved solely by a handful of professionals. For these reasons, there is in need of a shift towards a data-driven methodology, which could offer more objective and automatic results.

Several attempts have been made to suggest a data-driven approach of conducting morphological analysis, and patent data was generally incorporated as the main data source (Lee et al., 2007; Yoon et al., 2008; Yoon et al., 2014). However, one of the major limitations was that they were restricted to a single knowledge source of patent. Although it is renowned for its high reliability and usefulness, patent data often could not ensure up-to-dated, non-technological, or case-specific idea generation. To address this limitation, Geum and Park (2016) suggested WordNet-based morphological analysis. Their study highlighted the hierarchical network structure of WordNet, which has the capability of analytic and objective morphology development. However, WordNet was still subject to data rigidity and lacking domain-specific knowledge.

In response, this paper proposes a novel method of data-driven morphology building process using Wikipedia data. Wikipedia is a massive online repository of collective intelligence. The utilization of Wikipedia can be an excellent way to address the rigidity of WordNet. In terms of practicability, Wikipedia contains information from a wide range of fields and expands the data set to a new boundary. Unlike WordNet, Wikipedia offers a significant amount of case-specific knowledge that could better stimulate more specialized and practical idea generation. In terms of flexibility, Wikipedia provides hierarchical relationships of concepts, which enable a coherent matrix development of morphological analysis. One thing to note is that the objective of this research is neither to propose the most effective method of developing creative ideas nor to supersede conventional idea generation techniques. The approach is rather a heuristic process model, attempted to demonstrate the key role of Wikipedia in morphology building step for ensuring conceptual breadth and facilitating creativity.

6.2 Related Studies

6.2.1 Generating Creative Ideas

Generating new ideas has been a constant concern. A considerable literature has, thus, endeavored to address novel methods of idea generation. Starting with Osborn's brainstorming (Osborn, 1957), diverse idea generation techniques came into existence, including brainwriting, checklists, and synectics (VanGundy, 1981; Geschka, 1983; Ivanov & Cyr, 2014). They all made a tacit assumption: *quantity breeds quality* (Rietzschel et al., 2014). In other words, these techniques have assumed that as more ideas are created, the greater the possibility that creative ideas are found among them (Diehl & Stroebe, 1987; Girotra et al., 2010). As suggested by Shah et al. (2003), idea generation methodologies are classified into two primary groups: intuitive and logical. Intuitive methods are divided into germinal, transformational, SMAPER, progressive, C-sketch, organizational, fishbone diagram, and hybrid. Logical methodologies are divided into history-based methods and analytical methods such as TRIZ.

However, in recent years, the quality of the ideas has been questioned (Girotra et al., 2010, Rietzschel et al., 2014). Most studies underlined the concept of creativity as improving the quality (Rietzschel et al., 2014) and began to facilitate this improvement in idea generation process. Most studies demonstrated in a qualitative manner. For instance, Girotra et al. (2010) proposed the hybrid structure−in which individuals first work independently and then work together −and identified the structure's superior performance in terms of both quantity and quality. Rietzschel et al. (2014) applied a new brainstorming experiment and demonstrated that ideas became more creative as problems narrowed and instructions became more creative. In contrast, several studies attempted to demonstrate the improvement of quality in a quantitative manner. If a quantitative approach is taken, what is the best scale that determines a creative idea? According to Simonton (2013), the simplest but most useful answer is to treat criteria as dichotomous features. For example, a creative idea is original, rather than unoriginal, and useful, rather than useless. However, a definitional disagreement regarding the concept of creativity is inevitable given its subjective and versatile characteristics. Some research considered that creative ideas are both original and feasible (Sternberg, 1985; Diehl & Stroebe, 1987; Rietzschel et al., 2014). Other studies highlighted the concept of interventions during the generation process and demonstrated that ideas derived in the face of greater obstacles are considered more creative (Ward, 2004). Boden (2004) offered three criteria of novel, valuable, and surprising. As shown, the subjective concept of creativity was described through other subjective notions, such as original, novel, and valuable.

More recent studies focused on the methodology itself that generates new and creative ideas, and many of those stressed the underlying concept that creative ideas are crafted from existing knowledge by identifying novel combinations of previously separated ideas or concepts (Dosi, 1982; Geschka, 1983; Ward, 2004; Bjork & Magnusson, 2009; Schilling & Green, 2011; Nakamura et al., 2015). In other words, a method that can systematically combine valuable knowledge from novel data sources is expected to generate creative ideas.

6.2.2 Data-driven Morphological Analysis

Morphological analysis is the study of basic forms or patterns of a thing (Ritchy, 2011). When viewed from the perspective of *totality of things* (Zwicky, 1957), a certain object can be divided into multiple parts, and different arrangements of those parts could conform to create different wholes. The decomposition can be organized into a table, so-called morphological matrix, composed of dimensions and values. As an illustration, an unmanned aerial vehicle (UAV)−commonly known as drone−can be decomposed into four dimensions: body, sensors, actuators, and software. Generally, the body design of small UAV has 4 propelled rotors and is called quadcoptor. If the body changes from 4 to 8 rotors, the UAV is an octocopter and serves a very different purpose than that of a quadcoptor. Apparently, every dimension includes more than one value; for example, the body dimension comprises four, six, and eight rotor designs. When considering that a UAV is a combination of one value per dimension, what happens if every dimension includes 5 value options? The number of configurations increases exponentially to 625 (5 x 5 x 5 x 5) combinations. Because a small change in a value could lead to a large difference in a whole, each and every combination should be considered a candidate for creative ideas that may ultimately lead to an innovation (Zwicky, 1957).

According to Ritchey (2005), morphological analysis was first introduced in the field of biology and was used to generalize the combinatorial logic of the biological structure of organic bodies. Then, Zwicky (1957) proposed a generalized morphological analysis in the late 1940s that has been applied in a wide variety of scientific disciplines, including product planning (Geschka, 1983), technological forecasting (Jones, 1976), and futures study (Ritchey, 2005). However, such preliminary research was generally conducted in a qualitative manner, and some scholars underlined the limitations of an expert's subjectivity and lacking creativity. According to Zec et al. (2015), human bias was considered a critical bottleneck for effective collaboration for developing morphological matrix.

To address this issue, several quantitative studies have been conducted. Lee et al. (2007) proposed a new IT-based service concept generation method using patent analysis. Yoon et al. (2008) used information from product manuals and patent documents to develop a morphological analysis-based technology roadmap. Yoon et al. (2014) developed technology morphology and product morphology on the basis of patent information. However, the application of patent data is deemed unsuitable of the following reasons. First, there exist a time-lag between patent application date and patent grant date. Such an intrinsic drawback of patents cannot fully ensure up-to-dated knowledge source when constructing a morphological matrix and generating novel and non-obvious ideas. In fact, this is particularly critical considering today's fast-paced and ever-changing business environment. Second, patents are excellent proxy measures for technology, but they cannot be directly applied to non-technological and case-specific idea generation.

A recent study by Geum and Park (2016), in response, attempted to integrate WordNet into morphological analysis. WordNet was considered an excellent alternative to patent data (Geum & Park, 2016) since it is a large electronic database that forms a semantic network of words, interconnected through relations among meanings (Poli et al., 2010). The meronym/holonym relations and the hyponym/hypernym relations are employed to build a morphological matrix by providing dimensions and values. However, WordNet was still subject to knowledge rigidity, and the approach had a major limitation in terms of lacking domain-specific knowledge.

6.3 Technology Foresight using Wikipedia

6.3.1 Wikipedia as a Good Remedy

This study proposes a Wikipedia-based morphological analysis to address the limitations of prior studies. Wikipedia is a massive online repository of collective intelligence, which allows almost anyone to become contributors to information development (Milne et al., 2007; Mihalcea & Csomai, 2007). In fact, Wikipedia is the largest, fastest growing encyclopedia in the world (Wang

& Domeniconi, 2008; Milne et al., 2007). Wikipedia's distinguished features are as follows: (1) extensive topic coverage; (2) up-to-dated context; (3) domain-specific description; and (4) rich semantics (Strube & Ponzetto, 2006; Joorabchi et al., 2015). Apparently, Wikipedia holds millions of articles covering subjects in all areas of human knowledge and is continuously updated with new information (Joorabchi et al., 2015). It could be utilized as a great source for technology foresight activity since Wikipedia data is one of the most prominent example of *collective intelligence* (Schatzmann, 2013), and recently has it been mentioned in numerous futures research. Gheorghiu et al. (2008) mention that Wikipedia-like systems yield relatively reliable information in a future-related participatory approach. Moreover, Wikipedia is considered a great tool for achieving an open and collective approach in technology foresight (Cagnin & Konnola, 2014). Smith & Saritas (2011) discuss that Wikipedia is the main platform for an extensive collaboration and a crowd sourcing, which could promote more analytical foresight activities. Schatzmann (2013) asserts that it is one useful database capable of gathering individual intelligence and creating a bigger depiction. Furthermore, the methodological strength of Wikipedia is two-fold: its ability to provide domain-specific concepts and their semantic relationships, including equivalence, hierarchical, and associative relations.

Table 6-1 Characteristics of Wikipedia and a morphological matrix

Advantage of Wikipedia	Strength in building a morphological matrix
Case-specific knowledge	- Offers specialized and practical knowledge, thereby increasing the possibility of detecting new and practical ideas
Semantic relationships between concepts	- Allows to decompose the subject into parts in automatic and systematic manner given its semantic relations

First, Wikipedia provides specialized and practical concepts (Strube & Ponzetto, 2006). An idea could be valued only if it has been transformed into practical reality, an application across a variety of contexts. When Wikipedia is incorporated in morphological analysis, its case-specific knowledge may improve the quality of input, thereby increasing the possibility of detecting novel and useful ideas. Second, a rich semantic relationships allow for a systematic idea generation process. In terms of equivalence relations, Wikipedia offers alternative concepts, such as synonyms, acronyms, and common misspellings, of a certain concept through *redirect links*. For instance, *drone* includes a *redirect link* to *UAV*. Moreover, *disambiguation pages* resolves the conflict of articles having the same page titles by providing the list of concepts with identical names. The hierarchical relations of broader and narrower concepts are also demonstrated through *categories*. Every article is required to have at least one category, and these categories can be further categorized using other parent categories (Bunescu & Pasca, 2006). For example, *drone* is categorized in *robotics* and *emerging technologies*, and *emerging technologies* is further classified in *futurology* and *technology forecasting*. As is shown, higher-level concepts of a certain subject can be derived from *categories* page. Other than *categories*, certain pages such as *subcategories* and *pages in categories* provide useful hierarchical relations of the concept. Last but not least, associative relations are represented through *hyperlinks*. These relationships present the connections between the concepts that are closely related in a non-hierarchical fashion. *Drone*, for instance, is associated with the concepts of *antenna*, *analog-to-digital converter*, and *avionics*, and these terms constitute one of the sentences in the *communications* section of the *drone* article. In summary, Wikipedia can be a great complementary and valuable source for constructing a morphological matrix in idea generation, as listed in Table 6-1.

6.3.2 Preliminaries: How to Apply Wikipedia

Since there exist many different types of information in Wikipedia, this section thoroughly investigates the specific types, which could be useful for developing a morphological matrix. The information types to be investigated are *tables of contents, hyperlinks,* and *categories*, as shown in Figure 6-1. Each type has its distinctive strengths and weaknesses, as shown in Table 6-2.

Figure 6-1(a) Snapshot of *table of contents* of car

Figure 6-1(b) Snapshot of *hyperlinks* of car

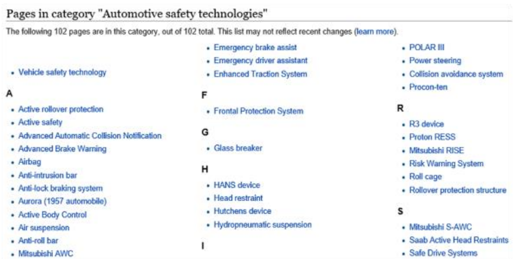


Figure 6-1(c) Snapshot of *pages in categories* of automotive safety technologies

Figure 6-1 Useful information types for morphological analysis

First, *table of contents* shows the summary of an article's contents, offering a clear overview of its structure. Such a feature allows to easily and precisely obtain the subject's composition information. Furthermore, various types of information are included in the list, and this may diversify input sources for matrix construction not only with technical but with non-technical contents, such as social, political and economic aspects. To understand the whole of a certain thing, one must grasp every aspect of it. The application of *table of contents* is thus powerful in dimension building process. The *table of contents* for "car", for instance, offers a solid list of dimension candidates, including "fuel and propulsion technologies", "user interface", "lighting", "weight", "seating and body style", and "safety", as indicated in Figure 6-2. The disadvantages, however, are its inconsistency and lacking contextual breadth. Whilst the secret to the abundance of knowledge is collective intelligence, a handful number of editors – from expert scholars to casual readers – cannot create fully coherent and consistent article structures. The layouts are often different from one article to another. For example, some sections are composed of multiple concepts, like "fuel and propulsion technologies" and "seating and body style". Moreover, the sections included in *table of contents* are subject to a limited contextual breadth, causing incoherency in developing a morphological matrix. Such a setting is a serious challenge when developing morphological matrix for a general use. A supplementary discussion is required to better organize the matrix.

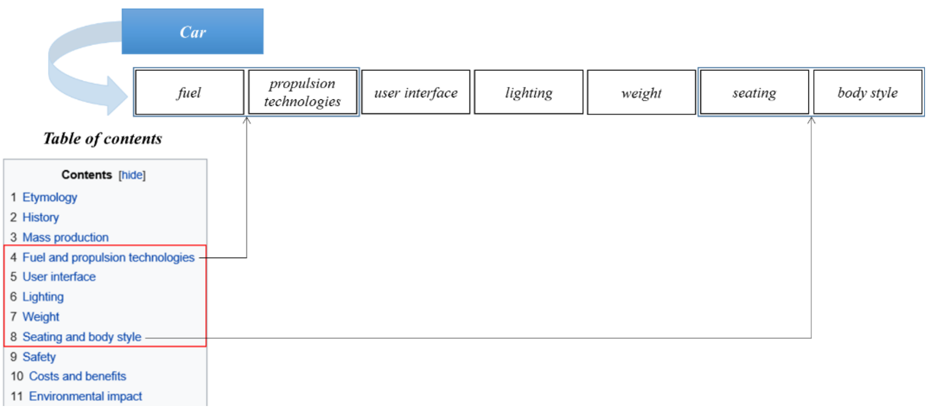


Figure 6-2 Example of dimension building using *table of contents*

Second, *hyperlinks* represent domain-related concepts closely associated to the corresponding sections. In a situation where each article is composed of thousands of words, *hyperlinks* are useful indicators for selecting suitable concepts in developing the matrix. It not only allows users to effortlessly retrieve the most relevant information but also serves as informative descriptors of the subject page (West et al., 2015). Therefore, we could consider them as the target's relatively essential elements. They would be particularly valuable in value development in morphological matrix. The *hyperlinks* in the section of "seating and body style", for instance, contain relevant concepts like "full-size cars", "sport utility vehicles", "sports cars", "sedan/saloon", "hatchback", "station wagon", and "minivan". These subsidiary concepts could be assigned as values for "seating and body style dimension", as shown in Figure 6-3. As shown, the main advantage of using *hyperlinks* is that they ensure conceptual significance and conceptual relevance. However, some *hyperlinks* may yield too much details with arbitrary hierarchical information. Such excessive description and non-hierarchical nature are considered primary disadvantages of its application.

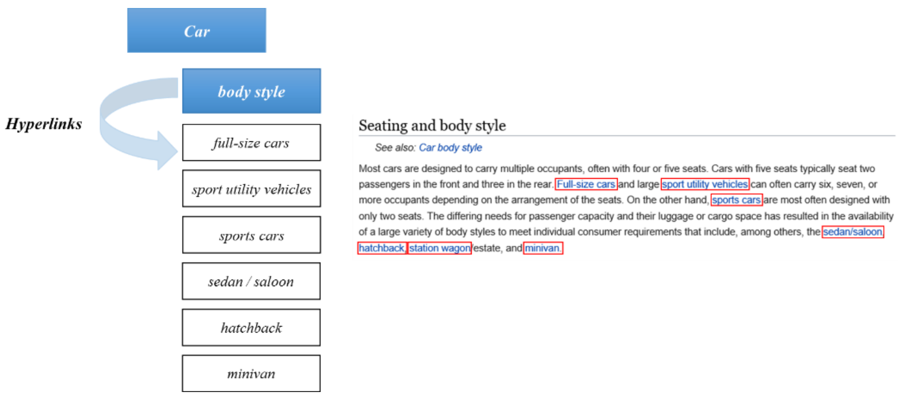


Figure 6-3 Example of dimension building using *hyperlinks*

Third, *categories* provide all concepts above a certain concept; *pages in categories* provide all concepts below the concept; and *subcategories* provide concepts from one level below the concept. Such a categorical system offers navigational links to Wikipedia articles. This is a great source for supplementing both dimensions and values of morphological matrix because both superordinate and subordinate concepts, which are related indirectly to the target subject, could be extracted based on their hierarchical relationship. Like the concepts of “alternative fuels” and “automotive lamps” in Figure 6-4, *categories* yield a list of concepts that could potentially play intermediary roles for supplementing subsidiary concepts, supplied by *pages in categories*. Two different methods exist for such a value expansion, as illustrated in Figure 6-4(a) and Figure 6-4(b). The former uses a same-level concept; whereas, the latter uses a higher-level concept from *categories* in order to generate the subsidiary concepts from *pages in categories*. Furthermore, *subcategories* could be utilized for expanding the number of dimensions, as presented in Figure 6-5. On the basis of dimensions constructed from the *table of contents*, *subcategories* could supplement additional concepts, including “automotive styling features”, “car body styles”, “car crime”, and “car culture”. However, there often exist an excessive number of concepts within *subcategories* and *pages in categories*. A group discussion is necessary to extract meaningful and relevant concepts. In addition, its proper application is quite difficult to achieve since *categories* provides quite complex and inconsistent conceptual structures.

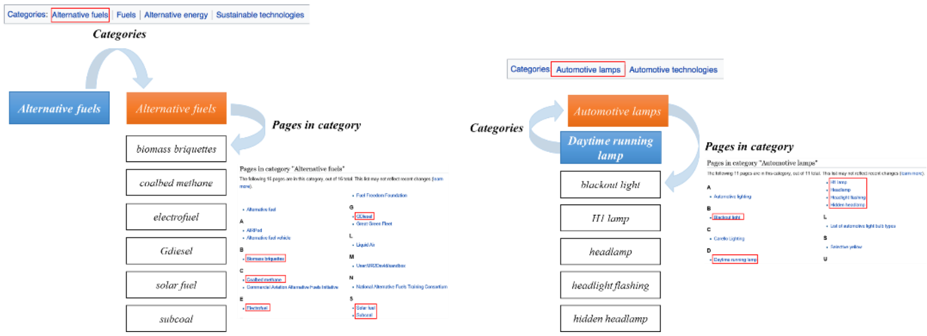


Figure 6-4 Example of expanding values using *categories*

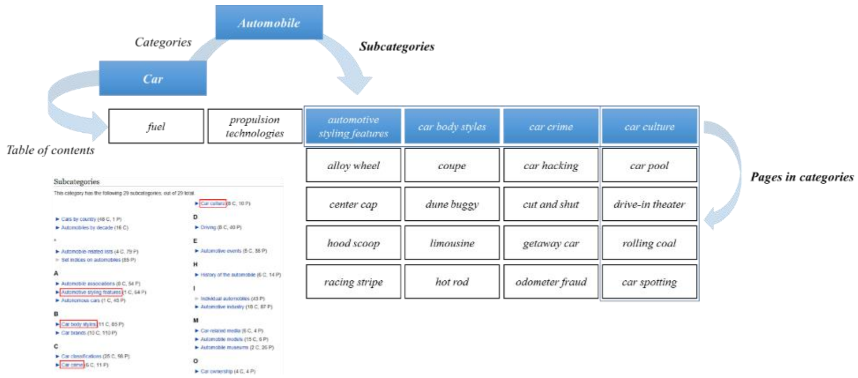


Figure 6-5 Example of expanding dimensions and values using *categories*

Considering their advantages, the utilization of *tables of contents*, *hyperlinks, and categories* is a great starting point for developing a data-driven morphological matrix. This article determines how to construct a detailed yet concise matrix by offsetting their respective limitations. First, information in *table of contents* is used to extract only the interested sections and to develop the main compositions, or dimensions, of a target. This preliminary process reduces noise and creates a more diversified but integrated structure of a morphological matrix. Second, *hyperlinks* are informative descriptors, which includes domain-specific information highly related to dimensions. They serve as a great supplementing source of these corresponding dimensions; however, *hyperlinks* often cannot cover all associated concepts of a certain dimension. In order to ensure comprehensiveness and relativeness in value development, nouns or noun phrases in the contents are extracted using text analysis. Each concept is thoroughly examined based on an experts' discussion. Third, *categories* are used to distinguish the hierarchies of concepts. The use of a structured information may significantly increase the number of dimensions and values, thereby making the matrix with a much richer information.

Table 6-2 Summary of applicability to dimensions and values

	Advantage	Disadvantage	Applicability to dimensions	Applicability to values
table of contents	- structured information - information diversity	- inconsistency - limited contextual breadth	very likely	unlikely
hyperlinks	- conceptual significance - conceptual relevance	- excessive description - non-hierarchical structure	likely	very likely
categories	- hierarchical structures - contextual depth	- excessive amount - complex structure	very likely	very likely

6.4 Research Framework

The article proposes two models for Wikipedia-based approach of morphology building process: basic model and extended model. As shown in Figure 6-6 and Figure 6-7, the basic model is composed of preliminary, dimension development, and value development phase, whereas the extended model takes the identical procedure with an additional sub-dimension development phase. Basic model is aimed at building morphological matrix in a more concise and intuitive manner, whereas extended model is aimed at building in a more systematic and extensive manner.

6.4.1 Basic Model

The framework of basic model is illustrated in Figure 6-6. The model adopts two of the abovementioned features of Wikipedia: *table of contents* and *hyperlinks*. *Table of contents* is used for developing dimensions, and *hyperlinks* and contents are used for developing values.

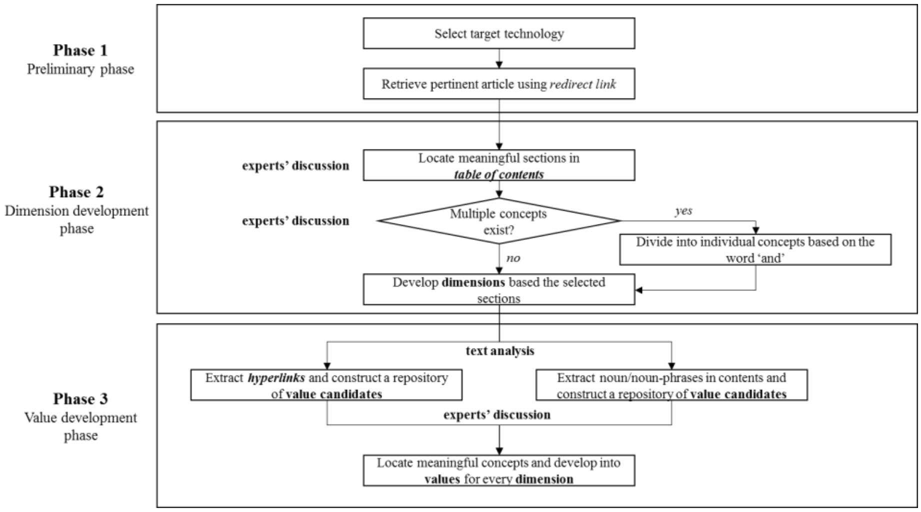


Figure 6-6 Overall framework of basic model

In preliminary phase, a target subject is selected and a proper name of the target subject is determined based on *redirect links* and *disambiguation pages*. In dimension development phase, *table of contents* is utilized. The overall structure of the target is first investigated through section titles. The sections that contain main composition information are located for dimension development. Due to its inconsistent and incoherent nature, however, a supplementary experts' discussion is required to subjectively select meaningful sections for the investigation. Second, a systematic process of dividing section titles is needed. For the purpose of assigning single concept in each dimension, the sections containing multiple concepts are automatically divided into single concept based on the word “and”. The resulting concepts are consequently assigned as dimensions. In value development phase, both *hyperlinks* and contents are used. Text analysis is conducted to extract all *hyperlinks* and nouns/noun phrases of each section, which consequently constitute a repository of value candidates. Once all associated concepts are collected, a thorough discussion is involved to select valuable terms from the repository.

6.4.2 Extended Model

The framework of extended model is illustrated in Figure 6-7. As indicated by its name, the model takes dimension and value generation process one step further from basic model by incorporating *categories*. The information of hierarchical structure allows the supplement of superordinate and subordinate concepts, thereby expanding breadth and depth of the morphological matrix. Moreover, the process involves multiple iterative processes (Eriksson & Ritchey, 2002). It is vital to continuously identify and further eliminate inappropriate components for the matrix development. As marked by shaded triangles and dotted arrows in Figure 6-7, several in-depth discussions are conducted to repeatedly modify the constituting dimensions and values. To offer a better conceptual understanding of our proposed approach, this sub-section describes the process with more details.

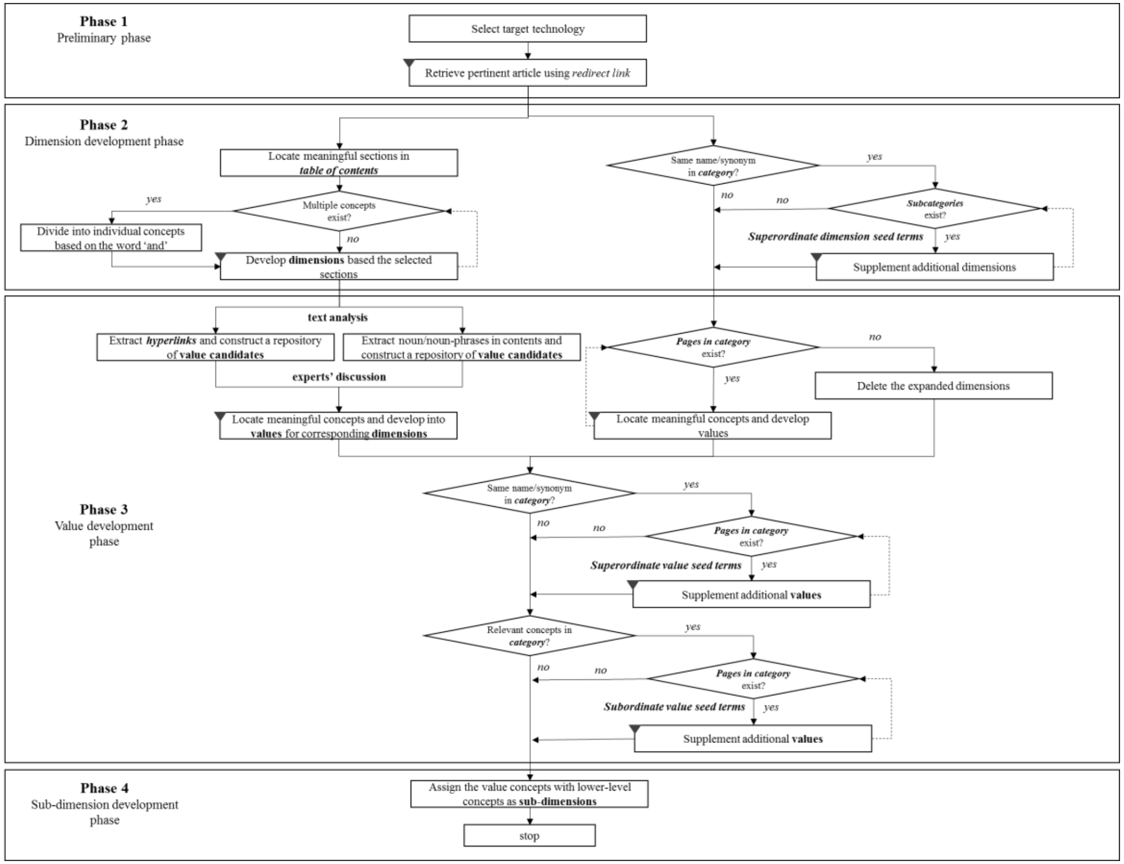


Figure 6-7 Overall framework of extended model

6.4.2.1 Phase 1: Preliminary Phase

1) Select a target subject and retrieve pertinent article

The target subject is first selected. As long as the subject is the main source of an innovation, the selection is highly versatile. It could be a technology, a service, or a company. For instance, emerging technologies are technological products, which are at the brink of commercialization. Because they are highly uncertain and are believed to hold substantial opportunities, emerging technology is a prominent example of an innovation source. This same philosophy could be applied to promising mobile application services or new enterprises. The approach, however, is critically depended on the availability and richness of relevant Wikipedia article. In order to avoid confusion and to maximize data quality, a proper article must be chosen on the basis of the lists provided in *redirect links* and *disambiguation pages*.

6.4.2.2 Phase 2: Dimension Development Phase

1) Extract the interested sections

Depending on the purpose, the composition of the subject will be completely different. For instance, if the interest lies in technological innovation of an emerging technology, the morphological matrix is constructed mainly using technological terms, rather than economic or social terms. To achieve this and for further analysis, we must look into the structure of the article based on *table of contents* and disregard information of history and regulation. Doing so increases the focus of idea generation because we are able to combine information only with that of which we are interested.

2) Develop basic dimensions of morphological matrix

Based on the premise that the selected sections represent the main structure of the target subject, each section title is assigned as a dimension of the morphological matrix. An iterative group discussion is required to include only the key sections and to ensure the integration of those dimensions form a coherent whole of the target. Moreover, a systematic process of assigning single concept in each dimension is needed. Some of the sections in *table of contents* comprise multiple concepts due to their contextual relatedness, like fuel and propulsion technologies in the “car” article. The word “and” is a great indicator for separating two noun/noun clauses, thus enabling a more automated matrix development.

3) Locate seed terms and expand dimensions

This step is aimed at two purposes: increasing the quantity of dimensions and developing a more structured matrix. *Categories*, *subcategories*, and *pages in categories* are accommodated since they yield information regarding the hierarchical relationships of concepts in Wikipedia. This is where seed terms come into play. The word “seed” indicates the concept’s capability of reproducing new individual concepts. Each concept in the target subject’s *category* list is examined to determine whether there exists a synonym, or whether the subject itself is defined as one of the target’s *categories*. The article of “car”, for instance, has a synonym of “automobile”, rather than “car”, as shown in Figure 6-5. If the target subject is confirmed to hold a *category* structure, numerous lower-level concepts from *subcategories* are extracted and further developed as supplementary dimensions. This is possible since *subcategories* offer various concepts from one step below the target subject, which are at a similar level as dimensions. The chosen concept, like "automobile", are noted as *superordinate dimension seed terms*, meaning they are superordinate concept capable of providing new subsidiary concepts. In order to supplement the most suitable concepts, a discussion session is repeatedly conducted.

6.4.2.3 Phase 3: Value Development Phase

1) Develop basic values of morphological matrix

The selection of meaningful values is vital in morphological matrix since creative ideas cannot be derived from a small amount of irrelevant and meaningless ingredients. The values, thus, must be assigned with maximized relatedness and diversity. *Hyperlinks* could be served as a great remedy since they hold a wide variety of domain-specific and informative descriptors highly related to dimensions. There exist, however, two major issues when developing values solely based on *hyperlinks*: an inclusion of excessive information and an omission of key information. Due to the inconsistent nature of article structures, they may discard some of the important concepts or, at the same time, contain too much unnecessary details. The "seating and body style" section, for example, encompasses the contents regarding the types of automobile seats and the types of body styles; however, only the key information concerning body style is expressed in the forms of *hyperlinks*. In response, all nouns/noun phrases within the contents must be extracted using text analysis. Such a process is required to prevent from excluding the essentials. Based on the concepts derived from *hyperlinks* and contents, a repository of value candidates is constructed, and an iterative experts' discussion is further performed to develop meaningful and diverse value concepts in the matrix.

2) *Develop values of expanded dimensions*

The value development is further achieved for the dimensions generated from *subcategories*. *Pages in categories* is particularly used in this step since it is capable of offering all concepts under certain dimension concepts. As in the case of Figure 5, numerous values, such as “alloy wheel”, “coupe”, “car hacking”, and “car pool”, are generated for the expanded dimensions.

3) Locate seed terms and expand values

Similar to the dimension expansion, this step is also aimed at two purposes: increasing the quantity of values and developing a more structured matrix. Value seed terms are identified to expand the number of values based on *categories* and *pages in categories*. These seed terms fall into two groups: *superordinate value seed term* and *subordinate value seed term*, as illustrated in Figure 6-8. *Superordinate value seed term* is necessary for expanding values based on high hierarchy value concepts, and *subordinate value seed term* is necessary for expanding based on low hierarchy value concepts. These two seed terms are sequentially utilized to capture both depth and breadth of matrix development.

Among the values derived from the previous step, certain value concepts that are capable of supplementing new subsidiary concepts are first identified. In details, each value concept is examined whether the concept itself is defined as one of the concept’s *categories*. This is achieved through tracing same category name or synonyms in the *category* list of the concept, as shown in Figure 6-4(a) and Figure 6-8(a). The use of *redirect pages* and *disambiguation* *page* is vital when dealing with the concepts derived from the contents because a new proper article must be first defined in order to obtain the right *categories* list. Once these value concepts are confirmed to hold a *category* structure, numerous lower-level concepts under the value concepts are extracted from *pages in categories* and further developed as supplementary values of the morphological matrix. This process is identical to that of dimension expansion. An iterative experts' discussion is also necessary to exclude the irrelevant concepts. The chosen value concepts that successfully meet these conditions are noted as *superordinate value seed terms*, meaning they are superordinate concept capable of providing new subsidiary concepts.

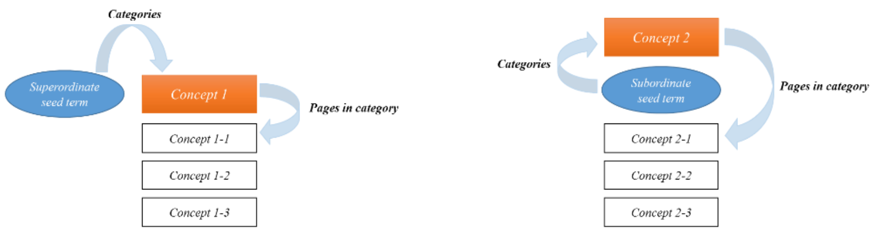


Figure 6-8(a) Value expansion process using *superordinate seed term* Figure 6-8(b) Value expansion process using *subordinate seed term*

Figure 6-8 Value expansion process using seed terms

However, the sole use of *superordinate seed terms* cannot quite cover all associated concepts since some values themselves are too low in the hierarchy. *Subordinate seed terms* are used to cope with such an issue. Certain value concepts are considered subordinate terms serving to locate alternative superordinate concepts, capable of providing additional low-level concepts. As shown in Figure 6-4(b) and Figure 6-8(b), the most suitable higher-level concept is first identified from the list of *categories* of the value concept. The use of *redirect links*, *disambiguation page*, and an iterative experts' discussion are vital for the concept selection. After confirming the availability of *pages in categories*, numerous lower-level concepts are extracted and further developed as the values of the morphological matrix. Such a value expansion process allows the retrieval of disparate information sources. By way of illustration, when the contents of “seating and body style” section is text analyzed, the concept “seat” is extracted from the contents. Then, the concept of “auto parts” can be used as the value seed term, which is selected from the *categories* of “car seat” article, to generate value concepts. As a result, the concepts of “bench seat”, “bucket seat”, “jump seat”, “rumble seat” and “third row seating” can be extracted from “*pages in categories*”. The intermediary concepts, like “seat”, are noted as *subordinate value seed terms*, meaning they are subsidiary concept capable of providing a key superordinate concept, like “auto parts”.

6.4.2.4 Phase 4: Sub-dimension Development Phase

1) Develop sub-dimensions

The structure of the matrix has been significantly extended from the value expansion process. Nonetheless, not only each dimension holds an excessive number of value concepts but multiple sub-hierarchies often exist in a single dimension. This is primarily caused by hierarchical gap between dimensions and values. In response, an intermediary concept of sub-dimension is introduced to bridge such a gap and adjust the matrix structure. All concepts that are capable of providing low-level concepts, specifically superordinate seed terms and the concepts derived from subordinate seed terms, are assigned as sub-dimensions. These concepts are presented with orange boxes in Figure 6-4 and 6-8.

6.5 Illustrative Case Study: Drone Technology

The process of our proposed methodology is demonstrated with an illustrative case study of an emerging technology–UAV, or so-called drone technology. As stated earlier, emerging technology is a perfect example for performing a creative idea generation process since it holds high level of both uncertainty and opportunity (Kwon et al., 2016). By coherently structuring and creatively recombining the components, new forms of technologies with novel functions or different application uses could possibly be derived. Drone technology, in particular, is considered a new wave of technology giant. It enables a whole new level of innovation by altering the existing business landscape, including delivery and recreation industries. In this section, Wikipedia article of drone technology is investigated, and a morphological matrix is systematically constructed on the basis of associated information. The case study is conducted in a consecutive manner: the development of basic model continued by that of extended model.

6.5.1 Basic Model

Drone technology is selected for the target subject. The analysis begins with retrieving the article of interest from Wikipedia website. The “drone” article regarding robotic vehicle of “unmanned aerial vehicle” is identified and retrieved from its *redirect page*. Then, dimension development phase is conducted using *table of contents*. Every section and sub-section of *table of contents* is thoroughly investigated and selected for dimension development of morphological matrix. In this particular case study, technical configuration of drone technology was of interest since such an ill-defined technology is in its embryonic stage with unclear technological standards and specifications. The sub-sections of “body”, “power supply and platform”, “computing”, “sensors”, “actuators”, “software”, “loop principles”, “flight controls”, and “communications” and the sections of “autonomy” and “functions” contained structured and fruitful technical information. They were considered as our basic dimensions of morphological matrix. Based on the term “and”, the dimension of “power supply and platform” is separated into two individual dimensions: “power supply” and “platform”. The next was value development phase. Basic model generated values using the concepts extracted from *hyperlinks* and contents. Text analysis was carried out to extract the noun/noun phrases in the contents. A group discussion was further conducted to select the meaningful concepts from the repository. The group comprised three experts having experience in technology forecasting and management from four years to five years and two experts having experience in mechatronics from three years to four years. A representative result of the morphological matrix is shown in Table 6-3. The full result is presented in Appendix A, where *hyperlinks* are noted in blue and nouns/noun phrases extracted from the contents are noted in black.

Table 6-3 Illustration of morphological matrix (basic model)

Dimensions	**power supply**	**platform**	**computing**	**sensors**	**actuators**	**communications**	**autonomy**
Values	lithium polymer batteries	battery elimination circuitry	system on a chip	exteroceptive sensors	actuators	radio frequency front-end	hierarchical control systems
		microcontroller unit	single board computers	non-coopertive sensors	digital electronic speed controllers	antenna	scripting language
			flight controller	collision avoidance	engines	analog-to-digital converter	finite-state machine
			flgiht controller board	gyroscopes	propellers	data link	behavior trees
				accelerometers	servomotors	telemetry	hierarchical task planners
				inertial measurement unit	weapons	satellite navigation	PID controller

compass	payload actuators	radio transmitter	motion planning
barometer	LEDs	ground control station	tree searches
GPS receiver	speakers	wearable devices	genetic algorithms
		electroencephalography	self-level
		uplink	altitude hold
		downlink	hover hold
		real-time video	headless mode
		downstream	care-free
		smartphone	take-off
		tablet	autoland
		computer	aerobatics
		human movement recognition	camera

6.5.2 Extended Model

The following model is extended from the result of the basic model. Based on the premise that the matrix is constructed as shown in Figure 6-4, the first part of Extended model is identifying *superordinate dimension seed terms* and expanding the number of dimensions by supplementing subsidiary values. The concept of “unmanned aerial vehicle” was located in the *categories* list of the article, and it was further identified that the article included several *subcategories* concepts. Among the many, only “unmanned aerial vehicle manufacturers” was considered meaningful and adequate enough to become the supplementary dimension. As conducted in the basic model, experts’ discussions were involved throughout the analysis.

The second part is expanding the number of values. The values of the expanded dimensions are first derived based on *pages in categories*. In this particular case example, the concepts were derived only from the dimension of “unmanned aerial vehicle manufacturers”, as shown in the far right column of Table 6-5. *Superordinate value seed terms* are then involved to expand the values based on high-hierarchy concepts. Each existing value concept was examined whether it be capable of supplementing lower-level concepts. The modified matrix on the basis of *superordinate value seed terms* is partially illustrated in Table 6-5 and fully presented in Appendix C. Given the space constraints, only four or five additional values were included in the result, and the rest were noted as "etc.". The supplemented values from s*uperordinate value seed terms* are noted in orange in Appendix C.

Table 6-4 List of seed terms (*hyperlinks*) of drone technology

superordinate dimension seed terms	unmanned aerial vehicle manufacturers
superordinate value seed terms	lithium polymer batteries, microcontroller unit, system on a chip, single board computers, gyroscopes, accelerometers, actuators, propellers, LEDs, aircraft controls, air brake, antenna, telemetry, electroencephalography, scripting language, finite-state machine, tree searches, genetic algorithms
subordinate value seed terms	lithium polymer batteries, microcontroller unit, compass, GPS receiver, digital electronic speed, controllers, LEDs, open loop, closed loop, plane flight dynamics, analog-to-digital converter, data link, satellite navigation, radio transmitter, hierarchical task planners, motion planning

This morphological matrix is constructed by generating additional values from the result of basic model. The resulting *superordinate value seed terms* are listed in Table 6-4. By way of illustration, the value concepts of "accelerometers" and "propellers" from "sensors" and "actuators" dimensions satisfy two conditions. They are the concepts that have same *category* names in the *categories* list and have numerous subsidiary concepts within *pages in categories*. These two seed terms resulted the augmentation of multiple concepts, including "gravimeter", "laser accelerometer", "liquid capacitive inclinometer", "modular propeller", "scimitar propeller", "cyclorotor", and "contra-rotating propellers".

Table 6-5 Illustration of modified morphological matrix using superordinate seed values (extended model)

Dimensions	power supply	sensors	actuators	communications	autonomy	manufacturers
Values	lithium polymer batteries	exteroceptive sensors	actuators	radio frequency front-end	hierarchical control systems	3D Robotics
		non-coopertive sensors	helical band actuator	antenna	scripting language	AeroDreams
		collision avoidance	linear actuator	loop antenna	ActivePerl	DJI
		gyroscopes	plasma actuator	antenna boresight	AMPL	Flirtey
		anti-rolling gyro	rigid chain actuator	batwing antenna	HyperTalk	Sky-Watch
		control moment gyroscope	etc.	array gain	Rexx	
		fibre optic gyroscope	digital electronic speed controllers	dipole antenna	etc.	
		quantum gyroscope	engines	halo antenna	finite-state machine	
		etc.	propellers	spiral antenna	alternating finite automation	
		accelerometers	modular propeller	ec.	asymmetric numeral systems	
		gravimeter	scimitar propeller	analog-to-digital converter	Buchi automation	
		laser accelerometer	cyclorotor	data link	Quotient automaton	
		liquid capacitive inclinometers	contra-rotating propellers	telemetry	etc.	
		PIGA accelerometer	etc.	EMR telemetry	behavior trees	
		etc.	servomotors	electronic data capture	hierarchical task planners	
		inertial measurement unit	weapons	remote data capture	PID controller	
		compass	payload actuators	remote terminal unit	motion planning	
		barometer	LEDs	wildlife radio telemetry	tree searches	
		GPS receiver	LED circuit	etc.	and-or tree	
			LED lamp	satellite navigation	k-ary tree	
			LED strip light	radio transmitter	suffix tree	
			etc.	ground control station	radix tree	
			speakers	wearable devices	trace tree	
				electroencephalography	etc.	
				beta wave	genetic algorithms	
				spike-and-wave	cultural algorithm	
				PGO waves	Fitness function	

evoked potential	genetic fuzzy systems
ear-eeg	genetic programming
etc.	etc.
uplink	self-level
downlink	altitude hold
real-time video	hover hold
downstream	headless mode
smartphone	care-free
tablet	take-off
computer	autoland
human movement recognition	aerobatics
	camera
	digital camera
	remote camera
	still camera
	video camera
	etc.

The last part of value expansion involves the use of *subordinate value seed terms*. Each existing value concept was examined to determine whether it is capable of detecting a higher-level concept, which could supply additional subsidiaries. This step is useful for expanding values based on relatively low-hierarchy concepts. The modified matrix using *subordinate value seed terms* is partially illustrated in Table 6-6 and fully presented in Appendix D. Given the space constraints, only four or five additional values were included in the result. The supplemented values from *subordinate value seed terms* are noted in red in Appendix D. The resulting *subordinate value seed terms* are also listed in Table 6-4.

To illustrate, the value concept of "radio transmitter" in "communications" dimension is quite low in hierarchy to have subsidiary concepts; however, it may be included in a meaningful high-level concept like "telecommunication equipment", which could further provide multiple subsidiary concepts, including "block upconverter", "hybrid coil", "radio spectrum scope", and "optical line termination". A higher-level concept is located in the categories list of "radio transmitter", and the subsidiary concepts are extracted from the *pages in categories* of that higher-level concept. However, there are some instances where a certain value concept could become both *superordinate* and *subordinate value seed terms*. The value concept of "LEDs", for instance, may provide numerous subsidiary concepts, like "LED circuit", "LED lamp" and "LED strip light"; at the same time, it may yield the higher-level concept of "optic diodes" to supplement the concepts of "crystal LED", "flexible OLED", and "phosphorescent OLED".

Table 6-6 Illustration of modified morphological matrix using subordinate seed values (extended model)

Dimensions	power supply	sensors	actuators	communications	autonomy	manufacturers
Values	lithium ion batteries	exteroceptive sensors	engines	radio electronics	hierarchical control systems	3D Robotics
	lithium polymer batteries	non-coopertive sensors	servomotors	radio frequency front-end	scripting language	AeroDreams
	solid-state lithium-ion battery	automotive safety technologies	weapons	RF power margin	ActivePerl	DJI
	18650 battery	collision avoidance	payload actuators	image response	AMPL	Flirtey
	dual carbon battery	drive by wire	speakers	feed line	AngelScript	Sky-Watch
	graphene foam	child safety lock	actuators	RF probe	HyperTalk	
	lithium hybrid organic battery	shock absorber	helical band actuator	transceiver	Rc	
	lithium ion manganese oxide	crosswind stabilization	linear actuator	radio frequency front-end	Rexx	
	lithium iron phosphate battery	gyroscopes	plasma actuator	antenna	finite-state machine	
	lithium-air battery	anti-rolling gyro	rigid chain actuator	loop antenna	alternating finite automation	
	lithium-sulfur battery	control moment gyroscope	propellers	antenna boresight	asymmetric numeral systems	
	lithium-titanate battery	fibre optic gyroscope	modular propeller	batwing antenna	Buchi automation	
	lithium-ion flow battery	quantum gyroscope	scimitar propeller	array gain	Krohn-Rhodes theory	
	nanoball batteries	accelerometers	cyclorotor	dipole antenna	Quotient automaton	
		gravimeter	contra-rotating propellers	halo antenna	permutation automation	
		laser accelerometer	power electronics	spiral antenna	behavior trees	
		liquid capacitive inclinometers	digital electronic speed controllers	digital signal processing	automated planning and scheduling	
		PIGA accelerometer	commutation cell	analog-to-digital converter	hierarchical task planners	
		inertial measurement unit	gate driver	aliasing	state space planning	
		navigational equipment	magnetic amplifier	bandlimiting	partial-order planning	
		compass	power module	infinite impulse response	kinodynamic planning	
		inertial navigation system	power semiconductor device	oversampling	multi-agent planning	
		pressure reference system	optic diodes	half-band filter	reactive planning	
		ecompass	LEDs	data transmission	PID controller	
		transfer alignment	LED circuit	data link	robot kinematics	
		barometer	LED lamp	adaptive aqualizer	motion planning	
		global positioning system	LED strip light	backward channel	kinodynamic planning	
		GPS receiver	crystal LED	bandwidth cap	kinematic chain	
		clock drift	flexible OLED	narrative traffic	articulated robot	
		digital anchor	phosphorescent OLED	parity bit	passive dynamics	
		vehicle tracking system	superluminescent diode	telemetry	serial manipulator	

positioning system
pseudorange

EMR telemetry
electronic data capture
remote data capture
remote terminal unit
wildlife radio telemetry
satellite navigation system
satellite navigation
timation
total electron content
vehicle tracking system
hybrid positioning system
automatic vehicle location
telecommunication equipment
radio transmitter
block upconverter
hybrid coil
radio spectrum scope
network termination
optical line termination
ground control station
wearable devices
electroencephalography
beta wave
spike-and-wave
PGO waves
evoked potential
ear-eeg
uplink
downlink
real-time video
downstream
smartphone
tablet
computer
human movement recognition

tree searches
and-or tree
k-ary tree
suffix tree
radix tree
trace tree
genetic algorithms
cultural algorithm
Fitness function
genetic fuzzy systems
genetic programming
truncation selection
self-level
altitude hold
hover hold
headless mode
care-free
take-off
autoland
aerobatics
optical devices
camera
digital camera
remote camera
still camera
video camera
range imaging
Wright camera
night vision devices
laser beam profiler
head-up display
electric eye

Finally, sub-dimensions are developed to prevent hierarchical value concepts from existing in a single dimension. All concepts expanded from the value seed terms may be involved with such an issue, including *superordinate value seed terms* and the higher level concept derived from *subordinate value seed terms*. For instance, “accelerometers” and “propellers” are assigned as sub-dimensions of “sensors” and “actuators”, in respective. Moreover, “telecommunication equipment” is developed into a sub-dimension, and the *subordinate value seed term* of “radio transmitter” is assigned as one of the comprising values. All concepts that are capable of providing low-level concepts are assigned as sub-dimensions. The modified matrix of using *sub-dimensions* is partially illustrated with “communications” dimension, as shown in Table 6-7. The full result is presented in Appendix E.

Table 6-7 Illustration of modified morphological matrix using sub-dimensions (extended model)

Dimension	**communications**					
Sub-dimension	**telecommunication equipment**	**satellite navigation system**	**data transmission**	**electroencephalography**	**telemetry**	**antenna**
Value	radio transmitter	satellite navigation	data link	beta wave	EMR telemetry	loop antenna
	block upconverter	timation	adaptive aqualizer	spike-and-wave	electronic data capture	antenna boresight
	hybrid coil	total electron content	backward channel	PGO waves	remote data capture	batwing antenna
	radio spectrum scope	vehicle tracking system	bandwidth cap	evoked potential	remote terminal unit	array gain
	network termination	hybrid positioning system	narrative traffic	ear-eeg	wildlife radio telemetry	dipole antenna
	optical line termination	automatic vehicle location	parity bit			halo antenna
						spiral antenna

6.6 Comparative Analysis

6.6.1 Experimental Setup

To demonstrate the effectiveness and internal validity of the proposed approach, we have conducted a comparative analysis between the proposed approach and the existing classical approach. A new morphological matrix solely depending on a focus group discussion was conducted on drone technology. The focus group was composed of three experts having a minimum of four years of experience in technology forecasting and management and two experts having a minimum of three years of experience in mechatronics. We have followed an interactive-based nominal group discussion process since a group interactive discussion was prone to several human biases and dysfunctional behaviors (Sutton & Arnold, 2013). The process involved three sessions: (1) introduction session, (2) individual brainstorming session, and (3) interactive discussion session.

The discussion was led by a moderator. First, a short introduction was conducted to provide the background and objective of this research and to demonstrate general overview of morphological analysis. The specifics regarding the technique were presented by the experts in technology forecasting. Second, the participants were asked to individually brainstorm and generate a list of technical components that could be included in the parts of drone technology. The list was constructed for both dimensions and values. Finally, the moderator collected the result generated from each participant and further constructed an aggregated list to initiate an interactive discussion. The participants gave questions or comments of each component in order to develop a complete set of dimensions and values that all members can agree upon.

Table 6-8 Illustration of morphological matrix using discussion-based approach

Dimension	Controllers	Frames	Motors	Propellers	Batteries	Safety systems	Landing gears	Electronics speed controllers (ESC)	GPS modules	Antenna	Camera	Sensors	Communications
Value	AVR	tricopter	brushed	standard	Lithium polymer (LiPo)	Battery monitor	fixed landing gears	DYS	Glonass - latest one	loose wire whip	front-facing	thermal sensor	radio control (RC)
	PIC	quadcopter	brushless	pusher	lithium	parachute	retractable landing gears	Diatone	Unmanned Tech	helical "rubber ducky"	down-facing	sonar ranging	bluetooth
	ARM	hexacopter	coreless	plastic	lithium sulphide (LiSTM)	GPS tracking	Speading wings S1000+	Eachine	Emlid	circularly polarized cloverleaf antenna	3d camera	accelerometer	WiFi
	Pixhawk	y6	servo	carbon fiber	CoreTexRC	black box	Inspire 1	Emax	Hex Technology	linear polarized antenna	CMOS camera	gyroscope	radio frequency
	The Flip32+	octocopter	electric	Gemfan	DJI	low battery alarm	Voyager 3	Sunrise Model		circular polarized antenna	CCD camera	inertia measurement unit (IMU)	
		x8	AltiGator	Hqprop	Fatshark		Scout x4	Unmanned Tech		directional antenna	depending on latency	compass / magnetometer	
			T-Motor	APC	Gens Ace		QR X900	DJI Inspire 1		omnidirectional antenna	resolution	pressure / barometer	
			Axi	Luminier	SkyRC					duck	dynamic range	infrared	
			Roxxy	3-blade	Unmanned Tech					skew planar	IR sensitive camera		
			Dualsky	2-blade	Yuneec					cloverleaf	IR blocked camera		
										array	night camera		
										helical			
										patch			
										crosshair			
										Yagi			

6.6.2 Comparison of Results

The result is presented in Table 6-8. In terms of dimension development, the morphological matrix derived from the discussion-based approach was composed of 12 dimensions. They were the most representative and general concepts that fundamentally comprise the primary structure of a drone technology. When compared with the dimensions of Wikipedia-based approach, the proposed approach not only successfully covered every dimension of the discussion-based method but offered more specific and comprehensive concepts, as shown in Table 6-9. To illustrate, whilst the classical approach generated a single dimension of "communications", the Wikipedia-based approach offered more specific component-related dimensions, including "telecommunication equipment", "digital signal processing", "satellite navigation system", "data transmission", "electroencephalography", "telemetry", etc. Moreover, Wikipedia-based methodology offered wider ranging concepts, such as "optic diodes", "software", "loop principles", "flight controls", "robot kinematics", "scripting language", and so on. The generation of both dimensions and sub-dimensions is expected to increase the number of specified and extensive range of concepts.

Table 6-9 Comparison of dimensions between discussion-based approach and Wikipedia-based approach

Discussion-based approach	**Wikipedia-based approach**	
Dimension	**Sub-dimensions**	**Dimensions**
controllers	embedded systems, microtechnology	platform, computing
frames	rotorcraft	body
motors	actuators	actuators
propellers	propellers	actuators
batteries	lithium ion batteries	power supply

safety systems	automotive safety technologies	sensors
landing gears	aircraft landing systems	autonomy
electronics speed controllers (esc)	power electronics	actuators
gps modules	global positioning system	sensors
antenna	antenna	communications
camera	optical devices	autonomy
sensors	automotive safety technologies, gyroscopes, accelerometers, navigational equipment, global positioning system	sensors
communications	telecommunication equipment, satellite navigation system, data transmission, digital signal processing, electroencephalography, telemetry, antenna, radio communications stubs	communications

In terms of value development, the overall value structure of discussion-based approach was simple and concise. Compared to the result of Wikipedia-based approach, the method offered more domain-specific and generalized value information. For example, discussion-based method provided the value concepts regarding widely-accepted types, compositions, and brands of drone “propellers”. There were “standard”, “pusher”, “3-blade”, or “2-blade” types of propellers, which are generally composed of either “plastic” or “carbon fiber”. Such a domain-specific input data is capable of generating relatively reliable ideas without much of noise. Nonetheless, it was difficult to expect any unanticipated and novel results since the input itself was composed of known and obvious information.

The effect of Wikipedia is quite different from that of the discussion-based approach. The most noticeable advantage was that the number of values were considerably greater, compared to that of the conventional method. Furthermore, the major strengths of using Wikipedia data were of two kinds: (1) diversity and (2) specificity. In terms of diversity, the proposed approach offered innovation-related and interdisciplinary information, while ensuring generalized and widely-used information. For instance, we have identified that “lithium polymer batteries” is a *subordinate value seed term*, which could derive a higher-level concept of “lithium ion batteries”. Even though lithium polymer battery (Li-Po) is a common battery type of drone technology, there were numerous other variations of the lithium-ion batteries could potentially be applied in drone technology. Based on the seed term of “lithium ion batteries”, several meaningful value concepts of “nanowire battery”, “thin film lithium-ion battery”, and “graphene foam” were identified. “Nanowire battery” is not commercially available; however, they are considered one of the promising technologies hoping to replace traditional graphite anode. “Thin film lithium-ion battery” and “graphene foam” can be applied to nano- or micro-drones based on its small size and flexible nature. Moreover, the use of “electroencephalography”, which was noted as “human brain waves” in the article, is an unconventional way of communicating and further controlling hobby drones. When such unprecedented and novel concepts are included, much diverse and innovative ideas will result from a reconfiguration process.

Furthermore, the approach occasionally generated the information from pertinent but different domain. For example, the sub-dimension of “automotive safety technologies” was extracted based on *subordinate value seed term* of “collision avoidance”. Despite its general use in field of autonomous cars, the value concepts like “shock absorber”, “drive by wire”, and “child safety lock” could be effectively used and further integrated in drone technology. Other value concepts include “aerodynamics” and “aircraft landing systems” from avionics; “finite-state machine”, “robot kinematics”, and “digital signal processing” from electronics and robotics; and “fibre optic gyroscope” and “ring laser gyroscope” from satellite missile industry. Such an off-domain information may spark creativity by providing solutions from disparate disciplines, thereby stimulating a more interdisciplinary development of a drone technology.

In terms of specificity, the proposed approach not only provided more product-related information but also offered specific component- or logic-related information of corresponding dimensions. By way of illustration, the dimension of “controller” in discussion-based approach is composed of very rudimentary product-related values, such as “AVR”, “PIC”, “ARM”, “Pixhawk”, etc. These value concepts were the most commonly used microcontrollers when building drones. However, the sub-dimensions of “embedded systems” and “microtechnology”, which were developed on the basis of seed values in the proposed methodology, generated an innumerable number of controlling products while covering each and every value concept in discussion-based approach. The values of “AVR2” and “Arduino” from Wikipedia-based method corresponded to “AVR” from discussion-based method; “PIC microcontroller” corresponded to “PIC”; “STM8” and “STM32” corresponded to “ARM”, “Pixhawk” corresponded to “PX4 autopilot”. Additionally, the proposed approach further produced more specified value concepts: language-related information of “embedded c”, “embedded java”, “Ada”, “Lua”, “BasicX”, etc.; system-related information of “slugs”, “low-voltage detect”, “priority inversion”, etc.; component-related information of “oscillator start-up timer”, “sensor node”, “interdigital transducer”, etc. Such an increase of values concepts may help technology designers to think outside the box. For example, the values of “remote terminal unit” and “wildlife radio telemetry” in the sub-dimension of “telemetry” may suggest novel application areas, such as wildlife monitoring, mine sites, or swimming pool control.

6.7 Intrinsic Limitations of Applying Wikipedia

Along with aforementioned benefits, our findings in this article are subject to at least three intrinsic limitations. The most critical shortfall lies in the fact that the process is highly dependent on the availability and richness of relevant Wikipedia articles. The cells of the morphological matrix are developed based on the contents existing in the selected page. However, the articles regarding new and emerging issues may hold a very small amount of, or even no, information; whereas the ones regarding matured and established issues may hold an excessive amount of information. For example, the article of “car” hold an abundance of contents; whereas the article of “e-textiles” hold insufficient amount of information to even develop the dimensions. The inclusion of a proper amount of information is vital. To cope with such an issue, the right model must be selected depending on the quality of the article: basic model for a matured subject and extended model for an emerging subject. Moreover, if the article contains only little content, the process can rely only on the development of value development and sub-dimension development. Since the derived sub-dimensions could substitute the role of dimensions, the extended model could skip dimension development and dimension expansion parts and start with text analyzing *hyperlinks* and contents. If the article is missing, other substitutable subject matter must be thoroughly sought.

Second, we have highlighted the importance of an expanded knowledge pool of Wikipedia data and its role in making the matrix more fruitful; however, this could be pointed out as the limitation of using Wikipedia. Because the contents of Wikipedia are edited by everyone−not just experts but also anonymous novices−the reliability and accuracy of a crowd-sourced knowledge source can be questioned (Lavsa et al., 2011). Moreover, the problem regarding context inconsistency must be highlighted. Self-looping problems, for instance, exist given the random usage of *hyperlinks*, and no standardized structures of *table of contents* exist in Wikipedia. Such issues may result the omission of valuable information and the inclusion of unnecessary information in dimension and value development processes. To deal with, we have conducted a text analysis to extract noun/noun phrases directly from the contents and have incorporated focus group discussion sessions. Yet, a more systematic and sophisticated procedure is necessary for a general use of the proposed approach. For instance, the information within the contents and *table of contents* should be exploited more effectively when expanding the dimensions and values.

Finally, the proposed approach cannot fully replace human intervention. Every step from generating dimensions to expanding values requires expert judgments, as illustrated in Figure 7. To better reconcile different opinions and refine the structure of morphological matrix, other existing idea generation techniques dealing with extracting knowledge from experts, including Delphi (Dalkey, 1969), brainstorming (Osborn, 1957), or brainwriting (VanGundy, 1981), must be incorporated. These structured methodologies are useful in supporting decision making process and enhance group creativity. An integration with a web-based software could also be useful in promoting shared understanding among experts and stakeholders (Zec et al., 2015; Zec & Matthes, 2017). In addition, a more sophisticated text mining techniques could be integrated to minimize human involvement. Such an incorporation could be particularly useful in combining the values and generating ideas, which are the most challenging steps of morphological analysis. The proposed approach accommodated only the basic techniques of text analysis: part-of-speech tagging and extraction of terms. However, if the concepts' semantic relationships are identified based on their co-occurrence measures, relatively logical and reasonable combinations could be derived. Furthermore, a visualization-based idea generation process will be possible on the basis of these measures, which could better stimulate communication between the participants and support decision-making process (Veryzer & Borja de Mozota, 2005).

6.8 Conclusion

Morphological analysis is deemed appropriate and necessary to generate new creative ideas given its objective, impersonal, and systematic nature. Conventionally, the process was only conducted by a handful number of experts and stakeholders. The participants had to be assembled in a certain place and be instructed to subjectively decompose a target matter into multiple components. Despite its many advantages, the approach was conducted with the input source subject to a limited spectrum of biased knowledge bases. Furthermore, such a knowledge pool seems quite unsuitable considering a highly complex and uncertain nature of today's matters. This research, in response, proposes a Wikipedia-based approach to the development of morphological matrix, which could be served as a supporting and supplementary tool for generating creative ideas. The fundamental premise is that creative ideas are crafted from a novel combination of existing but previously separated ideas. Wikipedia data source was incorporated considering its case-specific information and well-coordinated knowledge structure. Two models are presented: basic model and extended model. Basic model was aimed at constructing the morphological matrix in a more concise and intuitive manner; whereas extended model was aimed at constructing in a more systematic and extensive manner.

As a result, the proposed approach generated more specified and diversified morphological matrix, and three major intrinsic features of Wikipedia have been identified as major contributing factors: extended knowledge pool, case-specific knowledge content, and well-coordinated knowledge structure. First, Wikipedia platform enabled individuals from different background and expertise to participate in defining and structuring concepts. It has obtained and further converged a variety of perspectives to minimize subjectivity. Second, Wikipedia offered an abundance of case-specific knowledge that could better stimulate specialized and practical idea generation. Finally, Wikipedia had a well-coordinated knowledge structure that enabled the systematic decomposition of a certain target. In details, *table of contents* allowed overviewing the structure of certain subject; *hyperlinks* allowed extracting the most useful and relevant information; and *categories* allowed understanding complex hierarchical structure of the concepts. However, the current study still cannot replace classical expert rounds in creative idea generation. The proposed approach is rather a preliminary or exploratory attempt to apply other valuable knowledge sources in morphological analysis, thereby increasing the possibility of generating more innovative and novel ideas. If the debate is to be moved forward, further experimental investigations regarding more systematic and effective procedures of knowledge extraction and creative component combination are needed.

Chapter 7

Concluding Remarks

This dissertation has endeavored to formalize the notion of data-driven technology foresight – a systematic process of envisioning future technology developments and their impacts on our society based on a large-scale data analysis. Given the collective and inclusive nature of technology foresight, we have greatly expanded the knowledge boundary horizon by incorporating distinct data sources, like *future-oriented web data*, Wikipedia data, and scientific publication data, and have further developed theoretical and methodological frameworks for the proactive management process of emerging technologies. The process comprises impact identification, impact analysis, plan development, and technology ideation.

As mentioned earlier, this research is aimed at addressing the following questions: (1) What kinds of data sources are available on the web and which of those are considered useful in foresight studies? (2) Where could we incorporate these data sources and which techniques are most suitable for the given purposes? (3) Which foresight-related fields would particularly benefit from applying a data-driven approach and what exactly are the positive effects? The primary contribution of this dissertation lies in the preliminary attempt to gain insights into the utilization of different data sources and mining techniques in the field of technology foresight. In particular, each study has made the following contributions for theory and practice, as listed in Table 7-1. The detailed implications of each chapter are also summarized in Table 7-2.

Table 7-1 Theoretical and practical implication of this dissertation

	Theoretical Implication	**Practical Implication**
OVERALL DISSERTATION	· A preliminary and exploratory attempt to use an unprecedented method in envisioning the future of new technologies · Contributes to the foundation of futures and foresight research	· Stimulates a better user-acceptance of new and emerging technologies
Chapter 3. ***Foresight for Impact Identification***	· Contributes to the field of technology assessment by offering a more quantitative method of exploring and assessing new technologies, while efficiently incorporating the voices of the public	· Relieves public skepticism from the emergence of unfamiliar technologies · Supports policymakers to foresee unprecedented issues arising from new technologies and to design more balanced regulations between tech companies and potential customers
Chapter 4. ***Foresight for Impact Analysis***	· Contributes to the field of scenario studies by proposing a supplementary way of quantitatively encapsulating an extended range of voices, not only general but latent perspectives, when building scenarios	· Overlooks a wide range of perspectives, including both the many and the few, regarding the future of emerging technologies to promote a more responsive management of new technologies
Chapter 5. ***Foresight for Plan Development***	· Contributes to the field of responsible development by yielding an alternative and detailed guidance of conducting responsible development	· Offers a guidance or a clear-cut methodology was in need for engineers to develop them with some sort of responsibility · Enables new technologies to be yield only the right impacts to our society, by resolving the issues of rejection and non-acceptance of unfamiliar technologies
Chapter 6. ***Foresight for Technology Ideation***	· Contributes to the field of idea generation by applying a new valuable knowledge source in morphological analysis and demonstrating its effectiveness in providing insightful information	· Increases the possibility of generating more innovative and novel ideas associated with emerging technologies

Chapter 3. Foresight for Impact Identification

A data-driven technology foresight methodology was proposed for identifying and understanding the holistic overview of future societal consequences arising from the use of emerging technologies. In theoretical point of view, this research contributes to the field of technology assessment by offering a more quantitative method of exploring and assessing new technologies, while efficiently incorporating the voices of the public. It particularly solves two of the intrinsic issues within technology assessment: (1) encapsulating the voices of the general public in a foresight activity and (2) exploring futures in a social context while minimizing the involvement of human biases. In practice, it can be served as an essential tool for those seeking for a wide-ranging use of new and emerging technologies. It reduces the distrust of technology's unfamiliarity, thereby stimulating a better user acceptance. Furthermore, it may support policymakers to foresee unprecedented issues arising from new technologies and to design more balanced regulations between tech companies and potential customers.

Chapter 4. Foresight for Impact Analysis

A data-driven technology foresight methodology was proposed for examining the details of future impacts associated with emerging technologies. The methodology aimed at envisioning far-reaching societal consequences for our society through analyzing texts within *future-oriented web data*. In theoretical point of view, this research contributes to the field of scenario studies by proposing a supplementary way of quantitatively encapsulating an extended range of voices, not only general but latent perspectives, when building scenarios. It particularly seeks to address following limitations of previous studies: (1) a difficulty in ensuring both plausibility and unexpectedness within the scenarios and (2) a high reliance on the participatory way of building scenarios. The proposed approach is one of the preliminary attempts of building scenarios in a quantitative manner. An immediate improvement may seem small in scenario planning community; however, our result may open the possibility of incorporating futuristic concern-related voices from the web and applying text mining techniques into exploring the holistic overview of emerging technology's future uncertainties. In practice, it enables the overlook of a wide range of perspectives, including both the many and the few, regarding the future of emerging technologies to promote a more responsive management of new technologies.

Chapter 5. Foresight for Plan Development

A data-driven technology foresight methodology was proposed for providing a technical guidance to develop emerging technologies to better cope with future impacts. In theoretical point of view, this research contributes to the field of responsible development by yielding an alternative and detailed guidance of conducting responsible development. It particularly addresses the following questions: (1) how would the notion of responsible development be put into practice? and (2) how could we develop a methodology of efficiently capturing the perspectives of the general public and of scholars. When it comes to the responsible development of emerging technologies, there has been a prime gap between philosophizing and existing. In practice, the proposed process offers a guidance or a clear-cut methodology was in need for engineers to consider social concerns and human norms alongside more typical engineering ideals and develop them with some sort of responsibility. Furthermore, it enables new technologies to yield only the right impacts to our society, by resolving the issues of rejection and non-acceptance of unfamiliar technologies.

Chapter 6. Foresight for Technology Ideation

A data-driven technology foresight methodology was proposed for generating ideas regarding the future structures of emerging technologies, which may yield significant impetus for innovative product development. The proposed models reveal that the incorporation of Wikipedia into morphological analysis may promote a more innovative and creative idea generation. In theoretical point of view, this research contributes to the field of idea generation. Two major drawbacks have been addressed in this research: (1) involvement of human subjectivity in idea generation and (2) incorporation of unsuitable data sources considering highly complex and uncertain nature of today's matters. The proposed models have expanded the applied data set to a new boundary and showed its effectiveness in yielding insightful and future-oriented information related to emerging technologies.

However, we have only scratched the surface of the shift towards data-driven technology foresight, and there still remains a substantial amount of work to be done for the development of a stronger theoretical and methodological foundation. This dissertation highlights three notable limitations: (1) limited application areas, (2) missing evaluation process, and (3) lacking methodological sophistication. First, the range of application areas is quite limited. Technology foresight aims at a systematic observation and investigation of future technology developments and their interaction with society and environment (Yazan 2016). However, this research focuses solely on the negative implications and neglects the positive aspects of new technologies. An application in different futures research, particularly in the fields of new product development, new service development, and opportunity identification, remains to be seen for the future work.

Table 7-2 Detailed implication of each chapter

Chapter	Detailed Implication
Chapter 3. ***Foresight for Impact Identification***	· Provides information regarding direct impacts to the individuals, such as feelings, values, or states · Enables an understanding of a more comprehensive concepts of social impacts · Copes with intangibility - observed the situations that have not yet been occurred · Copes with complexity - takes unexpected involvement of players into account
Chapter 4. ***Foresight for Impact Analysis***	· Offers a new method of supporting a scenario building process based on keyword networks · Acquires generalized voice from the majority and this ensures plausibility in building scenarios · Acquires latent voice from the few and this ensures unexpectedness and creativity in building scenarios
Chapter 5. ***Foresight for Plan Development***	· Yields insightful details of impact scenarios, including human values and detailed information of unforeseen consequences · Provides technical guidelines or recommendations, specifying potential requirements that must be highlighted in the early design process · Identifies technical solutions of inter- or trans-disciplinary collaboration between engineering, social sciences, and applied ethics
Chapter 6. ***Foresight for Technology Ideation***	· Ensures diversity in morphology development by offering innovation-related and interdisciplinary knowledge while ensuring generalized information · Ensures specificity in morphology development by offering not only more product-related information but also specific component- or logic- related information of corresponding dimensions

Second, considering our applied data sources, an evaluation process seeking to assess validity, accuracy, and reliability of this research is essential. Both internal and external validity must be increased by comparing with existing literature in terms of both results and methodologies. Furthermore, the accuracy and reliability are also of paramount importance to increase the quality of the research. The notion of technology foresight was introduced based on a skepticism concerning the accuracy and utility of forecasting (Martin, 2010) and there is no common evaluation exercise in foresight practice. However, this research has relied on an abundance of unreliable data sources, like *future-oriented web data* and Wikipedia data. In fact, several studies highlight the significance of accuracy in short-term foresight activities. Further work needs to be done to analyze the reliability and realization of our results either by light self-evaluation or by fully-fledged formal evaluation involving a panel of diverse experts.

Third, existing text analysis techniques were adopted for the research frameworks without a major modification of the original process, and this has caused the involvement of human judgment in the proposed methodologies. Thus, more advanced and sophisticated methodologies must be designed to cope with human intervention and to extract a more insight information from the aforementioned data sources. Three major parts must be particularly modified: preprocessing, interpretation, and data collection processes. LSA, for instance, requires a much of subjectivity in selecting society- and technology-related terms for the input of actual analysis and in interpreting the resulting term clusters to develop future scenarios; IdeaGraph involves manual decision makings when filtering out unnecessary terms to extract latent structures and when setting a proper threshold values to construct impact scenarios.

In response, lexical databases, like WordNet, SentiWordNet, Affective Norms for English Words (ANEW), may be useful to build a more systematic preprocessing procedure for analyzing societal aspects of future technologies.

More advanced natural language processing techniques, like machine learning or deep linguistic analysis, can also be applied for clear and logical interpretation of the results. Furthermore, a systematic and detailed method of selecting proper search queries is required for this research to be applied to a more general use. This dissertation has demonstrated the potential of *future-oriented web data* and Wikipedia data as a new data sources useful in technology foresight activities. However, the proposed methodologies ought to be adjusted more radically depending on the research objectives to minimize subjectivity and to stimulate a more insightful knowledge in foresight activities.

Taken together, this dissertation is exploratory and open-ended. It is designed to determine the nature of the problem, rather than to offer definitive and conclusive answers. We cannot expect a drastic improvement in accuracy, reliability, quantity, creativity, or any other performances related to futures studies. Nor can we properly validate the results of this research. However, this research has observed a mere potential of a more quantitative approach in performing technology foresight. A data-driven technology foresight could be central to understanding uncertain and complex future of emerging technologies, and the results offered by this dissertation may well provide not just a rationale or justification but also a theoretical grounding for this newly introduced notion. This dissertation will hopefully provide a foothold for the readers to better comprehend and act on this new shift in the field of technology foresight.

Bibliography

Alcock, R. E., & Busby, J. (2006). Risk migration and scientific advance: the case of flame-retardant compounds. *Risk Analysis, 26*(2), 369-381.

Allahyari, M., & Kochut, K. (2016). Semantic tagging using topic models exploiting wikipedia category network. Proceedings of the 2016 IEEE Tenth International Conference on Semantic Computing (ICSC), Laguna Hills, California.

Amer, M., Daim, T. U., & Jetter, A. (2013). A review of scenario planning. *Futures*, *46*, 23-40.

Andersen, P. D., Jørgensen, B. H., Lading, L., & Rasmussen, B. (2004). Sensor foresight—technology and market. *Technovation, 24*(4), 311-320.

Andronis, C., Sharma, A., Virvilis, V., Deftereos, S., & Persidis, A. (2011). Literature mining, ontologies and information visualization for drug repurposing. *Briefings in Bioinformatics*, *12*(4), 357-368.

Arcos-Novillo, D. A., & Güemes-Castorena, D. (2017). Development of an additive manufacturing technology scenario for opportunity identification—The case of Mexico. *Futures*, *90*, 1-15.

Assmuth, T., Hildén, M., & Benighaus, C. (2010). Integrated risk assessment and risk governance as socio-political phenomena: a synthetic view of the challenges. *Science of the Total Environment, 408*(18), 3943-3953.

Ayres, R. U. (1969). *Technological forecasting and long-range planning*. New York: McGraw-Hill Book Company.

Ballard, S. C., & Hall, T. A. (1984). Theory and practice of integrated impact assessment: the case of the western energy study. *Technological Forecasting and Social Change, 25*(1), 37-48.

Bañuls, V. A., & Salmeron, J. L. (2008). Foresighting key areas in the information technology industry. *Technovation, 28*(3), 103-111.

Banuls, V. A., & Salmeron, J. L. (2011). Scope and design issues in foresight support systems. *International Journal of Foresight and Innovation Policy, 7*(4), 338-351.

Barré, R., & Keenan, M. (2008). Revisiting foresight rationales: what lessons from the social sciences and humanities?. In Cargnin, C., Keenan, M., Johnston, R., Scapolo, F., & Barré, R. (Eds.), Future-Oriented Technology Analysis (pp. 41-52): Springer Berlin Heidelberg.

Becker, H. A., & Sanders, K. (2006). Innovations in meta-analysis and social impact analysis relevant for tech mining. *Technological Forecasting and Social Change, 73*(8), 966-980.

Berry, M. W., Dumais, S. T., & O'Brien, G. W. (1995). Using linear algebra for intelligent information retrieval. *SIAM Review, 37*(4), 573-595.

Björk, J., & Magnusson, M. (2009). Where do good innovation ideas come from? Exploring the influence of network connectivity on innovation idea quality. *Journal of Production Innovation Management*, 26(6), 662-670.

Blake, R., & Ayyagari, R. (2012). Analyzing information systems security research to find key topics, trends, and opportunities. *Journal of Information Privacy and Security, 8*(3), 37-67.

Blok, V., & Lemmens, P. (2015). The emerging concept of responsible innovation. Three reasons why it is questionable and calls for a radical transformation of the concept of innovation. In B. Koops, I. Oosterlaken, H. Romijn, T.Swierstra, & J. van den Hoven (Eds.), Responsible Innovation 2: Concepts, apporaches, and Applications (pp.19-35). Basel: Springer International Publishing.

Boden, M.A. (2004). The creative mind: Myths and mechanisms. New York, NY: Routledge Press.

Borup, M., Brown, N., Konrad, K., & Van Lente, H. (2006). The sociology of expectations in science and technology. *Technology Analysis & Strategic Management*, *18*(3-4), 285-298.

Bradfield, R., Wright, G., Burt, G., Cairns, G., & Van Der Heijden, K. (2005). The origins and evolution of scenario techniques in long range business planning. *Futures*, *37*(8), 795-812.

Brem, A., & Bilgram, V. (2015). The search for innovative partners in co-creation: Identifying lead users in social media through netnography and crowdsourcing. *Journal of Engineering and Technology Management*, *37*, 40-51.

Britt, M. A., Wiemer-Hastings, P. M., Larson, A. A., & Perfetti, C. A. (2004). Using intelligent feedback to improve sourcing and integration in students' essays. *International Journal of Artificial Intelligence in Education, 14*(3-4), 359-374.

Buljan, M., & Nemet, Z. (2014). Latent Semantic Indexing. Text Analysis and Retrieval 2014 (Project Report). Croatia: University of Zagreb.

Bunescu, R.C., & Pasca, M. (2006). Using encyclopedic knowledge for named entity disambiguation. Proceedings of the 11th European Chapter of the Association for Computational Linguistics, Trento, Italy.

Burdge, R. J., & Vanclay, F. (1995). Social impact assessment. In F. Vanclay, D. Bronstein (Ed.), Environmental and Social Impact Assessment (pp. 31-65). Chichester: Wiley.

Callahan, E. S., & Herring, S. C. (2011). Cultural bias in Wikipedia content on famous persons. *Journal of the Association for Information Science and Technology*, *62*(10), 1899-1915.

Canongia, C., Antunes, A., & Pereira, M. D. N. F. (2004). Technological foresight—the use of biotechnology in the development of new drugs against breast cancer. *Technovation, 24*(4), 299-309.

Carleton, T., & Cockayne, W. (2009). The power of prototypes in foresight engineering. Proceedings of the 17th International Conference on Engineering Design, Palo Alto, CA, USA.

Carlsen, H., Dreborg, K., Godman, M., Hansson, S. O., Johansson, L., &

Wikman-Svahn, P. (2010). Assessing socially disruptive technological change. *Technology in Society, 32*(3), 209-218.

Carlsen, H., Johansson, L., Wikman-Svahn, P., & Dreborg, K. H. (2014). Co-evolutionary scenarios for creative prototyping of future robot systems for civil protection. *Technological Forecasting and Social Change, 84*, 93-100.

Cetindamar, D., Phaal, R., & Probert, D. R. (2016). Technology management as a profession and the challenges ahead. *Journal of Engineering and Technology Management*, *41*, 1-13.

Chen, K., Jarboe, K., & Wolfe, J. (1981). Long-range scenario construction for technology assessment. *Technological Forecasting and Social Change, 20*(1), 27-40.

Chermack, T. J., Lynham, S. A., & Ruona, W. E. (2001). A review of scenario planning literature. *Futures Research Quarterly*, *17*(2), 7-32.

Chiu, T. F., Hong, C. F., & Chiu, Y. T. (2008). Visualization of financial trends using chance discovery methods. Proceedings of the International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems, Berlin, Heidelberg.

Ciglan, M., Nørvåg, K. (2010). WikiPop: personalized event detection system based on Wikipedia page view statistics. Proceedings of the 19th ACM International Conference on Information and Knowledge Management, Toronto, Canada.

Coates, J. F. (1974). Some methods and techniques for comprehensive impact assessment. *Technological Forecasting and Social Change, 6*, 341-357.

Coates, V., Farooque, M., Klavans, R., Lapid, K., Linstone, H. A., Pistorius, C., & Porter, A. L. (2001). On the future of technological forecasting. *Technological Forecasting and Social Change*, *67*(1), 1-17.

Collingridge, D. (1980). The social control of technology. Michigan, USA: Open University Press.

Coussement, K., & Van den Poel, D. (2008). Integrating the voice of customers through call center emails into a decision support system for churn prediction. *Information & Management, 45*(3), 164-174.

Cuhls, K. (2003). From forecasting to foresight processes—new participative foresight activities in Germany. *Journal of Forecasting, 22*(2-3), 93-111.

Cuhls, K. (2003). From forecasting to foresight processes—new participative foresight activities in Germany. *Journal of Forecasting*, *22*(2-3), 93-111.

Cummings, M. L. (2006). Integrating ethics in design through the value-sensitive design approach. *Science and Engineering Ethics*, *12*(4), 701-715.

Dalkey, N. (1969). The Delphi method: An experimental study of group opinion, Rand Corporation, Santa Monica, CA.

Day, G. S., Schoemaker, P. J., & Gunther, R. E. (2004). *Wharton on managing emerging technologies*: New York, NY: John Wiley & Sons.

De Melo, G., & Siersdorfer, S. (2007). Multilingual text classification using ontologies. Proceedings of the 29th European Conference on Information Retrieval, Rome, Italy.

De Spiegeleire, S., van Duijne, F., & Chivot, E. (2016). Towards Foresight 3.0: The HCSS Metafore Approach—A Multilingual Approach for Exploring Global Foresights. In T. U. Daim, D. Chiavetta, A. Porter, and O. Saritas (Eds.), Anticipating Future Innovation Pathways Through Large Data Analysis. (pp. 99-117): Springer International Publishing, Switzerland.

Deerwester, S. C., Dumais, S. T., Landauer, T. K., Furnas, G. W., & Harshman, R. A. (1990). Indexing by latent semantic analysis. *Journal of the American Society for Information Science, 41*(6), 391-407.

Denning, T., Kohno, T., & Levy, H. M. (2013). Computer security and the modern home. *Communications of the ACM*, *56*(1), 94-103.

Denning, T., Kramer, D. B., Friedman, B., Reynolds, M. R., Gill, B., & Kohno, T. (2014). CPS: Beyond usability: Applying value sensitive design based methods to investigate domain characteristics for security for implantable cardiac devices. Proceedings of the 30th Annual Computer Security Applications Conference, New York, USA.

Dickson, P. H., & Weaver, K. M. (1997). Environmental determinants and individual-level moderators of alliance use. *Academy of Management Journal*, *40*(2), 404-425.

Diehl, M., & Stroebe, W. (1987). Productivity loss in brainstorming groups: Toward the solution of a riddle. *Journal of Personality and Social Psychology*, *53*(3), 497-509.

Dosi, G. (1982). Technological paradigms and technological trajectories: a suggested interpretation of the determinants and directions of technical change, *Research Policy*, *11*(3), 147-162.

Drucker, P. (1985). Innovation and entrepreneurship: practice and principles. Oxford, UK: Butterworth-Heinemann Press.

Dumais, S. T. (1994). Latent semantic indexing (LSI) and TREC-2. Proceedings of the Second Text Retrieval Conference (TREC-2), Gaithersburg, MD.

Dumais, S. T., Furnas, G. W., Landauer, T. K., Deerwester, S., & Harshman, R. (1988). Using latent semantic analysis to improve access to textual information. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. New York, USA.

Durance, P., & Godet, M. (2010). Scenario building: Uses and abuses. *Technological Forecasting and Social Change*, *77*(9), 1488-1492.

Durance, P., & Godet, M. (2010). Scenario building: Uses and abuses. *Technological Forecasting and Social Change*, *77*(9), 1488-1492.

Eaton, W., Wright, W., Whyte, K., Gasteyer, S. P., & Gehrke, P. J. (2014). Engagement and uncertainty: emerging technologies challenge the work of engagement. *Journal of Higher Education Outreach and Engagement*, *18*(2), 151-178.

Ehrenfeld, J. (2008). Sustainability by design: a subversive strategy for transforming our consumer culture. New Haven, CT: Yale University Press.

Eisenhardt, K. M. (1989). Building theories from case study research. *Academy of Management Review*, *14*(4), 532-550.

Eriksson, T., & Ritchey, T. (2002). Scenario development using computerised morphological analysis, Proceedings of Winchester International Conference, Winchester, England.

Evangelista, R., Sandven, T., Sirilli, G., & Smith, K. (1998). Measuring innovation in European industry. *International Journal of Economics and Business Research*, 5(3), 311-333.

Evangelopoulos, N., Zhang, X., & Prybutok, V. R. (2012). Latent semantic analysis: five methodological recommendations. *European Journal of Information Systems*, *21*(1), 70-86.

Featherston, C. R., Ho, J., Brevignon-Dodlin, L., & O'Sullivan E. (2016). Mediating and catalysing innovation: A framework for anticipating the standardisation needs of emerging technologies. *Technovation*, *48*, 25-40.

Firat, A. K., Woon, W. L. & Madnick, S. (2008). Technological forecasting–A review (CISL Working paper 2008-15). Cambridge, MA: Massachusetts Institute of Technology.

Fischhoff, B., Slovic, P., & Lichtenstein, S. (1982). Lay foibles and expert fables in judgments about risk. *The American Statistician*, *36*(3b), 240-255.

Flanagan, M., & Nissenbaum, H. (2014). Values at play in digital games. Cambridge, UK: MIT Press.

Forrester, J. W. (1997). Industrial dynamics. *Journal of the Operational Research Society*, *48*(10), 1037-1041.

Forsythe, G. E., Moler, C. B., & Malcolm, M. A. (1977). Computer methods for mathematical computations. Englewook Cliffs, NY: Prentice-Hall.

Franks, D. M., & Cohen, T. (2012). Social Licence in Design: Constructive technology assessment within a mineral research and development institution. *Technological Forecasting and Social Change*, *79*(7), 1229-1240.

Frewer, L. (1999). Risk perception, social trust, and public participation in strategic decision making: Implications for emerging technologies, *Ambio*, *28*(6), 569-574.

Friedman, B., Kahn, P., & Borning, A. (2002). Value sensitive design: Theory and methods (UW CSE Technical Report 02-12-01). University of Washington, Seattle.

Furnas, G. W., Deerwester, S., Dumais, S. T., Landauer, T. K., Harshman, R. A., Streeter, L. A., & Lochbaum, K. E. (1988). Information retrieval using a singular value decomposition model of latent semantic structure. Proceedings of the 11th International Conference on Research and Development in Information Retrieval, New York, NY.

Garechana, G., Rio, R., Cilleruelo, E., & Gavilanes, J. (2012). Tracking the evolution of waste recycling research using overlay maps of science. *Waste Management*, *32*(6), 1069-1074.

Gausemeier, J., Fink, A., & Schlake, O. (1998). Scenario management: An approach to develop future potentials. *Technological Forecasting and Social Change*, *59*(2), 111-130.

Geschka, H. (1983). Creativity techniques in product planning and development: a view from West Germany. *R&D Management*, *13*(3), 169-183.

Geum, Y., & Park, Y. (2016). How to generate creative ideas for innovation: a hybrid approach of WordNet and morphological analysis. *Technological Forecasting & Social Change*, 111, 176-187.

Gheorghiu, R., Curaj, A., Paunica, M., & Holeab, C. (2009). Web 2.0 and the emergence of future oriented communities. *Economic Computation and Economic Cybernetics Studies and Research*, *43*(2), 117-127.

Girotra, K., Terwiesch, C., & Ulrich, K. T. (2010). Idea generation and the quality of the best idea. *Management Science*, *56*(4), 591-605.

Gordon, T. J., & Hayward, H. (1968). Initial experiments with the cross impact matrix method of forecasting. *Futures*, *1*(2), 100-116.

Gordon, T., & Pease, A. (2006). RT Delphi: An efficient, "round-less" almost real time Delphi method. *Technological Forecasting and Social Change*, *73*(4), 321-333.

Gray, P., & Hovav, A. (2008). From hindsight to foresight: Applying futures research techniques in information systems. *Communications of the Association for Information Systems*, *22*(1), 12.

Grunwald, A. (2011). Responsible innovation: bringing together technology assessment, applied ethics, and STS research. *Enterprise and Work Innovation Studies*, *31*, 10-9.

Grunwald, A., & Achternbosch, M. (2013). Technology Assessment and Approaches to Early Engagement. In N. Doorn, D. Schuurbiers, I. van de Poel, & M. E. Gorman (Eds.), Early engagement and new technologies: Opening up the laboratory (pp. 15-34): Springer, Dordrecht.

Halaweh, M. (2013). Emerging technology: what is it. *Journal of Technology Management and Innovation*, *8*(3), 108-115.

Han, K., & Shin, J. (2014). A systematic way of identifying and forecasting technological reverse salients using QFD, bibliometrics, and trend impact analysis: A carbon nanotube biosensor case. *Technovation*, *34*(9), 559-570.

Harper, J. C. (2013). Impact of technology foresight (Nesta Working Paper 13/16). Manchester, UK: Manchester Institute of Innovation Research University of Manchester.

Hassan, M. M., Karray, F., & Kamel, M.S. (2012). Automatic document topic identification using wikipedia hierarchical ontology. Proceedings of 11th International Conference on Information Science, Signal Processing and their Applications, Montreal, QC.

Hassanzadeh, A., Namdarian, L., Majidpour, M., & Elahi, S.B. (2015) Developing a model to evaluate the impacts of science, technology and innovation foresight on policy-making. *Technology Analysis & Strategic Management*, *27*(4), 437-460.

Healy, M. L., Dahlben, L. J., & Isaacs, J. A. (2008). Environmental assessment of single-walled carbon nanotube processes. *Journal of Industrial Ecology*, *12*(3), 376-393.

Hernández-Ortega, B. (2011). The role of post-use trust in the acceptance of a technology: drivers and consequences. *Technovation*, *31*(10), 523-538.

Hiltunen, E. (2008). Good sources of weak signals: a global study of where futurists look for weak signals. *Journal of Futures Studies*, *12*(4), 21-44.

Hofmann, T. (2001). Unsupervised learning by probabilistic latent semantic analysis. *Machine Learning*, *42*(1), 177-196.

Hu, X., Zhang, X., Lu, C., Park, E. K., & Zhou, X. (2009) Exploiting Wikipedia as external knowledge for document clustering. Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Paris, France.

Huang, S. H., Liu, P., Mokasdar, A., & Hou, L. (2013). Additive manufacturing and its societal impact: a literature review. *The International Journal of Advanced Manufacturing Technology*, *67*(5-8), 1-13.

Hueske, A. K., Endrikat, J., & Guenther, E. (2015). External environment, the innovating organization, and its individuals: A multilevel model for identifying innovation barriers accounting for social uncertainties. *Journal of Engineering and Technology Management*, *35*, 45-70.

Hung, S.-C., & Chu, Y.-Y. (2006). Stimulating new industries from emerging technologies: challenges for the public sector. *Technovation*, *26*(1), 104-110.

Islam, N., & Miyazaki, K. (2010). An empirical analysis of nanotechnology research domains. *Technovation*, *30*(4), 229-237.

Ittipanuvat, V., Fujita, K., Sakata, I., & Kajikawa, Y. (2014). Finding linkage between technology and social issue: A Literature Based Discovery approach. *Journal of Engineering and Technology Management*, *32*, 160-184.

Ivanov, A., & Cyr, D. (2014). Satisfaction with outcome and process from web-based meetings for idea generation and selection: The roles of instrumentality, enjoyment, and interface design. *Telematics and Informatics*, *31*(4), 543-558.

Jones, H. (1976). Morphology and creativity in technological forecasting. *R&D Management*, *6*(3), 125-130.

Joorabchi, A., English, M., & Mahdi, A.E. (2015). Automatic mapping of user tags to Wikipedia concepts: The case of a Q&A website–StackOverflow. *Journal of Information Science*, 1-4.

Kajikawa, Y., Ohno, J., Takeda, Y., Matsushima, K., & Komiyama, H. (2007). Creating an academic landscape of sustainability science: an analysis of the citation network. *Sustainability Science*, 2(2), 221.

Kakkonen, T., Myller, N., Timonen, J., & Sutinen, E. (2005). Automatic essay grading with probabilistic latent semantic analysis. Proceedings of the Second Workshop on Building Educational Applications Using NLP, Ann Arbor, USA.

Kamprath, M., & Mietzner, D. (2015). The impact of sectoral changes on individual competences: A reflective scenario-based approach in the creative industries. *Technological Forecasting and Social Change*, *95*, 252-275.

Keller, J., & von der Gracht, A. (2014). The influence of information and communication technology (ICT) on future foresight processes—results from a delphi survey. *Technological Forecasting and Social Change*, *85*, 81-92.

Keller, J., Markmann, C., & von der Gracht, A. (2015). Foresight support systems to facilitate regional innovations: A conceptualization case for a German logistics cluster. *Technological Forecasting and Social Change*, *97*, 15-28.

Keller, K. H. (2008). From here to there in information technology: the complexities of innovation. *American Behavioral Scientist*, *52*(1), 97-106.

Klavans, R. & Boyack, K. W. (2006). Identifying a better measure of relatedness for mapping science. *Journal of the American Society for Information Science and Technology*, *57*(2), 251-263.

Köhler, A. R., & Som, C., (2014). Risk preventative innovation strategies for emerging technologies the cases of nano-textiles and smart textiles. *Technovation*, *34*(8), 420-430.

Koivisto, R., Wessberg, N., Eerola, A., Ahlqvist, T., Kivisaari, S., Myllyoja, J., & Halonen, M. (2009). Integrating future-oriented technology analysis and risk assessment methodologies. *Technological Forecasting and Social Change*, *76*(9), 1163-1176.

Kontostathis, A., & Pottenger, W. M. (2006). A framework for understanding Latent Semantic Indexing (LSI) performance. *Information Processing & Management*, *42*(1), 56-73.

Kwon, H., Kim, J., & Park, Y., (2017). Applying LSA text mining technique in envisioning social impacts of emerging technologies: The case of drone technology. *Technovation 60-61*, 15-28.

Kwon, H., Park, Y., & Geum, Y. (2016). How to Generate Ideas for Emerging Technologies using Future-oriented Data. Proceedings of ISPIM Innovation Symposium (the International Society for Professional Innovation Management), Boston, USA.

Landauer, T. K., Foltz, P. W., & Laham, D. (1998). An introduction to latent semantic analysis. *Discourse Processes*, *25*(2-3), 259-284.

Lavsa, S. M., Corman, S. L., Culley, C. M., & Pummer, T. L. (2011). Reliability of Wikipedia as a medication information source for pharmacy students. *Currents in Pharmacy Teaching and Learning*, *3*(2), 154-158.

Lee, C., Seol, H., & Park, Y. (2007). Identifying new IT-based service concepts based on the technological strength: A text mining and morphology analysis approach. Proceedings of the Fourth International Conference on Fuzzy Systems and Knowledge Discovery, Haikou, China.

Lee, S., Baker, J., Song, J., & Wetherbe, J. C. (2010). An empirical comparison of four text mining methods. Proceedings of Hawaii International Conference on System Sciences (HICSS), Honolulu, HI.

Lee, S., Song, J., & Kim, Y. (2010). An empirical comparison of four text mining methods. *Journal of Computer Information Systems*, *51*(1), 1-10.

Leenes, R., & Lucivero, F. (2014). Laws on robots, laws by robots, laws in robots: regulating robot behaviour by design. *Law, Innovation and Technology*, *6*(2), 193-220.

Lindgren, M., & Bandhold, H. (2003). Scenario planning. London, England: Palgrave Press.

Littman, M. L., Dumais, S. T., & Landauer, T. K. (1998). Automatic cross-language information retrieval using latent semantic indexing. In G. Grefenstette (Ed.), Cross-language information retrieval (pp. 51-62). Boston, MA: Springer.

Lochbaum, K. E., & Streeter, L. A. (1989). Comparing and combining the effectiveness of latent semantic indexing and the ordinary vector space model for information retrieval. *Information Processing & Management*, *25*(6), 665-676.

Luhn, H. P. (1957). A statistical approach to mechanized encoding and searching of literary information. *IBM Journal of Research and Development*, *1*(4), 309-317.

MacKay, B., & McKiernan, P. (2010). Creativity and dysfunction in strategic processes: The case of scenario planning. *Futures*, *42*(4), 271-281.

Maine, E., Thomas, V. J., & Utterback, J. (2014). Radical innovation from the confluence of technologies: Innovation management strategies for the emerging nanobiotechnology industry. *Journal of Engineering and Technology Management*, *32*, 1-25.

Maletic, J., & Marcus, A. (2000). Using latent semantic analysis to identify similarities in source code to support program understanding. Proceedings of 12th IEEE International Conference on the Tools with Artificial Intelligence. Vancouver, BC.

Manning, C. D., Raghavan, P., & Schütze, H. (2008). Matrix Decompositions and Latent Semantic Indexing. Introduction to Information Retrieval. Cambridge, England: Cambridge University Press.

Markmann, C., Darkow, I. L., & von der Gracht, H. (2013). A Delphi-based risk analysis—Identifying and assessing future challenges for supply chain security in a multi-stakeholder environment. *Technological Forecasting and Social Change*, *80*(9), 1815-1833.

Martin, B. R. (2010). The origins of the concept of 'foresight' in science and technology: An insider's perspective. *Technological Forecasting and Social Change*, *77*(9), 1438-1447.

Martino, J. P. (1993). *Technological forecasting for decision making*: McGraw-Hill, Inc.

Marx, L. (2010). Technology: The emergence of a hazardous concept. *Technology and Culture*, *51*(3), 561-577.

Matsumura, N., Matsuo, Y., Ohsawa, Y., & Ishizuka, M. (2002). Discovering emerging topics from WWW. *Journal of Contingencies and Crisis Management*, *10*(2), 73-81.

Meijer, I. S., Hekkert, M. P., & Koppenjan, J. F. (2007). The influence of perceived uncertainty on entrepreneurial action in emerging renewable energy technology; biomass gasification projects in the Netherlands. *Energy policy*, *35*(11), 5836-5854.

Mihalcea, R., & Csomai, A. (2007). Wikify!: linking documents to encyclopedic knowledge. Proceedings of the sixteenth ACM conference on Conference on Information and Knowledge Management, Lisbon, Portugal.

Miles, I. (2010). The development of technology foresight: a review. *Technological Forecasting and Social Change*, *77*(9), 1448-1456.

Miller, J. K., Friedman, B., Jancke, G., & Gill, B. (2007). Value tensions in design: the value sensitive design, development, and appropriation of a corporation's groupware system. Proceedings of the International ACM Conference on Supporting Group Work, Florida, USA.

Millett, L. I., Friedman, B., & Felten, E. (2001). Cookies and web browser design: Toward realizing informed consent online. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Washington, USA.

Milne, D. (2007). Computing semantic relatedness using wikipedia link structure. Proceedings of the New Zealand Computer Science Research Student Conference, Hamilton, New Zealand.

Nakamura, H., Suzuki, S., Sakata, I., & Kajikawa, Y. (2015). Knowledge combination modeling: The measurement of knowledge similarity between different technological domains. *Technological Forecasting and Social Change, 94*, 187-201.

Navarro, J., Hayward, P., & Voros, J. (2008). How to solve a wicked problem? furniture foresight case study. *Foresight*, *10*(2), 11-29.

Newman, N. C., Porter, A. L., Newman, D., Trumbach, C. C., & Bolan, S. D. (2014). Comparing methods to extract technical content for technological intelligence. *Journal of Engineering and Technology Management*, *32*, 97-109.

O'Connor, T. J. (1971). A methodology for analogies. *Technological Forecasting and Social Change*, *2*(3-4), 289-309.

Ohsawa, Y. (2005). Data crystallization: chance discovery extended for dealing with unobservable events. *New Mathematics and Natural Computation*, *1*(03), 373-392.

Ohsawa, Y., & Hayashi, T. (2015). Visualizing History for Qualitative Explanation of Valuable Events using Tangled String. *Procedia Computer Science*, *60*, 1178-1185.

Okazaki, N., & Ohsawa, Y. (2003). Polaris: an integrated data miner for chance discovery. Proceedings of The Third International Workshop on Chance Discovery and Its Management, Crete, Greece.

Oosterlaken, I. (2015). Applying value sensitive design (VSD) to wind turbines and wind parks: An exploration. *Science and Engineering Ethics*, *21*(2), 359-379.

Osborn, A. F. (1957) Applied imagination. Oxford, England: Scribner Press

Ostertag, K., & Hüsing, B. (2008). Identification of starting points for exposure assessment in the post-use phase of nanomaterial-containing products. *Journal of Cleaner Production*, *16*(8), 938-948.

Owen, R., Macnaghten, P., & Stilgoe, J. (2012). Responsible research and innovation: From science in society to science for society, with society. *Science and Public Policy*, *39*(6), 751-760.

Palm, E., & Hansson, S. O. (2006). The case for ethical technology assessment (eTA). *Technological Forecasting and Social Change*, *73*(5), 543-558.

Pang, A. S. K. (2010). Futures 2.0: rethinking the discipline. *Foresight*, *12*(1), 5-20.

Peterson, G. D., Cumming, G. S., & Carpenter, S. R. (2003). Scenario planning: a tool for conservation in an uncertain world. *Conservation Biology*, *17*(2), 358-366.

Phaal, R., Farrukh, C. J. P., & Probert, D. R. (2001). Technology management process assessment: a case study. *International Journal of Operations & Production Management*, *21*(8), 1116-1132.

Pidgeon, N., Harthorn, B., & Satterfield, T. (2011). Nanotechnology risk perceptions and communication: emerging technologies, emerging challenges. *Risk Analysis*, *31*(11), 1694-1700.

Pîrjan, A., & Petroşanu, D.-M. (2013). The impact of 3D printing technology on the society and economy. *Journal of Information Systems & Operations Management*, *7*(2), 360-370.

Poli, R., Healy, M., & Kameas, A. (2010). Theory and applications of ontology: Computer applications. Dordrecht, Netherlands: Springer Press.

Porter, A. L., & Cunningham, S. W. (2004). Tech mining: exploiting new technologies for competitive advantage. MA, USA: John Wiley & Sons Press.

Porter, A. L., & Roper, A. T. (1991). Forecasting and management of technology. New York, NY: John Wiley & Sons Press.

Raford, N. (2014). Online foresight platforms: evidence for their impact on scenario planning & strategic foresight. *Technological Forecasting and Social Change*, *97*, 65-76.

Rao, B., Gopi, A. G., & Maione, R. (2016). The societal impact of commercial drones. *Technology in Society*, *45*, 83-90.

Reeves, S. (2012). Envisioning ubiquitous computing. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, New York, NY.

Reger, G. (2001). Technology foresight in companies: from an indicator to a network and process perspective. *Technology Analysis & Strategic Management*, *13*(4), 533-553.

Rehder, B., Schreiner, M., Wolfe, M. B., Laham, D., Landauer, T. K., & Kintsch, W. (1998). Using latent semantic analysis to assess knowledge: some technical considerations. *Discourse Processes*, *25*(2-3), 337-354.

Ribeiro, B. E., & Quintanilla, M. A. (2015) Transitions in biofuel technologies: an appraisal of the social impacts of cellulosic ethanol using the Delphi method. *Technological Forecasting and Social Change*, *92*, 53-68.

Rietzschel, E. F., Nijstad, B. A., & Stroebe, W. (2014) Effects of problem scope and creativity instructions on idea generation and selection. *Creativity Research Journal*, *26*(2), 185-191.

Ringland, G., & Schwartz, P. P. (1998). *Scenario planning: managing for the future*: John Wiley & Sons Press.

Ritchey, T. (2005). Futures studies using morphological analysis. Adapted from an article for the UN University Millennium Project: Futures Research Methodology Series.

Ritchey, T. (2011). Modeling alternative futures with general morphological analysis. *World Futures Review*, *3*(1), 83-94.

Rittel, H. W., & Webber, M. M. (1973). Dilemmas in a general theory of planning. *Policy sciences*, *4*(2), 155-169.

Robinson, D. K. (2009). Co-evolutionary scenarios: an application to prospecting futures of the responsible development of nanotechnology. *Technological Forecasting and Social Change*, *76*(9), 1222-1239.

Rosario, B. (2000). Latent semantic indexing: An overview (Final paper INFOSIS 240). Berkley, CA: University of California, Berkley.

Rotolo, D., Hicks, D., & Martin, B. R. (2015). What is an emerging technology?. *Research Policy*, *44*(10), 1827-1843.

Ruan, Y., Hang, C. C., & Wang, Y. M. (2014). Government′ s role in disruptive innovation and industry emergence: The case of the electric bike in China. *Technovation*, *34*(12), 785-796.

Šabanović, S. (2010). Robots in society, society in robots. *International Journal of Social Robotics*, *2*(4), 439-450.

Saritas, O., & Nugroho, Y. (2012). Mapping issues and envisaging futures: An evolutionary scenario approach. *Technological Forecasting and Social Change*, *79*(3), 509-529.

Schatzmann, J., Schäfer, R., & Eichelbaum, F. (2013). Foresight 2.0-definition, overview & evaluation. *European Journal of Futures Research*, *1*(1): 1-15.

Schilling, M. A., & Green, E. (2011) Recombinant search and breakthrough idea generation: An analysis of high impact papers in the social sciences. *Research Policy*, *40*(10), 1321-1331.

Schoemaker, P. J. (2004). Forecasting and scenario planning: the challenges of uncertainty and complexity. In D. J. Koehler, & N. Harvey (Eds.) Blackwell handbook of judgment and decision making (pp. 274-296). Oxford: Blackwell Press.

Seear, K., Petersen, A., & Bowman, D. (2009). The social and economic impacts of nanotechnologies: a literature review. Victoria, Australia: Monash University Press.

Shah, J.J., Smith, S. M., & Vargas-Hernandez, N. (2003) Metrics for measuring ideation effectiveness. *Design Studies*, *24*(2), 111-134.

Shibata, N., Kajikawa, Y., Takeda, Y., & Matsushima, K. (2008). Detecting emerging research fronts based on topological measures in citation networks of scientific publications. *Technovation*, *28*(11), 758-775.

Shluzas, L. M. A., & Leifer, L. J. (2014). The insight-value-perception (iVP) model for user-centered design. *Technovation*, *34*(11), 649-662

Shurig, R. (1984). Morphology: A tool for exploring new technology. *Long Range Planning*, *17*(3), 129-140.

Sidorova, A., Evangelopoulos, N., Valacich, J. S., & Ramakrishnan, T. (2008). Uncovering the intellectual core of the information systems discipline. *MIS Quarterly*, *32*(3), 467-482.

Simons, L. P., & Verhagen, W. P. (2008). Applying value-sensitive design and quality function deployment to healthcare ICT: the case of Dutch primary care unit dossiers. *Journal of Design Research*, *7*(2), 155-176.

Simonton, D. K. (2013). What is a creative idea? Little-c versus Big-C creativity. In K. Thomas, J. Chan (Eds.), Handbook of Research Creativity, Northamption, MA: Edward Elgar Publishing Inc.

Smith, J. E., & Saritas, O. (2011). Science and technology foresight baker's dozen: a pocket primer of comparative and combined foresight methods. *Foresight*, *13*(2), 79-96.

Smith, M. R., & Marx, L. (1994). Does technology drive history?: The dilemma of technological determinism. Cambridge, England: MIT Press.

Sollie, P. (2007). Ethics, technology development and uncertainty: an outline for any future ethics of technology. *Journal of Information, Communication and Ethics in Society*, *5*(4), 293-306.

Stahl, B. C. (2011). What does the future hold? a critical view of emerging information and communication technologies and their social consequences In M. Chiasson, O. Henfridsson, H. Karsten, & J. DeGross (Eds.), Researching the Future in Information Systems (pp. 59-76): Springer Press.

Stefaniak, A. B., Hackley, V. A., Roebben, G., Ehara, K., Hankin, S., Postek, M. T., Thünemann, A. F. (2013). Nanoscale reference materials for environmental, health and safety measurements: needs, gaps and opportunities. *Nanotoxicology*, *7*(8), 1325-1337.

Stemerding, D., Swierstra, T., & Boenink, M. (2010). Exploring the interaction between technology and morality in the field of genetic susceptibility testing: a scenario study. *Futures*, *42*(10), 1133-1145.

Sternberg, R. J. (1985). Implicit theories of intelligence, creativity, and wisdom. *Journal of Personality and Social Psychology*, *49*(3), 607-627.

Stilgoe, J., Owen, R., & Macnaghten, P. (2013). Developing a framework for responsible innovation. *Research Policy*, *42*(9), 1568-1580.

Strube, M., & Ponzetto, S. P. (2006). WikiRelate! Computing semantic relatedness using Wikipedia. Proceedings of the American Association for Artificial Intelligence (AAAI), Boston, MA.

Surry, D. W., & Farquhar, J. D. (1997). Diffusion theory and instructional technology. *Journal of Instructional Science and Technology*, *2*(1), 24-36.

Sutton, S. G., & Arnold, V. (2013). Focus group methods: Using interactive and nominal groups to explore emerging technology-driven phenomena in accounting and information systems. *International Journal of Accounting Information Systems*, *14*(2), 81-88.

Swanson, D. R. (1986). Fish oil, Raynaud's syndrome, and undiscovered public knowledge. *Perspectives in Biology and Medicine*, *30*(1), 7-18.

Szlávik, Z., Tombros, A., & Lalmas, M. (2007). Feature-and query-based table of contents generation for XML documentsm. Proceedings of the 29th European Conference on Information Retrieval, Rome, Italy.

Taleb, N. N. (2010). The black swan:: The impact of the highly improbable fragility (Vol. 2). New York, NY: Random House Press.

Tarko, V., & Aligica, P. D. (2011). From “broad studies” to internet-based "expert knowledge aggregation". notes on the methodology and technology of knowledge integration. *Futures*, *43*(9), 986-995.

Timmermans, J., Zhao, Y., & van den Hoven, J. (2011). Ethics and nanopharmacy: Value sensitive design of new drugs. *Nanoethics*, *5*(3): 269-283.

Tran, T. A., & Daim, T. (2008). A taxonomic review of methods and tools applied in technology assessment. *Technological Forecasting and Social Change*, *75*(9), 1396-1405.

Tuominen, A., & Ahlqvist, T. (2010). Is the transport system becoming ubiquitous? socio-technical roadmapping as a tool for integrating the development of transport policies and intelligent transport systems and services in Finland. *Technological Forecasting and Social Change*, *77*(1), 120-134.

van de Poel I. (2015). Design for Values. In P. Kwalec & R. P. Wierzchoslwski (Eds.), Social Responsibility and Science in Innovation Economy (pp. 115-164): Wydawnictwo KUL, Lublin.

Van Den Ende, J., Mulder, K., Knot, M., Moors, E., & Vergragt, P. (1998). Traditional and modern technology assessment: toward a toolkit. *Technological Forecasting and Social Change*, *58*(1), 5-21.

van der Heijden, K. (2000). Scenarios and forecasting: two perspectives. *Technological Forecasting and Social Change*, *65*(1), 31-36.

van Merkerk, R. O., & Smits, R. E. (2008). Tailoring CTA for emerging technologies. *Technological Forecasting and Social Change*, 75(3), 312-333.

Vanclay, F. (2002). Conceptualising social impacts. *Environmental Impact Assessment Review*, *22*(3), 183-211.

VanGundy, A. B. (1981). Techniques of structured problem solving. Van Nostrand Reinhold, New York: Springer Press.

Veryzer, R.W., & Borja de Mozota, B. (2005). The impact of user-oriented design on new product development: An examination of fundamental relationships. *Journal of Production Innovation Management, 22*(2), 128-143.

von der Gracht, A., & Stillings, C. (2013). An innovation-focused scenario process—A case from the materials producing industry. *Technological Forecasting and Social Change*, *80*(4), 599-610.

von der Gracht, A., Bañuls, V. A., Turoff, M., Skulimowski, A. M., & Gordon, T. J. (2014). Foresight support systems: The future role of ICT for foresight. Te*chnological Forecasting and Social Change*, *97*, 1-6.

von der Gracht, A., Bañuls, V. A., Turoff, M., Skulimowski, A. M., & Gordon, T. J. (2015). Foresight support systems: The future role of ICT for foresight. *Technological Forecasting and Social Change*, *97*, 1-6.

Von Schomberg, R. (2013). A Vision of Responsible Research and Innovation. In R. Owen, M. Heintz, & J. Bessant (Eds.), Responsible innovation: Managing the responsible emergence of science and innovation in society. (pp. 51-74): Wiley Press, London.

Wang, C., Guo, X., & Han, H. (2012). Crime detection using latent semantic analysis and hierarchical structure. Proceedings of the 3rd International Conference on Software Engineering and Service Science, Beijing, China.

Wang, H., & Ohsawa, Y. (2013). Idea discovery: A scenario-based systematic approach for decision making in market innovation. *Expert Systems with Applications*, *40*(2), 429-438.

Wang, H., Lu, S., Zhang, C., Wang, Q., & Xu, F. (2016). Timing-ideagraph: a directed cognition graph approach for decision making based on temporal event sequences. Proceedings of 20th International Conference on the Computer Supported Cooperative Work in Design (CSCWD), Nanchang, China.

Wang, H., Ohsawa, Y., & Nishihara, Y. (2012). Innovation support system for creative product design based on chance discovery. *Expert Systems with Applications*, *39*(5), 4890-4897.

Wang, H., Xu, F., Hu, X., & Ohsawa, Y. (2013, October). Ideagraph: a graph-based algorithm of mining latent information for human cognition. Proceedings of 2013 IEEE International Conference on the Systems, Man, and Cybernetics (SMC), Manchester, UK.

Wang, P., & Domeniconi, C. (2008). Building semantic kernels for text classification using Wikipedia. Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Las Vegas, NV.

Ward, T. B. (2004). Cognition, creativity, and entrepreneurship. *Journal of Business Venturing*, *19*(2), 173-188.

Wardak, A., Gorman, M. E., Swami, N., & Deshpande, S. (2008). Identification of risks in the life cycle of nanotechnology-based products. *Journal of Industrial Ecology*, *12*(3), 435-448.

Webster, A. (1999). Technologies in transition, policies in transition: foresight in the risk society. *Technovation*, *19*(6), 413-421.

West, R., Paranjape, A., & Leskovec, J. (2015). Mining missing hyperlinks from human navigation traces: A case study of Wikipedia. Proceedings of the 24th international conference on World Wide Web, Florence, Italy.

Wild, F., Stahl, C., Stermsek, G., & Neumann, G. (2005). Parameters driving effectiveness of automated essay scoring with LSA. Proceedings of the 9th International Computer-Assisted Assessment Conference, Longhborough, UK.

Winsor, C. P. (1932). The Gompertz curve as a growth curve. Proceedings of the National Academy of Sciences, USA.

Wohlers, T., & Caffrey, T. (2013). Additive manufacturing: going mainstream. *Manufacturing Engineering*, *151*(6), 67-73.

Wright, D. (2011). A framework for the ethical impact assessment of information technology. *Ethics and Information Technology*, *13*(3), 199-226.

Xu, H., Crossler, R. E., & BéLanger, F. (2012). A value sensitive design investigation of privacy enhancing tools in web browsers. *Decision Support Systems*, *54*(1), 424-433.

Yazan, A. M. (2016). Methods Used in Future Technology Analysis and its Selection: an application to VTOL transportation system (IET Working Papers Series 1-25). Portugal: IET.

Yoon, B., Park, I., & Coh, B. Y. (2014). Exploring technological opportunities by linking technology and products: Application of morphology analysis and text mining. *Technological Forecasting and Social Change* *86*, 287-303.

Yoon, B., Phaal, R., & Probert, D. (2008). Morphology analysis for technology roadmapping: application of text mining. *R&D Management*, *38*(1), 51-68.

Zec, M., Matthes, F. (2017). Web-based software-support for collaborative morphological analysis in real-time. *Technological Forecasting and Social Change*. In press.

Zec, M., Schneider, A., & Matthes, F. (2015). Towards a process model for computer-supported collaborative morphological analysis. Proceedings of the Twenty-first Americas Conference on Information Systems, Fajardo, Puerto Rico.

Zhang, C., Wang, H., Wang, W., & Xu, F. (2014). An improved ideagraph algorithm for discovering important rare events. Proceedings of the Systems, Man and Cybernetics (SMC), San Diego, CA.

Zwicky, F. (1957). Morphological astronomy. Berlin, Germany: Springer Press.

Appendix

Appendix A Result of overt and latent structures of each impact scenario

Impact scenario	Overt structures	Latent structures	Impact scenario	Overt structures	Latent structures
1. Counterfeited money	card	fake, bezel, coordinated, debit, slots, hotel, trading, connectivity, inked, adhesives, keys, usb	6. IP problems	copyright	violation, betamax, provisions, moral
	patient			intellectual	filings, thefts, protracted, betrayal, secrets, affiliated, provisions, cgi, monopoly, repartriating
	counterfeit	authentication, luxury, marketplaces		property	filings, thefts, protracted, betrayal, secrets, affiliated, provisions, cgi, pirating
	power	crypto		rights	franchises, invention
	credit	skimmers, teller, instant		business	cyberbullying
	blueprint	possess, encrypted, liable, goods, molecule, helmet		manufacturer	untested, defective, liability, sporting
	illegal	downloads, drm, goods		trademark	filings, violation, defective
	digital	liable, circumvention		legal	betamax, defective
	atm	beknownst, keys, automatic, membership		industry	continent, demise, cyberbullying
	money	fake, bitcoin, lubricate, unemployed, poker		market	loyalty, secrets
	copyright	borrowing, inducement, issuing		licensing	franchises, liability
	economic	worldbank, crypto, ecosystem		owner	untested, inventio, sue
	cash	insurance		policy	disclosure, borrow
	currency	volatility, bitcoin		brand	loyalty, compartment
	scanner	microscopy		economic	governmental, affairs
2. Unhealthy air emission	environmental	preliminary, renewable, reefs, fauna	7. Bioprinting issues	cells	hydrogel, pluripotency, parenchymal, hepatocytes, bioink, ear, adipose, mammalian, endothelial,

	environment	rainwater, hepa
	air	currents, displace, porpoise, preheat, pdp, dbp, nacelle, filteration
	emission	burning, filteration, cigarette
	pollution	reefs, fauna
	scale	microstructure
	toxic	renewable, antimony, ventilated, lactide, ammable, polyvinyl
	particle	clogging, ventilated, filteration
	green	interstitial, foetus, hepa
	consumption	
	carnage	
	chemicals	vocs, carcinogenic, fumes
	harmful	fumes
	scale	financials
	citizen	aid
	food	sealant, utensils, pizza, pasta, crevices, tableware, toxic, nutrition, chemicals, approved, certified, grilling, meat, diapers, spacetech, cracks
	kitchen	chemicals, appliances, cutlery
	clean	oils, denatured, porous
3. Printed food	bacteria	cracks, harmful, spaces, porous, contamination, glazed
	people	nutrition, protective
	eat	plates, fruit, genetically, engineered
	safe	undermine, coatings, chemicals, crevices, cutlery

		progenitor, cancer, hipscs, cow, veterinary, cartilage, osteochondral, pulmonary
	tissue	canine, transfection, angiogenesis, bovine, decellularized
	organ	aging, extracellular, bladder, liver, kidney, transplantable, regenerative, rejection
	bioprinting	aortic, conduit, microextrusion, microenvironment, biomimetic
	medicine	veterinary, osteochondral, pulmonary, bone
	ethics	nascent, terrifying, restoration, darker, politics, crimes
	patient	stent, sternum, prescribed, cancer
	medical	
	muscle	
	fda	ear, cartilage, matured
	pharmaceutical	epilepsy, spirtam
	biomedical	
	health	obesity
	function	
	lung	vascular, kidney, cancer, asthma, strokes
	policy	disclosure, advocates
	heat	jam, deform, spacious, deflection, dissipate
	energy	recyclability, diffraction, ghg, pulsed, reliance
8. Energy hog	clean	denatured, minimizes, coal, burnout
	consumption	
	security	flaw, bash, cipher, esoteric
	power	concentrators, abrupt, motherboards, ccs, lca, unplugging, coal

	retail	employees, appliances
	business	competitiveness, tax, prognostications
	policy	competitiveness, tax, insurance, privacy, borrow, terrifying
	safety	mug, alchohol, regulators, msds, harmful
	health	debating, threaten, pregnant, chemicals, ingestion, cholestrol, bloodstream, contamination
	tax	ripple, services, online
	income	freedom, authorized
	cash	traverse, services
	legal	counterfeiters, virtual
	consumption	trade, problematic
	carnage	
4. Taxation problem	currency	counterfeiters, virtual, online
	retail	employees, shoes, appliances
	government	drought, bureaucratic, governance, stake
	economic	bioeconomic, crisis, premature
	legislation	rush, traded
	harm	fears
	old	congenital, nonprofit, drainage, untreated
	manufacturing	bioeconomic, fears, infrastructure
	financial	crisis, synthesis
5. Recyling plastics	plastic	prototypers, recyclebot, byproduct, petroleum, thermoplastic, carcinogen, acid
	medical	thermoplastic

	excessive	jam, friction, refraction, sticking
	green	interstitial, shades
	risk	flaw, healthier
	loss	currents, profits, dissipate, fertility
	economic	wasteful, recovery
	reduce	healthier
	environmental	stabilize, recovery
	ethical	
	gun	nra, untraceable, airport, penalty, enshrined, gunsmithing, rifle, homemade
	legal	uav
	firearms	nra, ray, penalty, atf, gunpowder, gunsmithing, terrorist, handgun, distribute, silencer, explosives, receivers
	weapon	semiautomatic, classficiation, ammunition, serial, rifle, nuclear, pistol
	rights	communist
9. Printed firearms	property	distribute
	security	untraceable, airport, identity, nuclear, uav, monitors, hackers
	illegal	stealth, downloads, gangs
	legislation	serial
	license	gpl, distribute
	undetectable	nra, untraceable, ray, ammunition
	act	entrepreneurially, defiance, inducer, circumvention
	policy	airport
	crime	cops, identity

biodegradable	cornstarch, tapioca, sugercane, potato, starch, polyvinyl, biocompatible, thermoplastic		ban	outright
manufacturing	bioeconomy, industrialization, stagnated		drug	digitise, regenerative, exoskeletons, bna, pill, adjunctive, recipes, glycolic, ebola, jawbone, hepatitis, ibuprofen
recycled	prototypers, recyclebot		fda	levetiracetam, jurisdiction, bna, premarket
policy	antiquated		filament	
economic	ecosystems, renewal		pharmaceutical	epilepsy, formulations, recipes
sector	bioeconomy		legal	
traditional	polyvinyl		heated	
security	sabotage		medicine	narcotics, regenerative, pill, pulmonary, bioprinter
environmental	ecosystems, fauna, coral, reefs, cod, ecology, carcinogen, habitat	10. Printed drugs	patient	prosthetis, preoperative
metal	thermoplastic, acid		safe	msds
implication			chemical	formations, reagents, microsterelithography, ecstasy, inorganic
health	occupational, bloodstreams, public		blueprint	digitise, bioprinter
policymaker			counterfeit	
			property	
			bacteria	deformed
			compound	parabolic, invisible

Appendix B Result of Wikipedia-based morphological matrix (basic model)

Dimensions	body	power supply	platform	computing	sensors	actuators	software	loop principles	flight controls	communications	autonomy
Values	tailless quadcopter	lithium polymer batteries	battery elimination circuitry	system on a chip	exteroceptive sensors	actuators	real-time	open loop	plane flight dynamics	radio frequency front-end	hierarchical control systems
	rotary wing		microcontroller unit	single board computers	non-coopertive sensors	digital electronic speed controllers	Raspberry Pis	closed loop	aircraft controls	antenna	scripting language
	mono-			flight controller	collision avoidance	engines	Beagleboards	hybrid control	air brake	analog-to-digital converter	finite-state machine
	bi-			flgiht controller board	gyroscopes	propellers	NavIO	positive	autopilot	data link	behavior trees
					accelerometers	servomotors	PXFMini	faster	helicopter flight dynamics	telemetry	hierarchical task planners
					inertial measurement unit	weapons	Nuttx	slower	multirotor flight dynamics	satellite navigation	PID controller
					compass	payload actuators	Xenomai	left		radio transmitter	motion planning
					barometer	LEDs	DDS-ROS 2.0	up		ground control station	tree searches
					GPS receiver	speakers	KKMultiCopter	down		wearable devices	genetic algorithms
							ArduCopter	feedback		electroencephalography	self-level
								tailwind		uplink	altitude hold
								feed forward		downlink	hover hold
										real-time video	headless mode
										downstream	care-free
										smartphone	take-off
										tablet	autoland
										computer	aerobatics
										human movement recognition	camera

Appendix C Result of Wikipedia-based morphological matrix using *superordinate seed terms* (extended model)

Dimensions	body	power supply	platform	computing	sensors	actuators	software	loop principles	flight controls	communications	autonomy	manufacturers
Values	tailless quadcopter	lithium polymer batteries	battery elimination circuitry	system on a chip	exteroceptive sensors	actuators	real-time	open loop	plane flight dynamics	radio frequency front-end	hierarchical control systems	3D Robotics
	rotary wing		microcontroller unit	advanced microcontoller bus architecture	non-coopertive sensors	helical band actuator	Raspberry Pis	closed loop	aircraft controls	antenna	scripting language	AeroDreams
	mono-		embedded controller	MPSoC	collision avoidance	linear actuator	Beagleboards	hybrid control	airbrake	loop antenna	ActivePerl	DJI
	bi-		Arduino	Amlogic	gyroscopes	plasma actuator	NavIO	positive	dive brake	antenna boresight	AMPL	Flirtey
			PIC microcontroller	ARC	anti-rolling gyro	rigid chain actuator	PXFMini	faster	flap	batwing antenna	AngelScript	Sky-Watch
			single-board microcontroller	Atom	control moment gyroscope	digital electronic speed controllers	Nuttx	slower	gouge flap	array gain	HyperTalk	
			Netduino	PSoC	fibre optic gyroscope	engines	Xenomai	left	air brake	dipole antenna	Rc	
				Rockchip	quantum gyroscope	propellers	DDS-ROS 2.0	up	balanced rudder	halo antenna	Rexx	
				InvenSense	accelerometers	modular propeller	KKMultiCopter	down	control loading system	spiral antenna	finite-state machine	
				single board computers	gravimeter	scimitar propeller	ArduCopter	feedback	controllable slope soaring	analog-to-digital converter	alternating finite automation	
				Banana Pi	laser accelerometer	cyclorotor		tailwind	servo tab	data link	asymmetric numeral systems	
				Kano	liquid capacitive inclinometers	contra-rotating propellers		feed forward	autopilot	telemetry	Buchi automation	
				P112	PIGA accelerometer	servomotors			helicopter flight dynamics	EMR telemetry	Krohn-Rhodes theory	
				VoCore	inertial measurement unit	weapons			multirotor flight dynamics	electronic data capture	Quotient automaton	
				flight controller	compass	payload actuators				remote data capture	permutation automation	
				flgiht controller board	barometer	LEDs				remote terminal unit	behavior trees	
					GPS receiver	LED circuit				wildlife radio telemetry	hierarchical task planners	
						LED lamp				satellite navigation	PID controller	

LED strip light	radio transmitter	motion planning
speakers	ground control station	tree searches
	wearable devices	and-or tree
	electroencephalography	k-ary tree
	beta wave	suffix tree
	spike-and-wave	radix tree
	PGO waves	trace tree
	evoked potential	genetic algorithms
	ear-eeg	cultural algorithm
	uplink	Fitness function
	downlink	genetic fuzzy systems
	real-time video	genetic programming
	downstream	truncation selection
	smartphone	self-level
	tablet	altitude hold
	computer	hover hold
	human movement recognition	headless mode
		care-free
		take-off
		autoland
		aerobatics
		camera
		digital camera
		remote camera
		still camera
		range imaging

Appendix D Result of Wikipedia-based morphological matrix after applying *subordinate value seed terms* (extended model)

Dimensions	body	power supply	platform	computing	sensors	actuators	software	loop principles	flight controls	communications	autonomy	manufacturers
Values	tailless quadcopter	lithium ion batteries	battery elimination circuitry	microtechnology	exteroceptive sensors	engines	real-time	control theory	aircraft controls	radio electronics	hierarchical control systems	3D Robotics
	rotorcraft	lithium polymer batteries	embedded systems	system on a chip	non-coopertive sensors	servomotors	Raspberry Pis	open loop	air brake	radio frequency front-end	scripting language	AeroDreams
	rotary wing	solid-state lithium-ion battery	microcontroller unit	advanced microcontoller bus architecture	automotive safety technologies	weapons	Beagleboards	closed loop	balanced rudder	RF power margin	ActivePerl	DJI
	co-axial rotors	18650 battery	embedded controller	MPSoC	collision avoidance	payload actuators	NavIO	networked control system	control loading system	image response	AMPL	Flirtey
	intermeshing rotors	dual carbon battery	Arduino	Amlogic	drive by wire	speakers	PXFMini	adaptive control	controllable slope soaring	feed line	AngelScript	Sky-Watch
	slowed rotor	graphene foam	PIC microcontroller	ARC	child safety lock	actuators	Nuttx	intelligent control	servo tab	RF probe	HyperTalk	
	cyclogyro	lithium hybrid organic battery	single-board microcontroller	Atom	shock absorber	helical band actuator	Xenomai	perceptual control theory	autopilot	transceiver	Rc	
	mono-	lithium ion manganese oxide battery	Netduino	PSoC	crosswind stabilization	linear actuator	DDS-ROS 2.0		helicopter flight dynamics	radio frequency front-end	Rexx	
	bi-	lithium iron phosphate battery	RAM image	Rockchip	gyroscopes	plasma actuator	KKMultiCopter		multirotor flight dynamics	antenna	finite-state machine	
		lithium-air battery	sensing floor	InvenSense	anti-rolling gyro	rigid chain actuator	ArduCopter		aerodynamics	loop antenna	alternating finite automation	
		lithium-sulfur battery	smart camera	deformable mirror	control moment gyroscope	propellers			plane flight dynamics	antenna boresight	asymmetric numeral systems	
		lithium-titanate battery	hardware reset	hydrogen sensor	fibre optic gyroscope	modular propeller			aerodynamic heating	batwing antenna	Buchi automation	
		lithium-ion flow battery	Bit banging	microreactor	quantum gyroscope	scimitar propeller			airspeed	array gain	Krohn-Rhodes theory	
		nanoball batteries	flash memory emulator	microthermoforming	accelerometers	cyclorotor			inertia coupling	dipole antenna	Quotient automaton	
				mask inspection	gravimeter	contra-rotating propellers			roaxial rotors	halo antenna	permutation automation	

single board computers	laser accelerometer	power electronics
Banana Pi	liquid capacitive inclinometers	digital electronic speed controllers
Kano	PIGA accelerometer	commutation cell
P112	inertial measurement unit	gate driver
VoCore	navigational equipment	magnetic amplifier
flight controller	compass	power module
flgiht controller board	inertial navigation system	power semiconductor device
	pressure reference system	optic diodes
	ecompass	LEDs
	transfer alignment	LED circuit
	barometer	LED lamp
	global positioning system	LED strip light
	GPS receiver	crystal LED
	clock drift	flexible OLED
	digital anchor	phosphorescent OLED
	vehicle tracking system	superluminescent diode
	positioning system	
	pseudorange	

spiral antenna	behavior trees
digital signal processing	automated planning and scheduling
analog-to-digital converter	hierarchical task planners
aliasing	state space planning
bandlimiting	partial-order planning
infinite impulse response	kinodynamic planning
oversampling	multi-agent planning
half-band filter	reactive planning
data transmission	PID controller
data link	robot kinematics
adaptive aqualizer	motion planning
backward channel	kinodynamic planning
bandwidth cap	kinematic chain
narrative traffic	articulated robot
parity bit	passive dynamics
telemetry	serial manipulator
EMR telemetry	tree searches
electronic data capture	and-or tree
remote data capture	k-ary tree
remote terminal unit	suffix tree
wildlife radio telemetry	radix tree

satellite navigation system	trace tree
satellite navigation	genetic algorithms
timation	cultural algorithm
total electron content	Fitness function
vehicle tracking system	genetic fuzzy systems
hybrid positioning system	genetic programming
automatic vehicle location	truncation selection
telecommunication equipment	self-level
radio transmitter	altitude hold
block upconverter	hover hold
hybrid coil	headless mode
radio spectrum scope	care-free
network termination	take-off
optical line termination	autoland
ground control station	aerobatics
wearable devices	optical devices
electroencephalography	camera
beta wave	digital camera
spike-and-wave	remote camera
PGO waves	still camera
evoked potential	video camera
ear-eeg	range imaging
uplink	Wright camera
downlink	night vision devices
real-time video	laser beam

	profiler
downstream	head-up display
smartphone	electric eye
tablet	
computer	
human movement recognition	

Appendix E Result of Wikipedia-based morphological matrix after developing sub-dimensions (extended model)

Dimensions	**body**		**power supply**	**platform**			**computing**		**sensors**				
Sub-dimensions		rotorcraft	**lithium ion batteries**		**embedded systems**		**microtechnology**		**automotive safety technologies**	**gyroscopes**	**accelerometers**	**navigational equipment**	**global positioning system**
	tailless quadcopter	rotary wing	**lithium polymer batteries**	battery elimination circuitry	microcontroller unit	flight controller	system on a chip	exteroceptive sensors	collision avoidance	anti-rolling gyro	gravimeter	compass	GPS receiver
	mono-	co-axial rotors	solid-state lithium-ion battery		embedded controller	flgiht controller board	advanced microcontoller bus architecture	non-coopertive sensors	drive by wire	control moment gyroscope	laser accelerometer	inertial navigation system	clock drift
	bi-	intermeshing rotors	18650 battery		Arduino	single board computers	MPSoC	inertial measurement unit	child safety lock	fibre optic gyroscope	liquid capacitive inclinometers	pressure reference system	digital anchor
		slowed rotor	dual carbon battery		PIC microcontroller	Banana Pi	Amlogic	barometer	shock absorber	quantum gyroscope	PIGA accelerometer	ecompass	vehicle tracking system
Values		cyclogyro	graphene foam		single-board microcontroller	Kano	ARC		crosswind stabilization	…	…	transfer alignment	positioning system
			lithium hybrid organic battery		Netduino	P112	Atom					…	pseudorange
			lithium ion manganese oxide battery		..	VoCore	PSoC						…
			lithium iron phosphate battery		RAM image	…	Rockchip						
			lithium-air battery		sensing floor		InvenSense						

lithium-sulfur battery	smart camera	…
lithium-titanate battery	hardware reset	deformable mirror
lithium-ion flow battery	Bit banging	hydrogen sensor
nanoball batteries	flash memory emulator	microreactor
…	…	microthermoforming
		…

Dimensions	**actuators**					**software**	**loop principles**	**flight controls**	
Sub-dimensions		**actuators**	**propellers**	**power electronics**	**optic diodes**		**control theory**		**aerodynamics**
	engines	helical band actuator	modular propeller	digital electronic speed controllers	LEDs	real-time	open loop	aircraft controls	plane flight dynamics
	servomotors	linear actuator	scimitar propeller	commutation cell	LED circuit	Raspberry Pis	closed loop	autopilot	aerodynamic heating
	weapons	plasma actuator	cyclorotor	gate driver	LED lamp	Beagleboards	networked control system	helicopter flight dynamics	airspeed
	payload actuators	rigid chain actuator	contra-rotating propellers	magnetic amplifier	LED strip light	NavIO	adaptive control	multirotor flight dynamics	inertia coupling
Values	speakers	…	…	power module	crystal LED	PXFMini	intelligent control	air brake	roaxial rotors
				power semiconductor device	flexible OLED	Nuttx	perceptual control theory	balanced rudder	…
				…	phosphorescent OLED	Xenomai	…	control loading system	
					superluminescent diode	DDS-ROS 2.0		controllable slope soaring	
					…	KKMultiCopter		servo tab	
						ArduCopter		…	

Dimensions	**communications**								
Sub-dimensions		**telecommunication equipment**	**satellite navigation system**	**data transmission**	**digital signal processing**	**electroencephalography**	**telemetry**	**antenna**	**radio communications stubs**
Values	ground control station	radio transmitter	satellite navigation	data link	analog-to-digital converter	beta wave	EMR telemetry	loop antenna	radio frequency front-end
	wearable devices	block upconverter	timation	adaptive aqualizer	aliasing	spike-and-wave	electronic data capture	antenna boresight	RF power margin
		hybrid coil	total electron content	backward channel	bandlimiting	PGO waves	remote data capture	batwing antenna	recurrent rotation
		radio spectrum scope	vehicle tracking system	bandwidth cap	infinite impulse response	evoked potential	remote terminal unit	array gain	thermal fade
		network termination	hybrid positioning system	narrative traffic	oversampling	ear-eeg	wildlife radio telemetry	dipole antenna	spurious emission
		optical line termination	automatic vehicle location	parity bit	half-band filter	…	…	halo antenna	M-ray transmission
		…	…	…	…			spiral antenna	…
								…	

Dimensions	**autonomy**									**manufacturers**
Sub-dimensions		**aircraft landing systems**	**robot kinematics**	**automated planning and scheduling**	**scripting language**	**finite-state machine**	**tree searches**	**genetic algorithms**	**optical devices**	
Values	hierarchical control systems	autoland	motion planning	hierarchical task planners	ActivePerl	alternating finite automation	and-or tree	cultural algorithm	camera	3D Robotics
	behavior trees	microwave landing system	kinodynamic planning	state space planning	AMPL	asymmetric numeral systems	k-ary tree	Fitness function	digital camera	AeroDreams
	PID controller	precision approach radar	kinematic chain	partial-order planning	AngelScript	Buchi automation	suffix tree	genetic fuzzy systems	remote camera	DJI
	self-level	simplified directional facility	articulated robot	kinodynamic planning	HyperTalk	Krohn-Rhodes theory	radix tree	genetic programming	still camera	Flirtey
	altitude hold	SCAT-I	passive dynamics	multi-agent planning	Rc	Quotient automaton	trace tree	truncation selection	video camera	Sky-Watch
	hover hold	transponder landing system	serial manipulator	reactive planning	Rexx	permutation automation	…	…	range imaging	
	headless mode	marker beacon	…	…	…	…			Wright camera	
	care-free	instrument landing system							night vision devices	
	take-off	…							laser beam profiler	

head-up
display

electric eye

초 록

본 학위논문은 유망기술의 미래를 예측하는 기술 포사이트 (technology foresight) 분야의 새로운 방향을 제시한다. 기술 포사이트의 이론적/방법론적 패러다임은 소수의 전문가에 의존하는 정성적 (qualitative)이고 참여적(participatory) 방법에서 보다 정량적(quantitative)이고 데이터 중심(data-driven)의 접근으로 전환하고 있다. 하지만 기존 문헌은 이러한 현상의 개념적인 정의를 확립하는 것에 그쳤고, 이를 실현하기 위한 명확한 방법을 제시하지 않는 한계점을 갖고 있다. 이를 위해, 본 논문은 방대한 양의 텍스트 정보와 다양한 텍스트 마이닝 기법을 활용한 데이터 기반의 기술 포사이트(data-driven technology foresight)라는 개념을 새롭게 조명한다. 또한, 영향 탐색(impact identification), 영향 분석(impact analysis), 기획 개발(plan development), 기술 아이디어 창출(technology ideation)로 구성된 유망기술의 사전적 관리 (proactive management) 프로세스에 적용하여 연구의 실효성을 확인한다.

본 논문의 목표는 세 가지이다. (1) 웹에서 어떤 종류의 데이터가 제공되며, 기술 포사이트 분야에서 유용하게 활용될 수 있는 데이터 소스는 무엇인가? (2)이러한 데이터는 어떤 구체적인 목적을 위해 적용될 수 있으며, 이를 수행하기 위한 기법은 무엇인가? (3) 데이터 기반의 기술 포사이트는 어떤 미래 연구분야에 기여할 수 있으며, 활용을 통한 긍정적인 효과는 무엇인가? 본 연구는 본질적으로 탐색적이고 실험적인 특성을 갖고 있기 때문에 알아보고자 하는 문제에 대한 최종적이고 결정적인 답을 제공하지 못할 수 있다. 하지만, 이와 같은 접근은 데이터 기반의 기술 포사이트라는 새로운 패러다임에 대한 깊은 이해와 폭 넓은 활용을 가능하게 하는 기반을 제공 할 것으로 기대된다.

주요어 : 기술 포사이트, 텍스트 분석, 유망 기술, 기술 경영, 데이터 기반 접근, 정량적 연구

학 번 : 2013-31007