Universiteit Leiden

ICT in Business and the Public Sector

Trend Research and
Strategic Forecasting

Name: Rashid Aziz Raja

Student No: s1902571

Date: December 10th, 2018

1st Supervisor: Dr Christoph Johann Stettina

2nd Supervisor: Dr Suzan Verberne

MASTER'S THESIS

Leiden Institute of Advanced Computer Science (LIACS)
Leiden University
Niels Bohrweg 1
2333 CA Leiden
The Netherlands

Acknowledgement

Alhamdulillah! All praises to Almighty God, who bestowed upon me a minute portion of His boundless knowledge by virtue of which I was able to accomplish this research.

I am greatly indebted to my supervisors Dr Christoph Johan Stettina and Dr Suzan Verberne. Without their personal supervision, advice and valuable guidance, completion of this project would have been doubtful.

I am also thankful to the director of Centre for Innovation (CFI), Michelle, for allowing me to organize the workshops with the experts of Centre for Innovation. I am grateful to the experts of both labs of CFI, Online Learning Lab and Future Work Lab, for taking out their time from their busy schedule. I am thankful to the CIO of Leiden University, Dr Jan-Willem, for giving me his precious time for an interview.

I want to say thank you to many other people who were, directly and indirectly, involved with this research. I am thankful to Helena Frijns, who was there from start, to guide me. Thanks to Lyanca, for helping me in organizing the workshops. Especially for the text-mining part, I am thankful to Hitesh and KNIME community, who were there to guide me and to answer my questions.

Last, but not least, I am thankful to my parents, friends and other family members, who have been a constant source of encouragement.

Executive Summary

Centre for Innovation of Leiden University, which works on the trending topics like future of work and future of education and how technological trends impact these topics. They require a method through which they can understand the trends well and which should help them in making plans and strategies to be the first one to gain competitive benefit by exploring those trends early.

Meanwhile, there is a need for forecasting by every organization or individual. There are a lot of sources which can be used to do forecast. A lot of textual data is around us; either it's in newspapers, web pages, books, reports, papers, … you name it. The best way to extract the textual information from all these sources for forecast is through text-mining. Using only single source is not as fruitful as combining different sources of information. That information can play a great role to forecast the trends.

That is why the main question for this research is: To what extent can text mining from different data sources (reports and Twitter) support experts in trend forecasting?

So, for that different mediums of information; databases, social media and people are combined in this research to see how good this method is for forecasting. Trend reports and Twitter data was processed through the text-mining tool to extract information out of them for experts to do strategic forecasting related to digital learning and future of work. This method was tested by the experts of Online Learning Lab and Future Work Lab, that how far this method can help them and how much it is useful for forecasting.

Keywords:

Trend Research, Forecasting, Text-mining, Experts, Digital Learning, Future of Work, Twitter.

Table of Contents

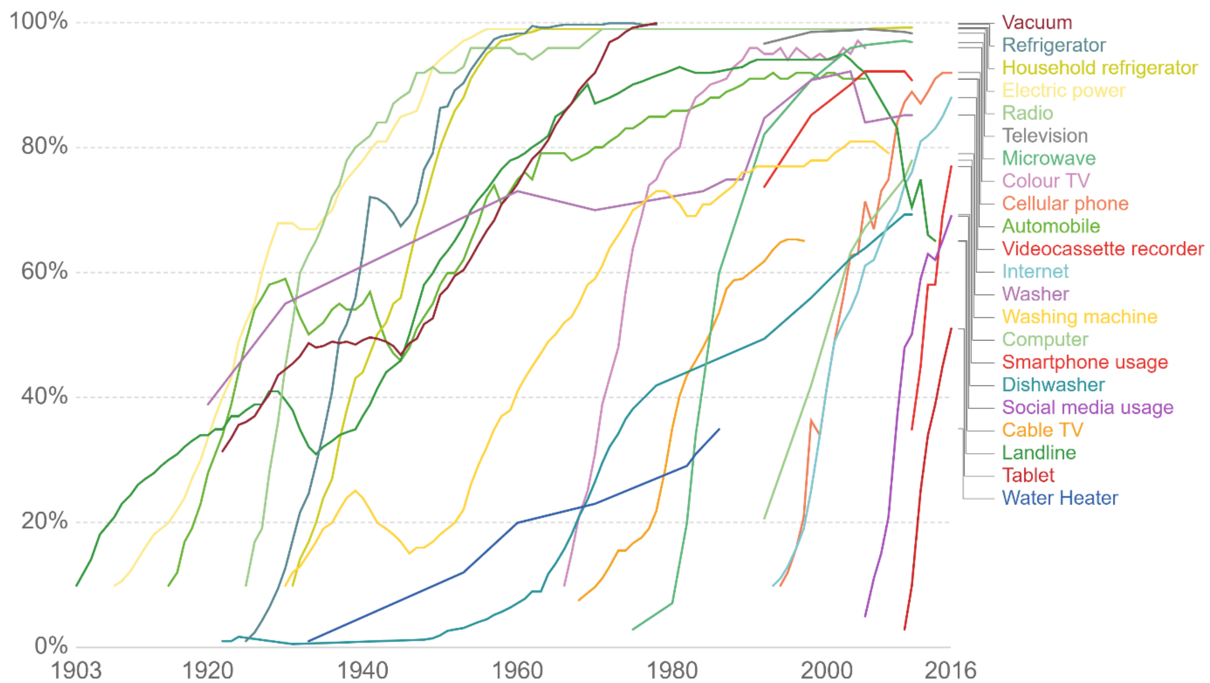
1. Introduction

"What's certain is that World is changing faster than at any time in human history."
Philip Stephens, FT, December 2012

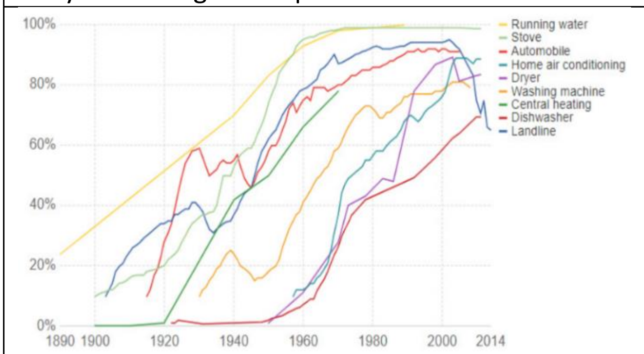
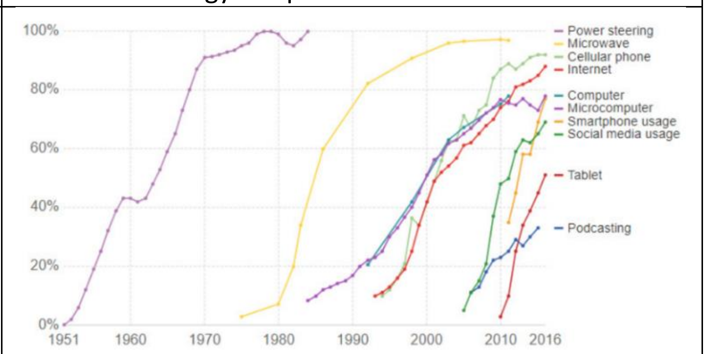
The world is evolving. Society is moving fast, and organizations need to move faster. Everything is changing around us so much that people are in sort of race where they are trying to be part of it. Either you are businessman, entrepreneur or an employee; you must be willing to evolve with the pace of the society around you. There is a lot of data available around us. Either it is data from books, reports, websites or social media, all these heterogeneous sources can increase the information which can be of help when analysing new aspects (Crawford & Boyd, 2012) (Kitchin, 2014). These new aspects are always important for businessmen and especially entrepreneurs, who are continuously looking for new opportunities. These new aspects and opportunities always come handy when you have to track things which you are interested in or things that require prediction.

For prediction, a track record about related product or thing is required. That record can have better information regarding the product or a thing. That information is required up to several time lapses, which is also known as trend analysis because you are analysing the previous data at different time periods in order to predict. Trend analysis is about which things are present now, which will be there in future and which will not remain in future. If we talk about technological gadgets, for the prediction of these gadgets, a lot of information is required and at different time spans. For example, in 2010, how many people were using these items. By 2011, who doubled them and who left using them and similarly, adding different categories in this information like what features are attracting people for these gadgets and how much they are going to pay for different features, several other information at different time lapses help experts to predict for next year or coming future.

Looking at history, according to the world learning organization - Pearson, in 19th century, it took Britain about 150 years to double their GDP (Gross Domestic Product) per capita. In the 20th century, it took United States about 20 years to double its GDP per capita, and it will take only 15 years in this 21st century for China and India to double their GDP per capita. It took decades for people to adopt various technologies but the ratio of adoption in this era concerning the ratio of adoption in 20th century or before that, is quite less. A graphical chart published by Horace Dediu, (Comin & Hobijn, 2004) and others, shows the adoption rate of technologies by households in United States.



Graph 1: Technology Adoption by households in US (published by Horace Dediu, (Comin & Hobijn, 2004) and others)

Early Technologies Adoption	Latest Technology Adoption
	

Graph 2: Early and Latest Technologies Adoption rate (published by Horace Dediu, (Comin & Hobijn, 2004) and others)

In the graphs above, the adoption rate can be seen. The technologies in the start of the 20th century took decades to cross even 20% of adoption by households (*Graph 2 Left*). Whereas by the end of 20th century and era of the 21st century, it took less than that to adopt technologies. So, rate of these adoptions is pretty much faster than before (*Graph 2 Right*).

In recent years, technologies have developed really fast. If we see through recent decades, thousands of technologies have been invented. Things which were only said to be the dreams, are becoming real. The things which people can't imagine are becoming into reality. Who would have thought in 19th century or even before that the answer of multiplication or addition of several digit numbers will just take less than the time to blink of an eye? Who would have thought a person can travel faster than ever before? The time it took to months and days to travel, has been reduced to hours and minutes. People can control appliances of their homes through smart portable devices, which are in their pockets (smartphones). They can see and feel the places through virtual reality, sitting from their homes. Smart credit cards are replacing cash notes, and similarly, those cards are being replaced to digital currency (cryptocurrency). Automation has dominated in industries and other fields. Robots are taking complex work as well as daily works of life.

In the era of all these innovations, no one can surely tell people on which side we are moving. A lot of technologies are coming out every year. Some are being replaced by previous technologies (like Infrared to Bluetooth), some are upgrading to the current technologies (like 2G network to 3G network, and 3G to 4G network and now 5G is under works), and some technologies are totally new (like 3D printing and more). Several questions arise in mind related to technology. Is the current technology sustainable? Does it have a future or is it going to be just like another floppy disk? Or how long will the new technology take to be accepted by people? How beneficial will it be for organizations or a businesses? Which are the emerging technologies so that maximum benefit should be exploited and lead the society? To answer all these questions is technological forecasting.

(Balachandra, 1980) sums up the purpose of technological forecasting. He described that technological forecasting tries to answer one or other questions like when would it be possible for the new development to be broadly accepted by the society? And what would be possible for the development in a given area of technology in the near future? These questions show the significance of technological forecasting.

A widespread practice is being used to collect information through historical data in order to predict what will happen in the future. This practice is known as trend analysis. Trend analysis is mostly done in the companies or organizations where a lot of prediction is required especially stock exchange, oil companies and many others. The benefit of this trend analysis is that one will have a prior knowledge of incoming future and they will be ready to get potential benefit out of it. They will be the one who enter in the market first, having an advantage over the ones entering late.

1.1 Problem Statement

People get insecure when they see the world around them is evolving. People around them move fast, the societies they live in are advancing day by day. There is a fear in between them – the fear of getting behind from others. Every organization wants to be the first one in the new market and wants to grab the major portion of it to gain maximum profit. Mostly businesses, industries and organizations want to get along with the new changes and new technologies so that they can stand parallel to others, as no one wants to stay behind in the field of technology. Once you lack behind a little, you are on edge of losing the race.

Similarly, if we don't know which changes are coming in society, either they are social changes or technological changes, how can one decide for the future? In order to deal with current and future problems and to help people to prepare for the future, there is a need to find a way to overcome this situation – a better forecasting for the future which can help individuals and organizations to achieve their goals so that they can help themselves and a society for betterment.

There are different methods and techniques to do forecasting. But which methods are more effective? Which mediums of information can be used while doing forecasting? Or the reports of the organization are enough to do forecasting? Or only experts' discussion can do enough forecasting? A lot of questions arise because, for forecasting, you need a lot of data to predict. That data is spread in different mediums such as in the companies' reports, its audits, in newspapers, on internet, in people's mind, in tech reports, databases, in other words – everywhere. But which method is more useful and can use multiple mediums of information for forecasting?

1.2 Research Question

As described, in the above *section – Problem Statement –* that there are a lot of methods and mediums to do forecasting. Around 80% of information lies in textual data (Yu, Wang, & Lai, 2005). To extract the information out of this huge data, which is mostly in the form of text and if we do text-mining on this textual-data, how far will that be effective for people who make strategies and plans and do forecasting. So, the main question of this research is:

To what extent can text mining from different data sources (reports and Twitter) support experts in trend forecasting?

1.3 Purpose of Research

There is a need to make good strategies which can help people to get ready for the future. For every decision which is made for future, it needs to be well forecasted. Who wants to invest in that business which is likely to be bankrupted in the coming months or a year(s)? Obviously, no one. Now, let the question be like, who want to invest in that business which seems like to be bankrupted, but business's forecast says (after implementing their best strategies) that it will revive and will bounce back? Yes, people will go for it. So, where do these strategies come from? The answer is forecast.

In order to do forecasting, a lot of data is required which should be relevant for the cause to forecast. Every year, big and well-known companies and institutes create future trend reports which are the collaborated work of thousands of experts around the globe. Thousands of reports are published every year, which predicts the incoming or future trends (including social and mostly technological). The normal technique which is applied mostly to make strategies and to do forecasting is Delphi technique (experts' opinions and decisions). Using only experts' opinion is not as fruitful as if it is combined with the knowledge taken from the documents (which includes the opinion of experts from all part of the world) and through the use of social media. Of course, this becomes huge data which includes number of trend reports collected from around the world plus gathering information through social media (Twitter) where people share the information (most likely) first. At the end, gathering this pile of information for experts from where they can derive strategies in order to prepare for the future and earlier than others.

Therefore, text-mining tool is the solution of gathering all of this information that lies in the form of text for the experts. Text-mining is by far best way to extract textual information out of the trend reports and from social media (Twitter). It can save a lot of time for people who read the trend reports and draw required information out of these reports. And similarly, it can also save time for those people who check social media manually in order to extract the required information. Text-mining do all this part more easily, efficiently and fast which can save a lot of time and resources.

In this thesis, a method of combining different mediums of information including expert's discussion was tested in order to see how good this method (reports + twitter + social media) is and what changes or revision will be required to make it better for the next time.

2. Literature Review

In this section, related work of other authors is discussed which is related to this research. This section is categorized into four section. The first section describes related work of forecasting. Second section describes related work of forecasting methods. Third section describes forecasting and text-mining, and fourth section describes the literature gap.

2.1 Forecasting

Prediction of an activity or incident accurately that will held in future is known as forecasting (Hyndman & Athanasopoulos, 2018). From thousands of years, forecasting has always attracted peoples' attraction. About 700 BC, in the book of Isaiah (Jewish Prophet) it's written as:

"Tell us what the future holds, so we may know that you are gods…"
(Isaiah 41:23)

Forecasting was forbidden in Rome during 357 AD, in which emperor Constantine issued a verdict "*to consult a soothsayer, a mathematician, or a forecaster … May curiosity to foretell the future be silenced forever*." (Hyndman & Athanasopoulos, 2018) (Goodwin, Ord, Öller, Sniezek, & Leonard, 2002). During 1736 in England, sentence of three months of custody was declared to the one who became an offence to defraud by charging money on prediction (Hyndman & Athanasopoulos, 2018). However, in recent years forecasting has been widely accepted and therefore, has been used in most of the fields.

It has a significant standing in our lives either we talk individually or collectively in the form of organizations. If we see for individual level, we want and always try our investments and our occupations successful. Similarly, every organization pay its significant time and amount on forecasting of its sales and profits. People who are the ones making important decisions for an organization forecast future of only those things and activities which has uncertainty about the future (Goodwin et al., 2002). Goodwin and his co-authors gave an example in their book that no one will forecast the sunrise of tomorrow because that's the certain thing.

When people talk about forecasting, they mix it up with planning and find a confusion in between planning and forecasting. Matter of the fact is, it is the difference of two words "should" and "will". It is the difference of *what the world "should" look* like (referred as planning) and *what the world "will" look like* (referred as forecasting) (Goodwin et al., 2002). Similarly, Hyndman and Athanasopoulos describe forecasting, goals and planning as three different things where forecasting is often confused with goals and planning. Planning consists of required actions determining to achieve forecast exactly like the goals. Goals are the required target or the requirements. And forecasting is, predicting the future with available information keeping historical information and data into consideration. For forecasting, variables are also kept in mind, that may affect the outcomes (Hyndman & Athanasopoulos, 2018).

2.2 Forecasting Methods

Latest book of Hyndman and Athanasopoulos, which is published earlier this year (April 2018) about forecasting, describes two major methods of forecasting which also advocates Sanders and Reid's methods and some other authors too. These two main methods are qualitative forecasting method and quantitative forecasting method (Sanders & Reid, 2012) (Hyndman & Athanasopoulos, 2018).

2.2.1 Qualitative forecasting method

Qualitative forecasting is that forecasting which is based on experts' views, built on their thoughts, opinions and decisions. Qualitative forecasting is mostly done at the time when data is not available for forecasting (Hyndman & Athanasopoulos, 2018) (Sanders & Reid, 2012). Delphi is one of the main examples of qualitative forecasting method as it involves judgements and opinions. Qualitative forecasting method is a method which includes brainstorming, experts panels, analysis, questionaries' and surveys (Popper, 2008). As Delphi method is mostly consider as qualitative method, but this method has been used to obtain quantitative results too whether it is quantifying risk or it is perception of quality process (Hallowell & Gambatese, 2010).

2.2.2 Quantitative forecasting method

Quantitative forecasting is more about numbers and is built on historic and present numeric data. Core of this forecasting method is based on mathematical modelling. Time series model is one of the quantitative forecasting method's example where patterns of past data is used (Sanders & Reid, 2012). Other than time series data collection, cross-sectional data collection also falls in quantitative forecasting method in which data is taken into consideration at a single point (Hyndman & Athanasopoulos, 2018).

There is also a third method which is described as semi-quantitative method. Semi-quantitative method is a hybrid of qualitative method and quantitative method (Popper, 2008). This method includes data, mathematical principles and modelling which includes data from different timelines. Other than this, it includes judgements and opinion of experts.

Each forecasting method has its own respective advantages and disadvantages. If we talk about qualitative forecasting method, Delphi technique is one of them which does its best when any other technique is not working. It is also used when forecasting is not going as it is planned (Modrak & Bosun, 2014). Similarly, there are also several concerns of planning and forecast in Delphi technique which can be bias and leads into reduction of accuracy (Sanders

& Reid, 2012). Likewise, quantitative forecasting method also have its own favours and drawbacks. Results from quantitative forecasting method are consistent and objective because they are highly reliable on present and historical data. Due to vastly dependable of quantitative forecasting method, it is very difficult to get quantifiable data. And all the results from this method are highly dependable on quality of data which matters the most in quantitative forecasting method. More the data is good, better the results will be (Sanders & Reid, 2012).

2.2.3 Technology Forecasting

"It is difficult to predict, especially the future."

A quote referred and attributed to Niels Bohr (Danish physicist) and many others which shows even the prior research found the prediction challenging. Due to uncertainty and unreliability of data, it is very difficult to predict the future of technology (Ayers, 1969).

(Jun, 2014) and (Coates, Vary, 2001) describe technology forecasting as an approach of predicting the technology in future. (Yoon & Park, 2007) define the technological forecasting as a term which predicts the direction of technology like; Where is it going? From where is it coming from? And how much time does it take to change technology? Major issues, mostly in organizations, like removing risk factors, prioritisation of several tasks including how much of resources should be allocated, technology forecasting enables in all of these decisions making and analysis process. Thus, need of individuals or groups either publicly or privately, can be fulfilled with technological forecasting (Goodwin et al., 2002) (Yoon & Park, 2007). Every organization or institute, who are concerned with future regarding if it is about investments, educations, lifestyle or others, they require technological forecasting. In this era, where technology is changing rapidly, even for the betterment of people and their plans and needs, government also requires technological forecasting to satisfy its peoples' requirements and to create multiple opportunities (Zhu & Porter, 2002).

(Porter, 2010) describe technology foresight as "multi-dimensional activity". Because it has different meanings and every meaning comprises of different objectives which should not be taken as single entity or "one size fits all".

Yoon and Park describe four components of technology forecasting in which technology forecasting is summarized. First component is comprised on the direction of technology. Like in which way it is growing. This component helps the researcher to guide and to see the plan of the development of technology. Next step, second, is to identify the technology alternatives that which other substitutes are available or could be there that can be taken into consideration. This step also keeps time into consideration that how much time will it take to develop new technology. Third step is of analysis in which several factors are taken into consideration including strengths, weaknesses, constraints, opportunities and threats.

Macro environmental scanning (PESTLED- Political Economic Social Technological Legal Environmental Demographic) is also performed in this step. Last step is of technology forecasting tool which help researchers and collaborators in collaboration of technology development. Communication in between them is consider as a key factor where they can collaborate to make better decisions (Yoon & Park, 2007).

In recent years, several tools and methods have been developed. By using the available information, they are not only helping in forecasting but also helping in identifying the factors which are related to foresight and can lead to shape better future (de Miranda Santo, Coelho, dos Santos, & Filho, 2006).

Martino describes the most used forecasting methods which came up as a result through several surveys are Delphi and Trend Extrapolation (Martino, 1980). Delphi is the process of experts' opinions and suggestions (Rowe & Wright, 2001) (Yun, Jeong, & Kim, 1991) while trend extrapolation is technique which uses mathematical methods in which time series analysis is top of the list (Gardner, Jr, & McKenzie, 1985).

There are different methods to analyze future-oriented technologies. *Table 1* shows the different method families (main methods) on left side while right side shows the sample of these main methods.

Methods families	*Sample methods*
Creativity approaches	TRIZ, future workshops, visioning
Monitoring and intelligence	Technology watch, tech mining
Descriptive	Bibliometrics, impact checklists, state of the future index, multiple perspectives assessment
Matrices	Analogies, morphological analysis, cross-impact analyses,
Statistical analyses	Risk analysis, correlations
Trend analyses	Growth curve modelling, leading indicators, envelope curves, long wave models
Expert opinion	Survey, delphi, focus groups, participatory approaches
Modelling and simulation	Innovation systems descriptions, complex adaptive systems modelling, chaotic regimes modelling, technology diffusion or substitution analyses, input-output modelling, agent-based modelling
Logical/Causal analyses	Requirements analysis, institutional analyses, stakeholder analyses, social impact assessment, mitigation strategising, sustainability analyses, action analyses (policy assessment), relevance trees, futures wheel
Roadmapping	Backcasting, technology/product roadmapping, science mapping
Scenarios	Scenario Management, Quantitatively based scenarios
Valuing/Decision-aiding/economic analyses	Cost-Benefit Analysis (CBA), Analytical Hierarchy Process (AHP), Data Envelopment Analysis (DEA), Multicriteria Decision Analyses
Combinations	Scenario-simulation (gaming), Trend impact analysis

Table 1: Future-oriented technology analysis methods (Porter, 2010)

Every method is useful for different type of requirements and outcomes. Several studies show that for technological forecasting, qualitative and subjective approaches are used, and Delphi technique is one of them (Roper et al., 2011) (Mitchell, 1992) (Yun et al., 1991) (Jun, 2014). Delphi technique is used when a human judgement is required to do the forecast, and it is more accurate than the individual experts (Rowe & Wright, 2001). But on the other hand, the results from Delphi technique due to highly dependent on the experts' experience ends into inconsistency (Yoon & Park, 2007) (Jun, 2014).

There are several methods for forecasting. *Table 2* describes the advantages and drawbacks of some of these forecasting techniques and methods, as described by (Meredith, Mantel, & Jr., 1995).

Technique	Advantages	Disadvantages
Monitoring	Unsophisticated First step to any good TF technique Low cost	Must review a great quantity of material Time-consuming Collection/summarization technique Not a "predictor"
Scenarios	Aids in understanding present Develops a plan of action for future	Dependent on select few (the writers) High cost Too general
Morphology	Goal-setting method Exhaustive Precise methods Breaks down whole into component parts	Extremely time-consuming Must know all alternatives Extremely high cost Impossible combinations must be recognized
Relevance trees	Goal-setting method Structures goal achievement	Unsophisticated Too general Bad project may not be easily seen
Delphi	First step to other TF techniques Many people can participate Eliminates personality conflict	Emphasis on consensus Time-consuming Does not relate final event to means to achieve final event
Cross-impact	Moderates some problems with Delphi Can be computerized Highlights lack of specific knowledge	Very laborious Requires Delphi analysis prior to use Time-consuming Must use same people as in Delphi study

Table 2: Advantages and Disadvantages of forecasting methods (Meredith et al., 1995)

2.3 Forecasting and Text Mining

2.3.1 Text Mining

Text-mining is the method of discovering knowledge from the textual or documented database (Feldman & Dagan, 1995) (Tan, 1999). Tan and Hearst described it as "Text" Data-Mining (Hearst, 1997) (Tan, 1999). Senellart and Blondel defines text mining, in Berry's book, as a method of automatically extracting previously unknown information and new information from different written sources which came after the discovery by computer (Berry, 2004).

As described earlier, around 80% of information is stored in textual documents (Yu et al., 2005). And the textual information, which is stored in the textual documents, plays an important role and is considered as most useful information. There are several reasons why it is considered most useful which are described by (Zhai & Massung, 2016). These reasons are *"Text(natural language) is the most natural way of encoding human knowledge"*, *"Text is by far the most common type of information encountered by people"* and *"Text is the most expressive form of information"* (Zhai & Massung, 2016). Due to large amount of data from mixed sources and day by day increasing sources, it is essential to study these published studies, articles and reports (Ortner, Pfurtscheller, Rizzolli, & Wiesinger, 2014) (Kayser & Blind, 2017).

Text-mining is widely used to extract the information out of the documents. As many researchers have used text-mining to extract the information for different purposes out of these textual documents (like patents, records etc.). As (Meystre, Savova, Kipper-Schuler, & Hurdle, 2008) have extracted information from electronic health records. (Tseng, Lin, & Lin, 2007) used text-mining for analysing and extracting the information from patents. Similarly, many other researchers have used text-mining for different purposes and on different types of textual documents.

Text mining, which is also known as "Intelligent Text Analysis", "Text Data Mining" or "Knowledge-Discovery in Text" (KDT), is generally referred to extracting the information and knowledge from unstructured text (Gupta & Lehal, 2009) (Banu & Chitra, 2015). (Banu & Chitra, 2015) describe text mining as,

> *"a young interdisciplinary field which draws on information retrieval, machine learning, data mining, statistics and computational linguistics"*.

The problem of KDT (which is also known as Text Data Mining), as described by (Gupta & Lehal, 2009) and (Banu & Chitra, 2015), is to

> "*extract explicit and implicit concepts and semantic relations between concepts by using Natural Language Processing (NLP) techniques. Its aim is to get insights into large quantities of text data. KDT, which is deeply rooted into Natural Language Process (NLP), draws on methods from reasoning, knowledge management, statistics, machine learning and others for its discovery process*".

KDT plays an important role in applications like Text Understanding and other emerging applications. The goal of text-mining is to extract the data and analyse the data via application of NLP and analytical methods[1].

There is a difference between data-mining and text-mining (Yu et al., 2005). Yu, Wang and Lai describe data mining, which works on overall data and mostly applied for finding patterns in the data and the relationship between data and patterns, whereas, text mining is the idea of dealing with unstructured text documents or less structured text documents to extract information out of it. Similarly, Navathe and his co-authors also describe the difference between data mining and text mining. As a variation between both forms where text mining is describes as KDT (Knowledge Discovery Text) or as Intelligent Text Analysis whereas data mining is a technique to find different and interesting patterns from huge data (Navathe, Shamkant, & Ramez, 2000). Although both methods, data mining and text mining, looks similar in order to extract the information, but the difference described by (Navathe et al., 2000) is the tools of data mining. As the main purpose of usage of data mining tools that they are designed in such a way that they can extract the information from structured data (structured data is that data which exist in organized form, written or present is straightforward way). On the other hand, text mining tools are designed in such a way that they can work on less structured data or unstructured data (opposite to structured data where data is not in organized form).

There are a lot of ways from where knowledge and required data can be extracted. One can find different methods to get useful and required patterns out of data. As mentioned above about the types of data (structured and unstructured), unstructured data is largely available data which can be converted into knowledge (Banu & Chitra, 2015). Because these days, major portion of researchers are spending their time in research and development in data mining because of the availability of structured data (Gupta & Lehal, 2009) and data mining tools are designed to handle structured data (Banu & Chitra, 2015). Therefore, unstructured data remains available which can be converted into knowledge.

Except from the data mining tools (which are designed to handle structured data), text mining is similar to data mining. Text mining can work with less structured (semi-structured) or

[1] https://en.wikipedia.org/wiki/Text_mining

unstructured data like HTML files, emails and full-text documents etc. As a result, text mining is better solution for organizations and companies (Fan, Wallace, Rich, & Zhang, 2006) (Gupta & Lehal, 2009).

As described by (Fan et al., 2006),

> *"Humans have the ability to distinguish and apply linguistic patterns to text and humans can easily overcome obstacles that computers cannot easily handle such as slang, spelling variations and contextual meaning. However, although our language capabilities allow us to comprehend unstructured data, we lack the computer's ability to process text in large volumes or at high speeds. Herein lays the key to text mining: creating technology that combines a human's linguistic capabilities with the speed and accuracy of a computer".*

So, purpose of text-mining, as described in the last sentence of above quotation by authors, is to create a technology that can combine capabilities of human's linguistics with the computer's speed and precision (Fan et al., 2006).

Figure 1 shows an example of text mining which describes the process of text-mining that how text mining works in different set of textual data (Fan et al., 2006). There are several steps involve in this process. It starts from "selection of documents" from which information is required to be extracted. In next step, text mining tool will get those documents, preprocess it. Preprocess involves several actions like punctuation removal and stop-words etc. It is also advocated by (Yu et al., 2005). Third step is to analyse the text in which several actions are performed up to several times until required information is extracted. These main actions in this step are information extraction, summarization and clustering. Fourth step is about storage of resulting information of step three. Last step shows the pile of knowledge which is available for a user, that is extracted from all these steps.

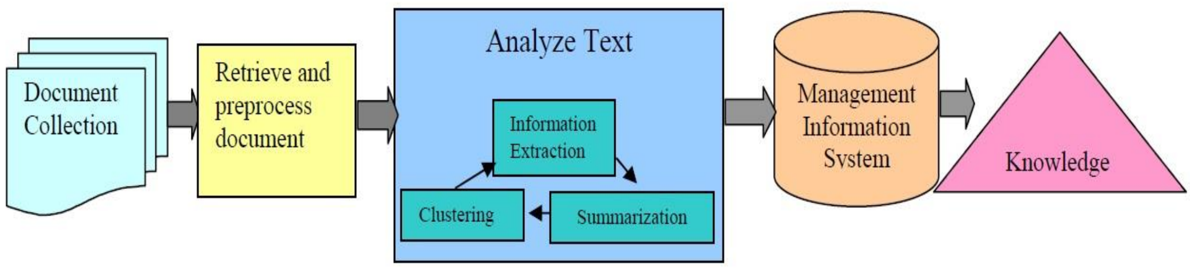


Figure 1: Example of Text Mining (Fan et al., 2006)

Similarly, (Yu et al., 2005) also describe process of text mining into four main stages including collection of documents, preprocessing of those documents, extraction of required features and last is mining of metadata and generation of rough knowledge.

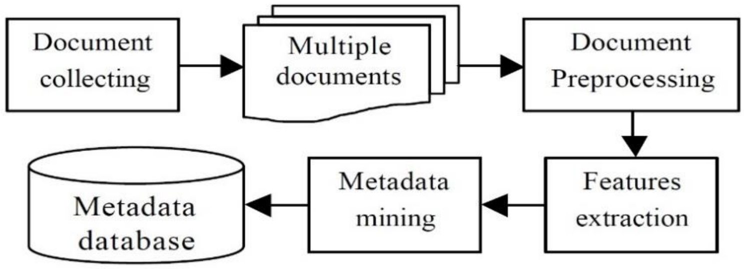


Figure 2: Process of Text Mining (Yu et al., 2005)

(Kayser & Blind, 2017) also describe these steps in more detail. *Figure 3* shows their version of process of text mining. Process of text mining is in four steps which are; text selection, text preprocessing, data analysis and interpretation. Text can be selected from different sources including social media, newspapers, patents or documents etc. After selecting the sources, text preprocessing is performed on the documents (all type of textual sources). Same kind of activities are performed as mentioned in (Fan et al., 2006), (Yu et al., 2005) and many other papers like stopword removal, tokenization, stemming and others. Next step is data analysis, which includes different type of analysis like clustering, association analysis and network analysis. After analysing the data, huge amount of knowledge is available which can be interpreted accordingly.

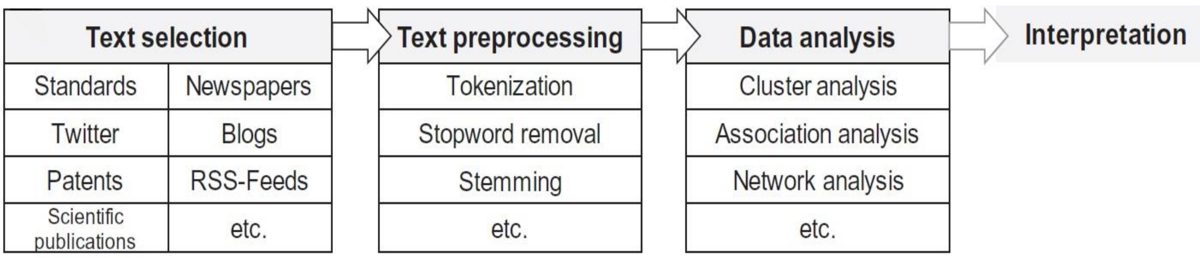


Figure 3: Text Mining Process (Kayser & Blind, 2017)

2.3.2 Text Mining in Forecasting

Forecasting plays a vital role where there are cases of decision making, planning and achieving target whether for one person or an organization (Yoon & Park, 2007) (Zhu & Porter, 2002) (Goodwin et al., 2002). For all the techniques or methods of forecasting (as some are mentioned in *Table 1*), relevant data is required. In the above section of Forecasting methods, forecasting has two main methods which are performed on qualitative data and quantitative data. As mentioned earlier, 80% of company's data is stored in textual form (Yu et al., 2005). A lot of knowledge can be extracted out of this data which is in textual form and can have significant value. Therefore, several forecasts have been done using text mining. Wang and his co-authors used text mining to do financial time series forecasting (Wang, Huang, & Wang, 2012). Jun used text mining using other visual apriori algorithm to create technology forecasting method (Jun, 2014). Yu and his co-authors did forecasting for crude oil market tendency by applying text mining approach (Yu et al., 2005). Mittermayer used text mining techniques to forecast intraday stock price trends (Mittermayer, 2004) and many others.

2.3.2.1 Text as Data

This section describes the sources, which are mostly used for forecasting, and can be analysed with text mining.

2.3.2.1.1 Standards, scientific publications and patents

Out of many sources, these sources are being used to determine any technological change. Mentioned sources are often used for doing forecasting. Data is extracted from these sources which is helpful in determining any changes.

As mentioned above, several papers have been published where researchers used these types of sources to forecast. Like (Tseng et al., 2007) and (Jun, 2014) used text mining on patents.

While using publications, text mining is done on abstraction, keywords, contents or full text of publication is used (Kayser & Blind, 2017) (Cunningham, Porter, & Newman, 2006). Text preprocessing is performed on texts like stemming, stopwords or PoS (Part of speech) grammar based term extraction (van Eck & Waltman, 2011) (Delen & Crossland, 2008) (Glenisson, Glänzel, Janssens, & De Moor, 2005). Similarly, several methodologies like network and mapping approach have been used by (van Eck & Waltman, 2011). (Yau, Porter, Newman, & Suominen, 2014) used statistical model to discover topics (topic modelling). Cluster analysis is done by (Delen & Crossland, 2008) and (Glenisson et al., 2005).

2.3.2.1.2 Social Media

Web-based applications like Twitter, Pinterest, Facebook, Instagram etc. which are used to share the contents or to do social networking, generally known as social media (Kietzmann, Hermkens, McCarthy, & Silvestre, 2011) (Kaplan & Haenlein, 2010). By providing multimodal platforms for millions of people around the globe for discussion, sharing contents and organizing events, online social networks gained much popularity (Cachia, Compañó, & Da Costa, 2007). Cachia and his co-authors used these online social networks as a tool to detect social and trend changes. People who are interested and have expertise in future and forecasting, known as "futurist", create contents and blogs or their views related to their interest and put it on these web platforms. (Pang, 2010) did work on analysing these contents, which are published by futurists that how to create value out of these contents for professional futurists. Whereas Pang work is limited to its theory and practical example is missing in his work. On the other hand (Albert, Moehrle, & Meyer, 2015) and (Glassey, 2012) analyse different contents from online social networks. Albert and his co-authors did technology maturity assessment which is based on blogs (Albert et al., 2015) whereas Glassey did analysis on "data about data" and how much it is possible to contribute to early detection objectives (Glassey, 2012).

There are many options for analysis due to vast variety of data available on these online social networks which need to be done in structural form in order to gain maximum information in real time (Kayser & Blind, 2017). Kayser and Blind believe that a lot has been done in this regard, but still, there are many options which are quite open, e.g. combination of social media and foresight by using scenario workshops or other experts' methods (Kayser & Blind, 2017).

2.4 Literature Gap

This section explains the missing and available thing from the literature. As mentioned in the above section, *Forecasting*, different sources are described in order to do forecast which have their pros and cons.

Based on what is currently known after following literature, there are several methods and techniques to do forecasting. Many are discussed in previous section, but the missing part is that these methods are rarely linked or concatenate with other methods and sources.

Some mentioned authors have used forecasting through textual data mainly from patents and scientific publications like (Yu et al., 2005) (Tseng et al., 2007) (Jun, 2014) and others. Some others have opted for expert opinions and decisions to do forecast like (Mitchell, 1992), (Rowe & Wright, 2001) and others. (Kayser & Blind, 2017) believes that there are many other options which are rarely used and make very little use of those data sources. They particularly describe usage of web contents are hardly use. Huge amount of textual data is available on these online social platforms, which is rarely considered in forecasting activities. If it is used, it's hardly combined with other forecasting methods which are based on expert discussions such as scenario development or roadmapping (Kayser & Blind, 2017).

Similarly, (Porter, 2009) also describes that researcher usually take one medium out of three mediums to do forecast. These mediums are databases (in the form of scientific publications, patents etc.), internet (social media platforms) and people (professional and technical experts).

Medium	**Technology**
A) Databases	Publications and Research Reports
B) Internet	Twitter
C) People	Technical Experts

Table 3: Information Type – Adapted from (Porter, 2009)

Many organizations fail to take technological advantage by neglecting the advantages of row A and B and completely rely on experts' judgements and predictions. According to Porter, this is "folly" of them by not taking advantage from the first two rows and he mentioned that the ones who used empirically based knowledge (row A and B of *Table 3*), and add human expertise (row C) in it, will dominate the ones who totally rely on experts only (Porter, 2009).

Combination of these mediums are very much useful, and one can lead the others who are using only single medium either databases or internet or just experts' reviews.

To summarize, we know there are a lot of mediums of information like reports, web, social media etc. and we also know that there are a lot of techniques to do forecasting, but the missing thing is combination of all the mediums with combination of forecasting techniques to make forecasting more accurate and fruitful.

To overcome this gap, Kayser and Blind describe a process model which is the combination of expert discussion method and text-mining. The techniques they mentioned like scenario planning or similar expert method and text-mining, are not done parallel. As scenario planning, roadmapping or similar (expert related) techniques are solely done by experts (Kayser & Blind, 2017).

Literature recommends that by combining the different mediums of information and applying expert related forecasting method along with text-mining will be a great combo for forecasting. So, the purpose and main question of this research is how effective this method is for forecasting if the information is extracted through different mediums (reports + Twitter) by a text-mining tool and adding expert related method in this process for planning strategies.

3. Research Method and Design

This section describes the research methodology which is used in this research to gather and process the data. This section is categorized into following: case selection, data collection and data sources and analysis.

3.1 Research Strategy

In order to determine, which research strategy and data collection method fits in this research, we follow a process which is described by (Saunders, Lewis, & Thornhill, 2007) and is known as "research onion".

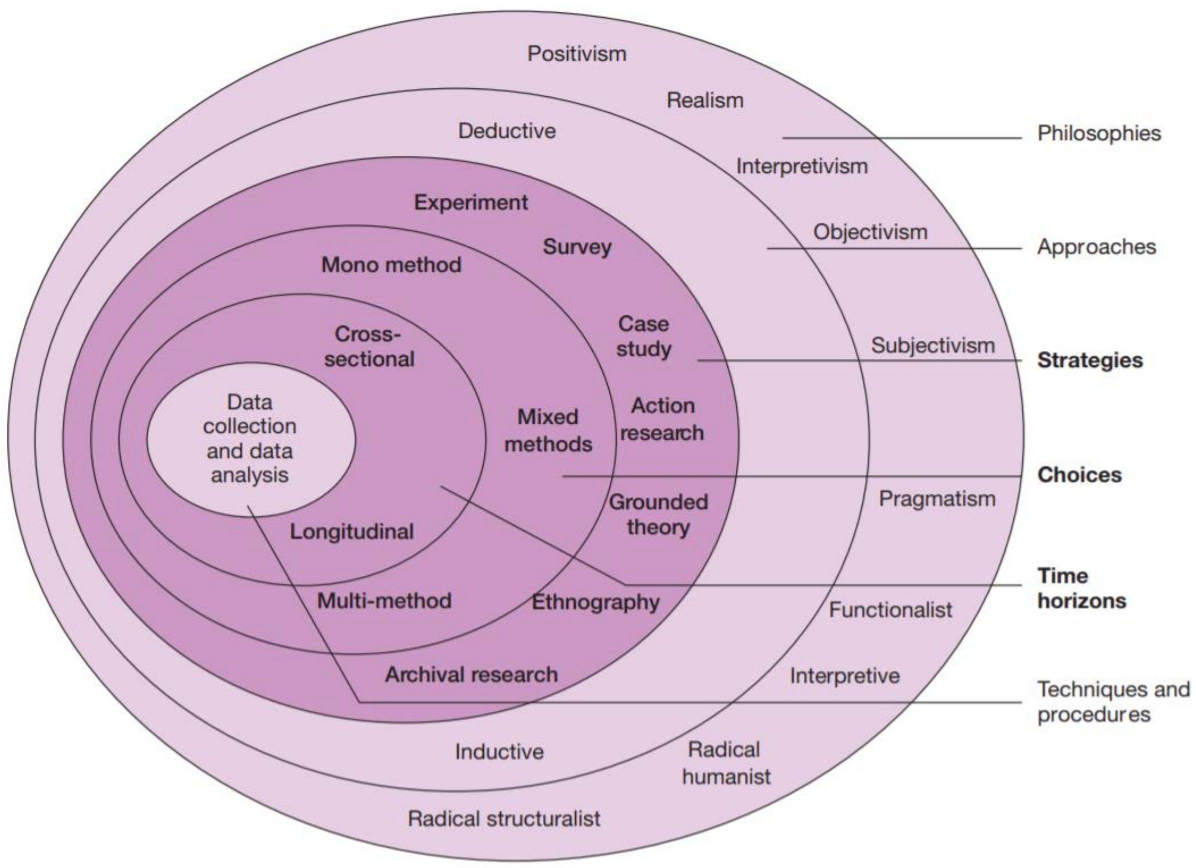


Figure 4: Research Onion (Saunders et al., 2007)

The research philosophy for this research is interpretivism as the objective of this research is to see what experts think that how much this method (by using different information mediums) can help them in forecasting. This requires the interpretation and understanding of the answers given by the experts and interpreting the interviews.

As the question of this research is to what extent can text mining from different data sources (reports and Twitter) support experts in trend forecasting?

The approach to this research is regarded as deductive as the theory is tested in this research. As described above, here the theory says, according to (Kayser & Blind, 2017) and (Porter, 2009), by combining different mediums of information; Database, Internet and People (experts), the outcome will be more beneficial than the outcomes of using any single medium out of these.

The research methodology of this research is action research in order to make the previous practice better (practice to do forecasting). Action research is also known as the spiral of steps because it involves different action cycle in it and new action cycle is based on the revised idea of previous action cycle and so on. Similarly, in this research, 2nd workshop used the revised/improved idea from the previous action cycle of 1st workshop.

Interviews and two workshops were arranged where observation and surveys (questionnaires) of the workshops were considered as a method of data collection. A semi-structured interview and several unstructured interviews (during meetings) were done. The participants of these workshop surveys (the surveys were based on questionnaires) were the experts of both labs; Online Learning Lab's experts in the 1st workshop and Future Work Lab's experts in the 2nd workshop. These surveys were done to see how useful this tool will be in the future. This research was concerned with the observation of these experts and questionnaires were filled out by the experts in the workshops. The time horizon for this research is cross-sectional. Because the data gathered during this research, is taken at one particular moment. And cross-sectional study always takes new sample of people every time it conducts. For example, workshops were organized, but each time, different participants participated in it. So, one workshop was done with each lab.

In order to understand the trend research and forecasting technique, which is inquire by many organizations, there is a need to have trend research and forecasting methodology. In order to pursue with this research, it is decided to do action-based research in order to help and to make the previous practice better.

3.1.1 Why Action Research?

Action research is a way of conducting social research collaboratively that satisfies hard and demanding scientific requirements and promote democratic social change simultaneously (Greenwood & Levin, 2007). (Thomas, 2017) described action research as a research which is undertaken by practitioners collectively for helping to develop their practice. Action research is done usually at the same time as performing and practising and its basic aim is to change and focus on problem-solving in any appropriate way (Thomas, 2017). In action research, a concept from science is used to address the challenge within the context. As described previously that the problem at many organizations is to have some sort of methodology which can help organizations in doing trend research and forecasting.

As described above, in this research, 2nd workshop used the revised/improved idea than the previous action cycle of 1st workshop. Hence, this research is improving the previous practice, which is the exact aim of action research.

3.2 Case Selection

This research was performed at Centre for Innovation of Leiden University. There are different labs at Centre for Innovation which includes Online Learning Lab, Future Work Lab and Humanityx. In these labs, number of experts work and collaborate with respect to their fields. In Online Learning Lab[2], the focus is to explore new way of learning, exploring new forms of education and new target groups within digital technology. Whereas Future Work Lab is more about the future of work like what are the future jobs, and it works in research and innovative projects through team-based project management methods. On the other hand, Humanityx[3], which is more focused to support organizations in the peace, justice and humanitarian sectors, and help them to adopt digital innovations to increase their impact.

Centre for Innovation is deeply working with technological and societal trends, and also impacts of these trends in the future. And the missing thing with Centre for Innovation is to have some sort of methodology which can help them in trend research. That is why this research is about a methodology: a method of using different mediums of information like database knowledge, social media and expert's knowledge to see how far it can to forecast.

3.3 Data collection and Data sources

In order to perform this research, different data sources were used to understand better context of Centre for Innovation. Following are the data collection methods and sources:

3.3.1 Interviews

Semi-structured interview and unstructured interviews (stand-up meetings) were conducted during this research. Interviews were done with CIO of Leiden University, program manager of Centre for Innovation and leader of Future Work Lab.

[2] https://onlinelearninglab.centre4innovation.org/
[3] https://www.humanityx.nl/

3.3.2 Experiment through the Workshops

Two workshops were organized in order to test the methodology of this research. These workshops were conducted with Online Learning Lab and Future Work Lab of Centre for Innovation.

3.3.3 Questionnaires

After the workshops, questionnaires were used in order to get the feedback on the methodology. Qualitative questions were used during in the questionnaires.

3.4 Data Analysis

Due to the nature of this research, which is qualitative, the interview was fully transcribed, and all the feedback from the workshops were also written down in order to be used in section *Data* and *Evaluation*.

4. Action Research

This section is going to follow a normal structure of action research consisting of diagnosing, action planning, action taking and evaluating. Several books and papers like (Mcniff & Whitehead, 2002), (Mcniff & Whitehead, 2010) describe and explain about action research and its structure. (Stettina & Smit, 2016) also used the same methodology which describes the similar and normal structure of action research.

4.1 Diagnosing

The analysis started in April 2018 until October 2018. It all started from playing around with the text-mining tool in order to get familiar and to understand the tool better. The tool used for text mining was "KNIME" – ***K**o**N**stanz **I**nformation **M**in**E**r,* which is an open source data analytics and integration platform. It does not require the knowledge of any programming language prior for using this tool because it's a GUI (Graphic User Interface) platform with drag and drop options, but one needs to have basic knowledge of text mining, like how text-mining works. During the testing with the tool, trend reports of several consultancies, educational organizations and economic organizations were used in order to detect the trends and to have better understanding with the tool and trends.

Meetings were arranged before proceeding with the analysis for Centre for Innovation with the people who work with Centre for Innovation. Information was gathered on the first hand to know about the current system and methods of Centre of Innovation's Labs of Leiden University and how they are currently proceeding with their method. Centre for Innovation know some of the trends and also know in depth. But there is a need to make good understanding on how to distribute the resources like what to implement and what not etc.

During the interviews and meetings with the people working with Centre of Innovation, we concluded that we need a better workshop format to do this distribution.

4.2 Action Planning

After the project briefing, related work was considered in order to do trend research and forecasting through text-mining technique. Why was text-mining technique considered? That is explained in the *Literature section*. Many of research papers related to our research were analysed and discussed to see the literature gap and the best way to do this trend research. From those research papers, (Porter, 2009) and (Kayser & Blind, 2017) were mainly focused and were considered as the main inspiration for this research.

After analysing research papers related to the trend research, KNIME (a text mining tool) was selected in order to proceed with the text analysis as it was easy to work with this GUI tool, especially when one is not expert in writing codes for extraction and analysis. Steps taken within this KNIME tool are further explained in the *section – 4.2.3 – Text Mining*. Outcomes were discussed in the meetings simultaneously while working on the text-mining part.

After the meetings with the people related to the trend process at CFI/LU (Centre for Innovation / Leiden University), several things were required to be adjusted. Source reports for the text-mining techniques were updated. New search terms for Twitter were added in order to see the better information for Online Learning Lab. Similarly, during the interview with the CIO of Leiden University, several things were more focused like discussion of experts.

Research process plan for this research is shown in the *Figure 5* down below:

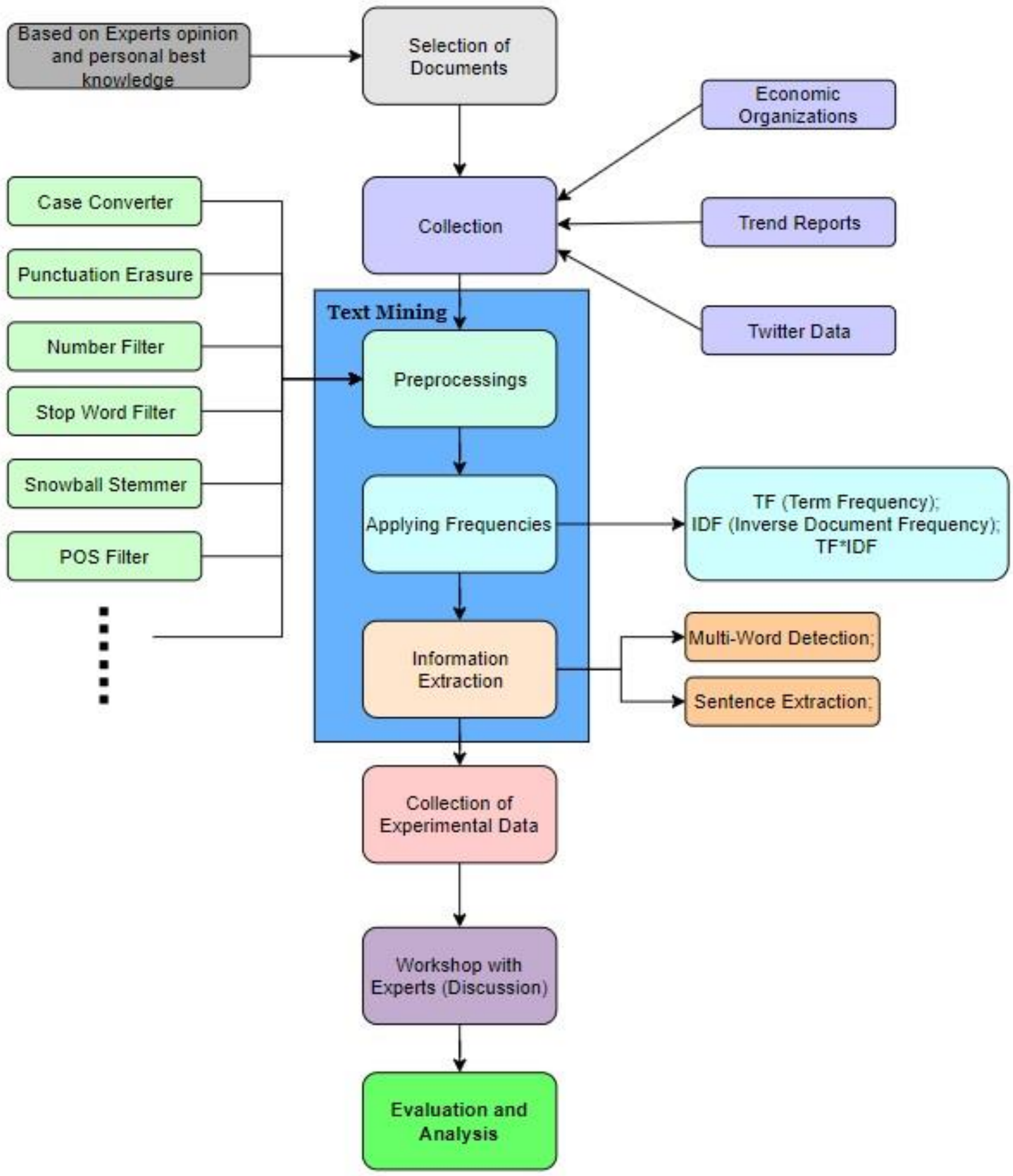


Figure 5: Research Process Plan

It begins with the selection of documents. Number of documents were selected for this research. Extracting information through multiple mediums was taken into account so that this research should comprise on the information from maximum number of reliable sources to have better understanding of future trends. A lot of research was done beforehand while considering the sources like asking from different experts etc.

4.2.1 Selection of documents

This is the first step of the study, to begin with the research. There are a lot of documents available through several organizations and companies, which are mostly accessible through internet. The main thing considered in this step is to select those documents which are authentic and keeping those organizations into consideration who have well reputation in the market for its authentication. These documents were chosen on the best level of personal experts' knowledge, related to the field.

4.2.2 Collection

This section describes which data sources were selected to extract information. This section is mainly divided into three categories of the collection set.

4.2.2.1 Intercontinental Economic Organizations

Reports from the organizations which used economic methods to see and operate the organizations, were used in this research. Purpose of using economic reports from these organizations to get to know the economic trends as well as technological trends. The economic reports used in this research are the reports from European Trade Unit Institute (ETUI) and Organization for Economic Co-operation and Development (OECD).

4.2.2.2 Consultancy, Educational and Future Work Trend Reports

Consultancies and Educational organizations are one of the priors who create and publish trend reports. As the nature of their work, consultancy companies have to deal with trends more in order to provide expert advice to individuals and organizations. Professional experts create these reports with collaboration of other experts and that's why it is worth taken source in order to get deep knowledge about the future trends. Some of the reports used in this research are reports from consultancy organizations such as Deloitte, Horizon, Accenture, PWC, Gartner, Forrester etc. On the other hand, educational departments who are more into the future of education are keen to note the trends related to education and technology in order to combine both educational and technological trends to get maximum advantage out for them. For that Horizon report was considered.

These reports were categorized into two groups, separate for Online Learning Lab workshop and Future Work Lab workshop. These sources are mentioned in the sections below.

Online Learning Lab Trend Reports

For the Online Learning Lab reports, the total collection size of reports (in words) was 158581 words. Following were the sources which were considered into account. These sources were suggested from the experts of Online Learning Lab and up to the best of our knowledge.

Report Name	Category	Publisher
Horizon Report 2018 **Higher Education Edition** **Brought to you by EDUCAUSE**	Education	NMC Horizon, EDUCAUSE NMC Beyond the Horizon
The Top 10 Strategic Technology Trends for 2017	Research and Advisory	Gartner
ELEARNING MARKET TRENDS AND FORECAST 2017-2021	Learning Management System	docebo
Hype Cycle for Education, 2018	Research and Advisory	Gartner
Accenture Technology Vision 2018	Consultancy	accenture
Tech Trends 2018	Consultancy	Deloitte.
Foresight Brief **Shaping the world of work in the digital economy**	Intercontinental Economic Organization	etui. european trade union institute

Tech breakthroughs megatrend: how to prepare for its impact	Consultancy	
Skills for a Digital World	Intercontinental Economic Organization	Organization for Economic Co-operation and Development

Table 4: Sources (Reports) used for Online Learning Lab

Future Work Lab Trend Reports

For Future Work Lab, the size of reports (in number of words) was 177779 words. Following were the sources which were referred by experts, were used for the text-mining. *Table 5* shows the sources for Future Work Lab reports.

Report Name	Category	Publisher
Future of work 2030 *A wake-up call for organisations, people and government*	Consultancy	
Digitally-enabled automation and artificial intelligence: *Shaping the future of work in Europe's digital front-runners*	Consultancy	
THE FUTURE OF SKILLS EMPLOYMENT IN 2030	Mix	
The Future of Jobs *Employment, Skills and Workforce Strategy for the Fourth Industrial Revolution*	Intercontinental Economic Organization	

Table 5: Sources (Reports) used for Future Work Lab

4.2.2.3 Twitter

Large amount of data is present on Twitter and millions of users use this platform for discussions and share their opinions. Almost every invention news break through this social media platform (Twitter) which is often neglected in the process to generate knowledge for forecasting. As described in the literature, it is much important to consider social media as a medium of information which can be used to forecast. The research on Twitter is done on the basis of specifics terms which were suggested by the experts and based on the best of personal knowledge related to the digital learning and future of work. Twitter API was used to extract tweets and Twitter data.

4.2.3 Text Mining

After collecting the required documents, information was extracted through text mining from these documents. The tool used for text mining is "KNIME" – KoNstanz Information MinEr which is an open source data analytics and integration platform. It is not required to have knowledge of any programming language for using this tool for use because it's a GUI (Graphic User Interface) platform. After extracting different files (documents) through KNIME tool, there was a need for preprocessing.

There were several steps taken while adding, analysing and taking output of the text data which includes:

- Input/output (Tika Parser);
- Transformation (Strings to document);
- Enrichment (POS Tagger);
- Preprocessing (Case converter, Punctuation Erasure, Number filter, stop word filter, etc.);
- Frequencies (TF, IDF);
- Write (for writing tables);
- Convert & Replace (Math Formula – TF-IDF);

Cleaning the data was one of the main things in this process – extracting the clean and meaningful data. For that preprocessing is explained a little more in the section down below:

4.2.3.1 Preprocessing

After extracting documents, preprocessing was done on the documents. This was done through KNIME. There are several steps which were done during preprocessing. Some main steps are as follows:

(1) Case Converter
Case converter is used to convert case of alphabets to upper or lower. It is used when similar case is required for all the documents.

(2) Punctuation Erasure
Punctuation erasure is used to remove all sort of punctuation from the text. Punctuation is erased while cleaning the data. Question marks, full stops, hyphens and other similar punctuations get remove from this.

(3) Number Filter
Number filter is used to filter all sort of number from the texts including decimal separators like "." (point) or "-" (hyphens). There is an option to filter those term which contain only one digit in it.

(4) Stop Word Filter
Stopword filtering is used in order to remove specific words from the results. For example, there is no need of words like "which", "the", "is", "at" etc. in the result because it is not related to the required data.

(5) Stemming
Stemming is the process of reduction of those words which either derived or inflected. E.g. Swimming to swim, running to run etc. Stemming was done in order to restrict similar words in results. All these derived or inflected words were converted into their stem word in order clean the data. But several complexities were raised from this as sometimes several words get stemmed which were not making any sense (it was hard to understand). So later, this step was skipped.

(6) POS Tagging
Part Of Speech (POS) tagging is the process of adding tags with each word like this word is noun, verb, adjective etc. KNIME POS tagging assign part of speech tags, such as person name, locations, organization name, chemical structure, bio-medical name entities etc. Each word is assigned a tag value. For example, for person, tag will be "NE", which is for name entity. POS tagging is used assign each term of document a POS tag. By using "Tag Filter", we can extract specific assigned tags (words).

4.2.3.2 Frequencies

There are two frequencies which were applied. These are as follows:

(1) TF
TF stands for Term Frequency. TF refers to number of times a term has occur in the document or in text which depicts how important the word is. And it will give the most repeated words and consider them as important.

(2) IDF
IDF stands for Inverse Document Frequency which works a little different to TF. There might be some words like "the", "and" which are occurring in the document quite frequently and TF will consider them as important words. But IDF will give these common words a low weight and concentrate on less common words.

IDF is defined as,

> *"a statistical weight used for measuring the importance of a term in a text document collection. The document frequency DF of a term is defined by the number of documents in which a term appears. In mathematical term, IDF is described as:*
>
> $$IDF = log_2 \frac{N}{DF}$$
>
> *Where N stands for Number of Documents and DF stands for Document Frequency[4]".*

(3) TF-IDF
It is the combination of both frequencies which was used in this research to get more un-common words out of the text. It assumes how important the word is in the set of documents and how often does this occurs.

4.2.3.3 Information Extraction

Some of the information can be extracted through the frequencies which are mentioned in above *section –4.2.3.2—* i.e. TF and TF-IDF. One can easily see that which term has occurred more often in the text documents and how much important that term is, in the whole set of documents. Other than these frequencies, there are several other ways through which information was extracted. Those are described below:

4.2.3.3.1 Term Co-occurrence

Term co-occurrence counter is used to see which terms have occurred together more often. It is a method through which it counts the number of co-occurrence in the document(s). As mentioned by KNIME platform, who have created these nodes, that these co-occurrence terms are not in ordered. So, term one can follow term two and term two follows term one, is considered the same.

4.2.3.3.2 Sentence Extractor with Row Filter

Sentence extractor can extract the sentences from a given documents and separates each sentence. After splitting these sentences, row filter was used to filter specific sentences. Several terms and multi-words were used to filter the rows. For example, "Future skills", "artificial intelligence" etc.

4.2.3.3.3 Gathering Information

After extracting the information through the above methods; frequencies, term occurrence and sentence extractor, all the extracted information was then converted into several statements and questions which were later discussed by the experts in the workshops.

---

[4] Inverse Document Frequency https://doi.org/10.1007/978-0-387-39940-9_933

4.3 Action Taking

Before testing the methodology, several meetings were held with the different people related to Online Learning Lab and Future Work Lab. An Interview with the CIO of Leiden University was conducted about the current trend selection process at Leiden University. Meetings were arranged with program manager of Centre of Innovation, experts of Online Learning Lab and head of Future Work Lab. Other than these meetings, working on text-mining tool was carried out throughout the whole process.

4.3.1 Timeline

Following is the timeline of the trend project, shown in the *Table 6* below:

DATE	ACTION TAKEN
15TH MARCH, 2018	Briefing on the trend project
3RD APRIL, 2018	Beginning with the project
20TH APRIL, 2018	Discussion on the usage of text-mining tool "KNIME."
15TH MAY, 2018	Discussion on the implementation plan of text-mining techniques and alternative methods
29TH JUNE, 2018	Discussion on the outcomes of text-mining method
23RD JULY, 2018	Discussion on the implementation plan of the project
3RD AUGUST, 2018	Discussion on the outcome results of text-mining tool
17TH AUGUST, 2018	Feedback on the text mining outcomes
31ST AUGUST, 2018	Meeting with the program manager of Centre of Innovation and expert of Online Learning Lab, getting their remarks, opinions and feedback on the project.
5TH SEPTEMBER, 2018	Discussion on the implementation plan and arranging the project according to the previous feedbacks
14TH SEPTEMBER, 2018	Discussion on organizing 1st workshop with Online Learning Lab and procedure.
20TH SEPTEMBER, 2018	Testing workshop (full demo) with program coordinator of Centre for Innovation
24TH SEPTEMBER, 2018	Interview with the CIO of Leiden University regarding trend process.
27TH SEPTEMBER, 2018	1st Workshop with the Online Learning Lab
3RD OCTOBER, 2018	Discussion on the outcomes of 1st workshop and planning for the 2nd workshop
12TH OCTOBER, 2018	Discussion and organizing plan for 2nd workshop with Future Work Lab
19TH OCTOBER, 2018	2nd workshop with the Future Work Lab
26TH OCTOBER, 2018	Result discussion of 2nd Workshop

Table 6: Project timeline

4.3.2 Workshop

Two workshops were organized with the collaboration of Centre for Innovation, Leiden University. Experts of Centre for Innovation took part in these workshops. They discussed the questions and statements which were provided to them in the workshops. Those questions and statements were based on the text-mining output.

While arranging the 1st workshop with Online Learning Lab, 2nd workshop was first not intended. Several things were changed for the 2nd workshop. Mostly these were suggestions from the 1st workshop to improve the next workshop. For the 1st workshop, several sources were taken into account like along with the educational reports, technological and skills reports were also included. 1st workshop was monitored by a coordinator of Online Learning Lab who was noticing each and every aspect of the workshop including the workshop's structure and the content used in it. Workshop structure was improved in the 2nd workshop and also the content shown or provided to the experts of Future Work Lab was much more and better than the 1st workshop. Similarly, 2nd workshop was also monitored, and several things were noticed to make the future workshop better than the 2nd workshop.

A cloud-based response system "Socrative" was used during workshops. All the questions were posted on Socrative platform first, and then participants had to log-in into a "session" through a link, which was provided to them during the workshop to participate in it. In the workshops, by the end of question session, answers were shown to participants (experts) without mentioning the name of the participants to keep their identity unknown from other participants. Participants were able to see all the answers of other experts. Out of all the answers, common answers were discussed first, and unique answers / less common were also debated later to see the other experts' perspective regarding those answers.

4.3.2.1 Workshop with Online Learning Lab

One of the workshops was performed with the Online Learning Lab of Centre for Innovation of Leiden University. Experts of Online Learning Lab participated in this workshop where they discussed, debated and provided their answers to the several statements and questions. The content of the workshop was gathered through the same process which is described in *Text Mining section*.

4.3.2.1.1 Online Learning Lab Workshop Design

First, a meeting was held with the experts. They recommended several documents and terms for Twitter search. The text mining process was run over those documents and given Twitter terms, which resulted into that information which was related to Online Learning Lab.

Figure 6 shows the process for Online Learning Lab.

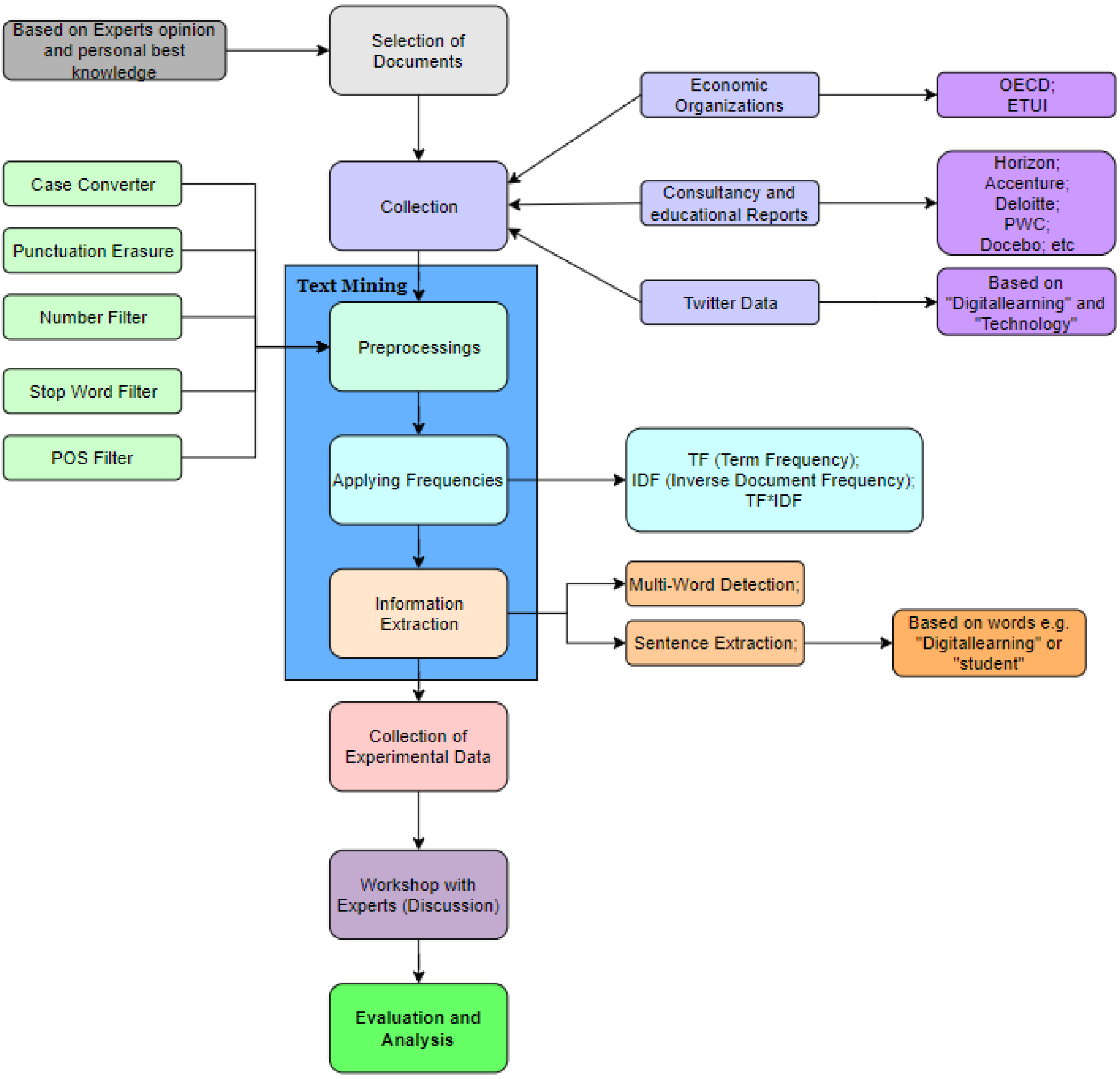


Figure 6: Process for Online Learning Lab

Workshop was of 2 hours with the experts of Online Learning Lab. After the introduction of the project, working of the method (Trend reports + Twitter + Experts) and text mining tool were discussed with them briefly. Statements and questions were discussed on the "Socrative" platform with the experts. Complete playbook for Online learning Lab workshop is mentioned in *section – Appendix G –*.

All the experts were log-in through their devices (cell phones/laptops) to Socrative portal. List of questions were presented to them. Before answering, they discussed about that statements/questions with the other experts first, and then answered those questions separately. Several statements were there in the form of questions, which they discussed first and answered later.

By the end of this activity, questionnaires were given to them where they were asked about the method (text mining + twitter + experts discussion) and about the workshop, which are discussed in *section 4.3.3.3*.

4.3.2.2 Workshop with Future Work Lab

Other workshop was conducted with the experts of Future Work Lab, which mostly focuses on future of work and jobs. As described in the research process, several recommendations were asked from the experts about the documents and reports for the Future Work Lab. The text mining process was run over those documents and on Twitter term (which were recommended by the experts), which resulted into that information which was related to Future Work Lab. This information was then converted into several statements and questions, which was discussed in the workshop.

4.3.2.2.1 Future Work Lab Workshop Design

For Future Work Lab's workshop, same procedure was applied in order to get the information beforehand. A meeting of similar type was held with the expert of Future Work Lab. Experts of Future Work Lab recommended several documents and terms for twitter search. The text mining process was run over those documents and Twitter (based on recommended terms), which resulted into information, that was related to Future Work Lab.

Figure 7 shows the process for Future Work Lab.

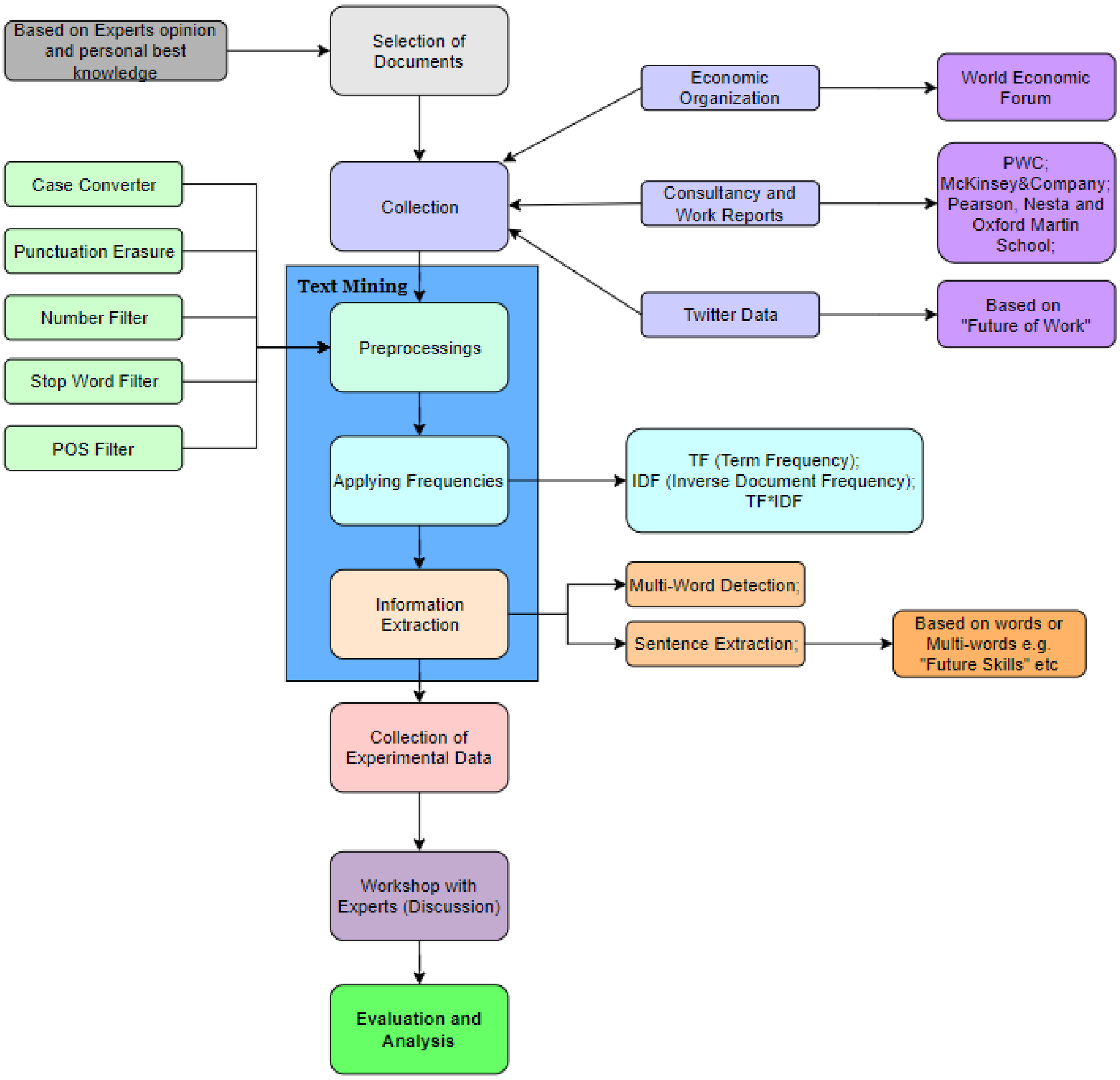


Figure 7: Process for Future Work Lab

2nd Workshop was also of 2 hours with experts of Future Work Lab. After the introduction of the project, working of the method (Trend reports + Twitter + Experts) and text mining tool were also part of discussion. Discussion about the statements and questions which were extracted through text-mining tool was done on the online platform of “Socrative”. Complete playbook for Future Work Lab workshop is mentioned in *section – Appendix H –.*

All experts were log-in through their devices (cell phones/laptops) to Socrative portal. List of questions were presented to them. This workshop was changed from the 1st workshop. In 2nd workshop, experts first answered the questions and did the discussion later.

By the end of questions activity, questionnaires were given to them where they were asked about the method (text mining + twitter + experts discussion) and the workshop, which are discussed in the *section 4.3.3.4*.

By the end, they were shown the answers which they had given during the “Socrative” session. Their identity was kept secret so that no one knows which expert has answered what. In discussion section, common answers (answers which got most votes) were selected first, and experts discussed those answers. After that, unique or less-common answers were debated to see other participants’ opinion on those answers.

4.3.3 Data

This section describes the results of data collection. Results were collected through text-mining tool, through interview with the CIO of Leiden University and workshops with experts of Online Learning Lab and Future Work Lab. This section is divided into four section: Text-mining results, outcomes of Interview, 1st Workshop results with Online Learning Lab and 2nd Workshop results with Future Work Lab.

The first section, text-mining results, covers the text-mining results which were generated through trend reports and Twitter. This section includes which terms and trends are more prominent and likely to occur more often in the future. This is based on trend reports and Twitter data. Other than this, it also describes which information is extracted through this tool. The second section describes the outcomes of interview. The interview is about the trend selection process at Leiden University, Centre for Innovation and recommendation in the process. The transcription of the interview is included in *section – Appendix A –*. The third section describes the results of the workshop with Online Learning Lab. The fourth section describes the results from the workshop with Future Work Lab.

4.3.3.1 Text-mining Results

The text-mining tool was used to analyze the texts of the reports, Twitter data and extract information out of them for trend research. Text-mining tool was used in order to extract and analyse the data related to digital learning and future of work. In the section below, text-mining results for digital learning (for Online Learning Lab) and results for future of work (for Future Work Lab) are described separately.

4.3.3.1.1 Text-mining results for Online Learning Lab

Using the sources which are mentioned in *Table 4 of section 4.2.2.2*, following word cloud was formed, which is indicating keywords. Size of the terms show the weight of that term with respect to other terms.

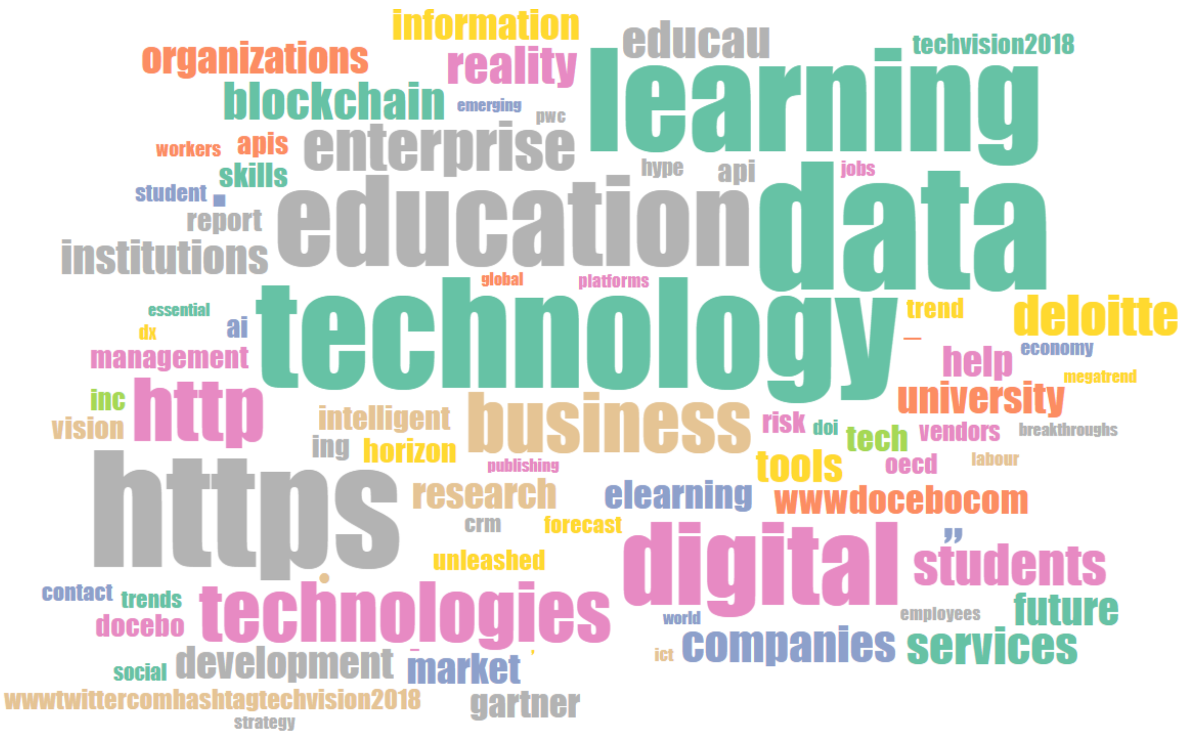


Figure 8: Word Cloud from Digital Learning and Trend Reports

From the above word cloud, several keywords (which were more technological trends) were extracted. Those keywords were extracted by getting all the information about the documents which include document name, its writer, publisher, pages, keywords, terms etc. Out of all the terms which are in above figure and adding those terms/trends which were occurring more in Twitter search, were combined. At the end, these (*figure 9*) were the trends which were provided to Online Learning Lab experts.

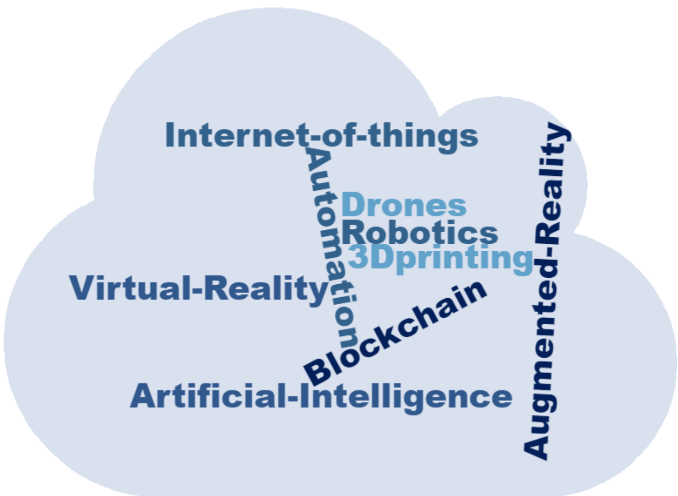


Figure 9: Trends provided to Online Learning Lab Experts

Artificial Intelligence	**Blockchain**	**Virtual Reality**
3D printing	**Robotics**	**Drones**
Internet of Things	**Automation**	**Augmented Reality**

After extracting the trends, to get more meaning out of these reports, multi-word detection was used to get those terms which are occurring quite often next to each other. Out of all the co-occurrence terms, several terms were extracted manually which were having higher co-occurrence number. These Multi-words were provided to experts during workshop. In the *Table 7*, some higher co-occurring terms are mentioned which were extracted. They were provided to the experts of Online Learning Lab.

Multi-words extracted (co-occurrence)	
Patent application Drug delivery Technology applications NATO technology Application Technology Economic Development Delivery Systems Landscape technology Business Skills Learning students	Intelligent vision Drug system Enabling technologies Barriers countries Management skills Digital economy Enterprise technology E-leadership skills Patent technologies Intelligent Technology Delivery Nano-enabled

Table 7: Multi-word table from trend reports

4.3.3.1.1.1 Online Learning Lab Twitter result

After the meeting with experts, several terms were suggested in order to look for Twitter results. As prior, only term “E-learning” was being used. But on expert’s recommendation, several other terms were used such as “digital learning” and “blended learning”. As these terms are new and are mostly in use for “Online Learning” at the moment. Following were the tweets, which were extracted by using the terms mentioned above and were provided to the experts. These tweets were selected on the basis of high number of “favourites” and “retweets”.

Extracted Tweets

1. Digital learning market trends forecast driven mobile smartphones, social networking, cameras learn content, video sharing moocs (massive open online courses) recommendation.

2. Moocs corporations established academic universities student paying fees accessing online courses.

3. China and India are higher rank ambient insight e-learning countries.

4. According to surveys, companies are adopting tools that China is using for employee skills development and recruitment.

5. E-learning market continuous evolve.

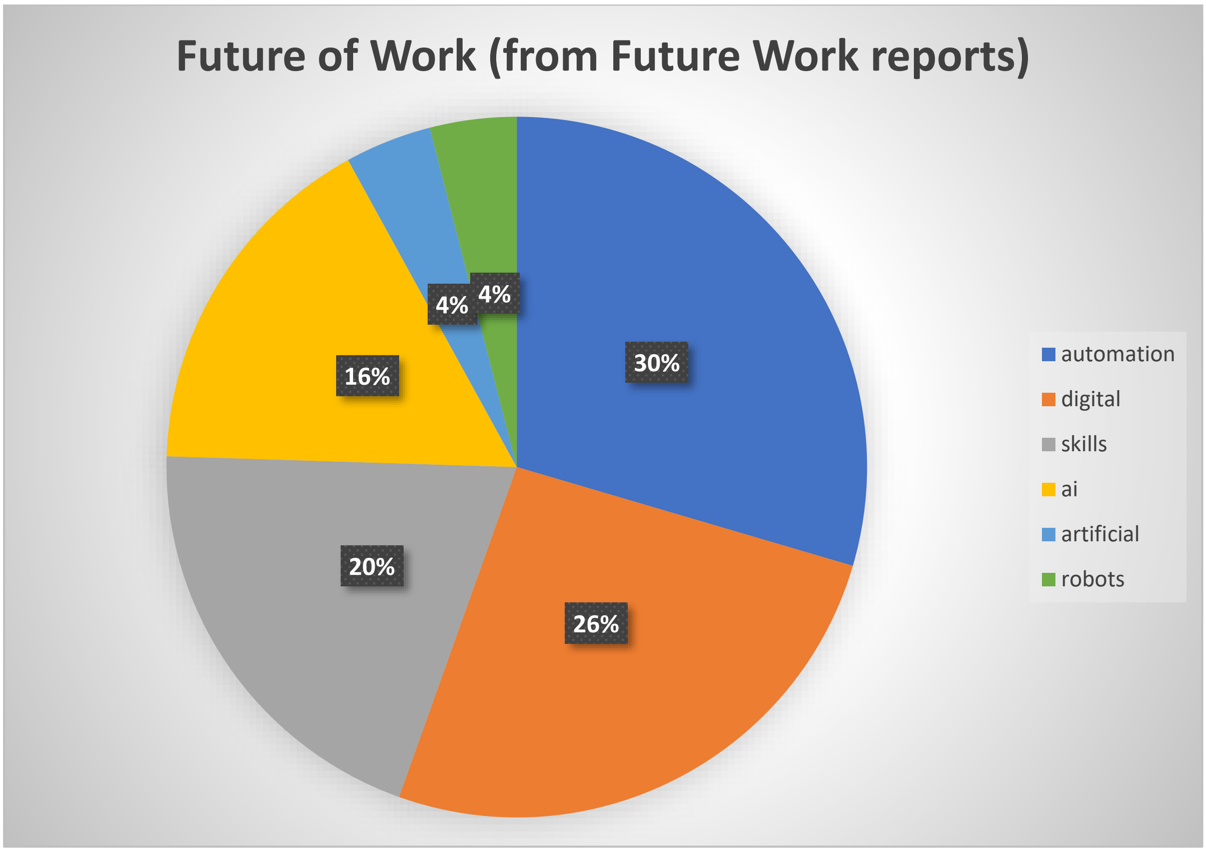
4.3.3.1.2 Text-mining results for Future Work Lab

Results were gathered in different forms. Overall “Word-Cloud”, which depicts the important and common terms in future work reports, is shown in *Figure 10*.

Figure 10: Word Cloud from Future Work reports

In *Figure 10*, these are the terms which occur more often and are common in between all these reports. By extracting trends out of these terms, “Term Frequency” was applied and then manually these trends were extracted. From Blue to red colour, and size shows higher weight of the term (Blue has the highest weight and red shows the low).

In the *graph 3* down below, it is shown that which trend has occurred more often in these future work reports.



Graph 3: Future Work Trends (Based on Reports)

This above graph is based on the Term frequency values. Images of some excel sheets, for term frequency, are added in *section – Appendix B –*. Some main trends have the following TF values which show percentage in the above graph and their values are shown in the *Table 8* below:

Trend	Term Frequency Values
Automation	0.01409
Digital	0.01235
Skills	0.009552
Artificial Intelligence	0.007874
Artificial	0.001906
Robots	0.001906

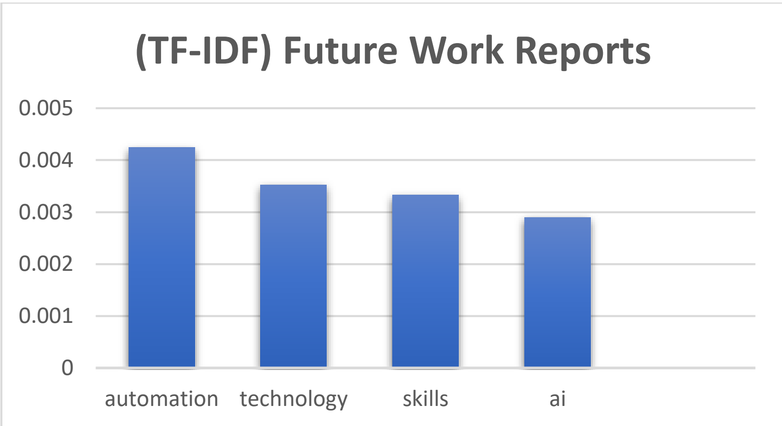
Table 8: Future Work Trends Value (TF) from reports

To see the important trends, which are not only in one report but equally important in all the reports, TF-IDF (Term Frequency-Inverse Document Frequency) was applied on the documents. Images of some excel sheets for TF-IDF are added in *section – Appendix C–*. Some main trends have the following TF-IDF values which are shown in the *Table 9* below:

Trend	TF-IDF value
Automation	0.004242
Technology	0.003525
Skills	0.003328
Artificial Intelligence (AI)	0.002897

Table 9: TF-IDF values (based on Future Trend Reports)

From the above TF-IDF table, following graph (*Graph 4*) is produced in order to see which trend is more likely common and important in all these documents:



Graph 4: Future Work Trends (TF-IDF) (Based on Reports)

After extracting the trends and important keywords, these terms (keywords, trends) were furthered divided to make them more meaningful. So, in order to make them meaningful and to get more sense out of these terms, multi-word detection (co-occurrence terms) were applied. *Table 10* shows terms, which are occurring more often together in these reports. These terms are manually taken on the basis of their relevance to Future Work Lab.

Multi-words extracted (co-occurrence)	
Future jobs Employment skills Future skills Emerging markets Basic infrastructure Communication technology Families job Communication information Employment skills Gender gap Emerging class Cloud technology Demand skills	Middle class Mobile internet Climate change Cloud internet Architecture engineering Media entertainment Gap wage Financial services Future management Innovation strategy Change management Job rotation Recruiting women

Table 10: Multi-word table from future work reports

These multi-words were extracted from the co-occurrence table. Some images of multi-words table, which were extracted, are included in *section – Appendix D –*.

After extracting multi-words, sentences were extracted to get deep meaning about the trends. Those sentences were based on the extracted trends and multi-words.

Several sentences were extracted basis on the topmost trends and topmost multi-words. In an image (*Figure 11*), as an example, sentences like these were extracted on the basis of "Future Skills".

Filtered - 2:58 - Row Filter

File Hilite Navigation View

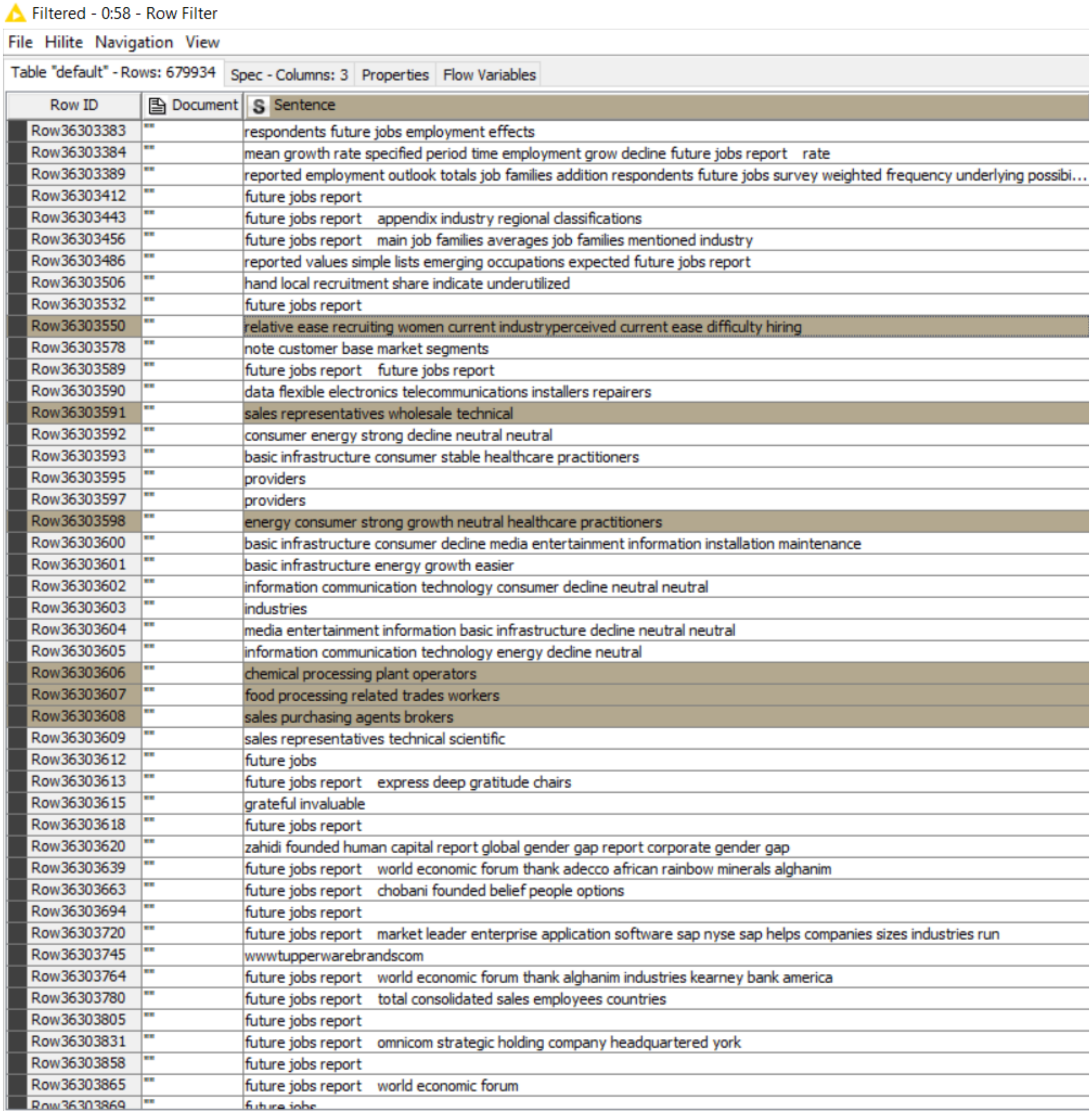
Table "default" - Rows: 1090647 | Spec - Columns: 3 | Properties | Flow Variables

Row ID	Document	Sentence	Number…
Row4768572	""	future skills employment finding robust range specifications future understood widest sense tests controls demographic industry climbing policy agenda	18
Row4768590	""	provide overview trends following display description trends analysis fits wider	34
Row4768612	""	future skills employment demographic change	19
Row4768636	""	future skills employment urbanisation	10
Row4768659	""	future skills employment political uncertainty	33
Row4768666	""	future skills employment approach	15
Row4768698	""	model predict future occupations defined hotspots demand associated existing occupation	14
Row4768729	""	future skills employment table list onet features study	352
Row4768765	""	future skills employment figure sample page trends analysis workshops	15
Row4768769	""	recession oil shock	36
Row4768776	""	monitoring	38
Row4769063	""	machine learning	17
Row4769471	""	note	5
Row4769721	""	future skills employment future skills employment results	12
Row4769726	""	future skills employment	3
Row4769739	""	future skills employment figure plot following frey osborne distribution current employment probability	15
Row4769752	""	future skills employment table minor occupation probabilities future	19
Row4769754	""	education training	26
Row4769762	""	future skills employment table minor occupation lowest probabilities future	20
Row4769765	""	transportation workers future skills employment results support importance future technological change	11
Row4769770	""	future skills employment	3
Row4769785	""	management directors senior officials	50
Row4769816	""	occupation probability experiencing fall workforce share detailed	30
Row4769843	""	building finishing trades agricultural related trades business finance related associate professionals caring personal services future skills employment table suggestive interpretations	20
Row4769850	""	future skills employment	3
Row4769869	""	future skills employment figure examples share employment extrapolations uk occupations	10
Row4769877	""	future skills employment table percentage pp differences trend extrapolation predictions probabilities future demand minor occupation level	25
Row4769884	""	teachers instructors health diagnosing treating practitioners future skills employment table continued	37
Row4769889	""	installation maintenance repair occupations construction trades workers drafters engineering technicians mapping technicians electrical electronic equipment mechanics installers repairers life scient…	78
Row4769899	""	future skills employment	3
Row4769902	""	table volume receiver operating surface model uk	29
Row4769915	""	occupations business finance related associate professionals engineering professionals assemblers routine operatives animal care control services quality regulatory professionals housekeeping rel…	186
Row4769925	""	table relative rankings major occupation uk model trained expert	82
Row4769938	""	future skills employment	3
Row4769939	""	table ranking pearson correlation importance onet variables future demand	384
Row4769941	""	active learning	11
Row4769943	""	response orientation	19
Row4769967	""	future skills employment	3
Row4769968	""	uk	395
Row4769969	""	monitoring	11
Row4769971	""	glare sensitivity	19
Row4769982	""	future skills employment	3
Row4769987	""	figure relative importance skills abilities knowledge assessed pearson correlation	50
Row4769995	""	proportion proportion	9
Row4770008	""	future skills employment table major occupation ranked lists ranked top lowest ranked	487
Row4770023	""	table uk occupation ranked lists ranked top lowest ranked bottom onet features complementary soc title current complementary features features features	568

Figure 11: Future Skills sentence extraction

From the above *Figure 11*, these sentences are a result for future skills. All these sentences are related to the future skills, and out of all these sentences, important and useful information was manually extracted. As shown in the image, three rows are highlighted which give a clear and useful information. For example, for "Future Skills", text-mining tool has extracted the information like: "machine learning", "education training", "active learning" etc.

Similarly, sentences were also extracted on the basis of term "Future Jobs", and for that, below *Figure 12* shows the outcome results.



Filtered - 0:58 - Row Filter

File Hilite Navigation View

Table "default" - Rows: 679934 | Spec - Columns: 3 | Properties | Flow Variables

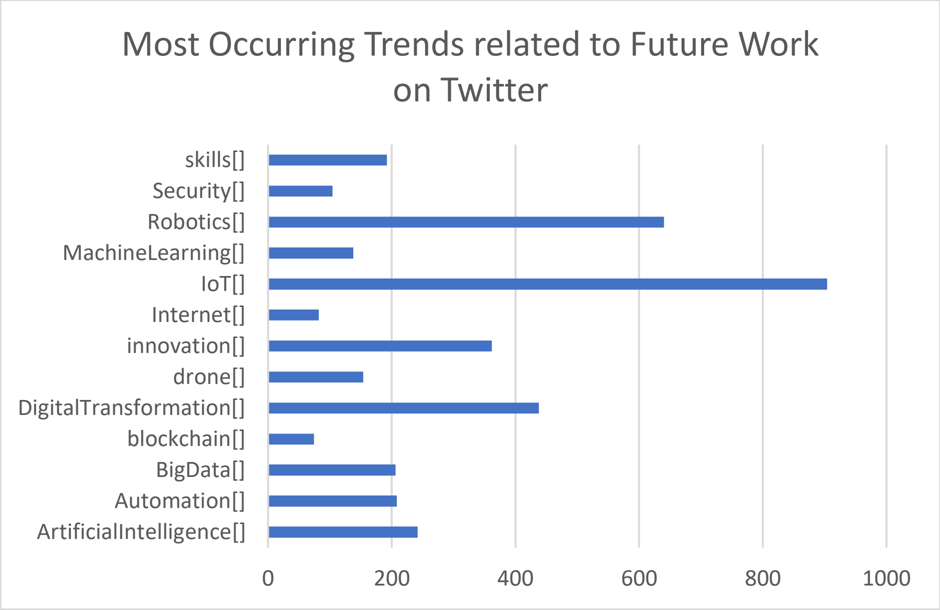
Row ID	Document	Sentence
Row36303383	""	respondents future jobs employment effects
Row36303384	""	mean growth rate specified period time employment grow decline future jobs report rate
Row36303389	""	reported employment outlook totals job families addition respondents future jobs survey weighted frequency underlying possibi...
Row36303412	""	future jobs report
Row36303443	""	future jobs report appendix industry regional classifications
Row36303456	""	future jobs report main job families averages job families mentioned industry
Row36303486	""	reported values simple lists emerging occupations expected future jobs report
Row36303506	""	hand local recruitment share indicate underutilized
Row36303532	""	future jobs report
Row36303550	""	relative ease recruiting women current industryperceived current ease difficulty hiring
Row36303578	""	note customer base market segments
Row36303589	""	future jobs report future jobs report
Row36303590	""	data flexible electronics telecommunications installers repairers
Row36303591	""	sales representatives wholesale technical
Row36303592	""	consumer energy strong decline neutral neutral
Row36303593	""	basic infrastructure consumer stable healthcare practitioners
Row36303595	""	providers
Row36303597	""	providers
Row36303598	""	energy consumer strong growth neutral healthcare practitioners
Row36303600	""	basic infrastructure consumer decline media entertainment information installation maintenance
Row36303601	""	basic infrastructure energy growth easier
Row36303602	""	information communication technology consumer decline neutral neutral
Row36303603	""	industries
Row36303604	""	media entertainment information basic infrastructure decline neutral neutral
Row36303605	""	information communication technology energy decline neutral
Row36303606	""	chemical processing plant operators
Row36303607	""	food processing related trades workers
Row36303608	""	sales purchasing agents brokers
Row36303609	""	sales representatives technical scientific
Row36303612	""	future jobs
Row36303613	""	future jobs report express deep gratitude chairs
Row36303615	""	grateful invaluable
Row36303618	""	future jobs report
Row36303620	""	zahidi founded human capital report global gender gap report corporate gender gap
Row36303639	""	future jobs report world economic forum thank adecco african rainbow minerals alghanim
Row36303663	""	future jobs report chobani founded belief people options
Row36303694	""	future jobs report
Row36303720	""	future jobs report market leader enterprise application software sap nyse sap helps companies sizes industries run
Row36303745	""	wwwtupperwarebrandscom
Row36303764	""	future jobs report world economic forum thank alghanim industries kearney bank america
Row36303780	""	future jobs report total consolidated sales employees countries
Row36303805	""	future jobs report
Row36303831	""	future jobs report omnicom strategic holding company headquartered york
Row36303858	""	future jobs report
Row36303865	""	future jobs report world economic forum
Row36303869	""	future jobs

Figure 12: Future Jobs sentence extraction

Several other sentences were also extracted on the basis of trends and multi-words, which are included in the *section – Appendix E –*.

4.3.3.1.2.1 Twitter results for Future Work

Other than this, text-mining tool was also used on Twitter for Future Work Lab. It was also done using KNIME tool, by using Twitter API. Outcomes were based on the search term like "Future of Work" which was recommended by the experts. In the *Graph 5* shown below, these trends came out as a result from Twitter:



Graph 5: Future Work Trends (Based on Twitter)

The *Graph 5* shows "Trends" on the y-axis (top to bottom), whereas trend occurred in how many tweets is shown on the x-axis (left to right). Numbers shown on x-axis represents the number of tweets. For example, "IoT" was mentioned and discussed in around 900 tweets when people on Twitter were talking about "Future of Work". Likewise, top trends related to future of work were extracted from the twitter. This data was collected on 2nd October 2018 and is limited time of data which changes time to time. This data was also part of the workshop with Future Work Lab.

4.3.3.2 Outcomes of Interview

An interview with the CIO of Leiden University was done to know how the current process of selecting trend is going at Leiden University. It was a semi-structured interview and was for short time because of the busy schedule of the interviewee. A complete transcription of interview is included in *section – Appendix A –*.

The interview was mainly focused on five questions. Primarily starting from the trend selection procedure at Leiden University. In the *figure 13* below of question 1, it is depicting the whole process like how and through which sources information is gathered from outside (outside of university). And then this information is shared with the CFI (Centre for Innovation) through meetings and casual talks with the people related to CFI. Meanwhile, university administration also gets updated by CFI about the trend research, which CFI is doing by themselves. CFI has their own maturity level, included in *section – Appendix F –*, which shows how CFI is currently working on trends. Because trend process at Leiden University is handle by CFI in the initial stages of the process, where they test the trends. And the later phases of process are looked by university administration, like which trend should they implement or focus more. Interviewee mentioned some hurdles too like there are several hurdles in the trend process like time and budget, which are an essential need for realization of any trend.

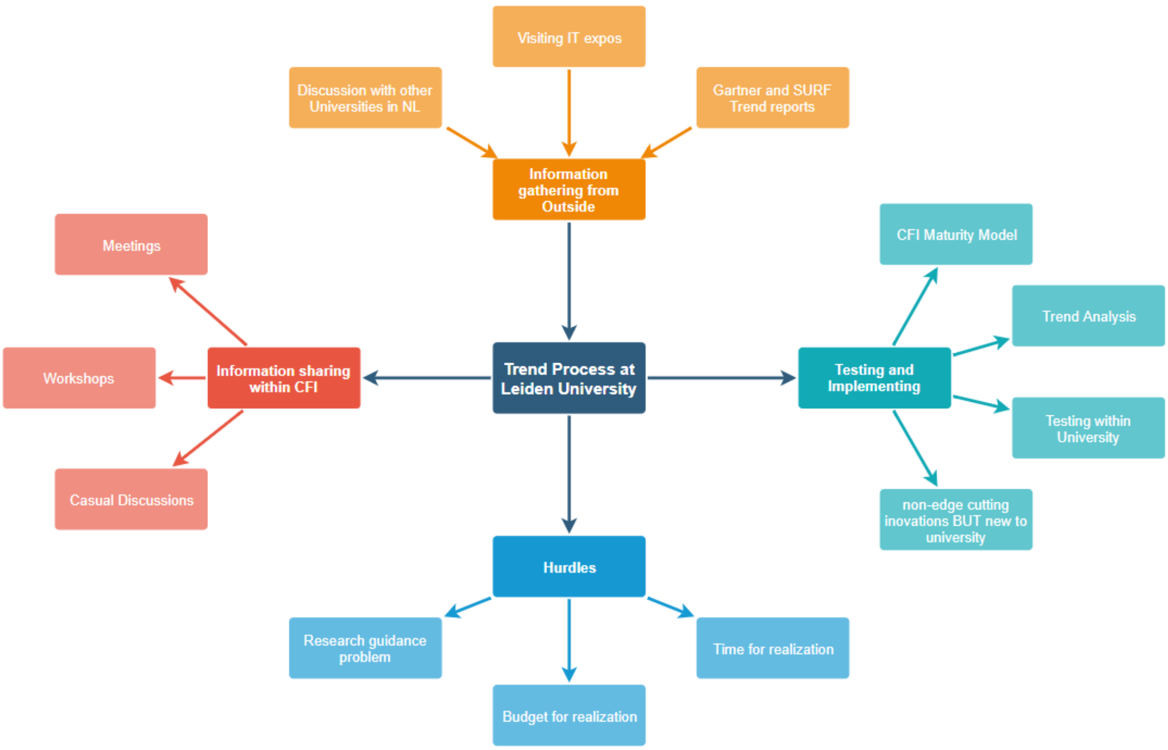
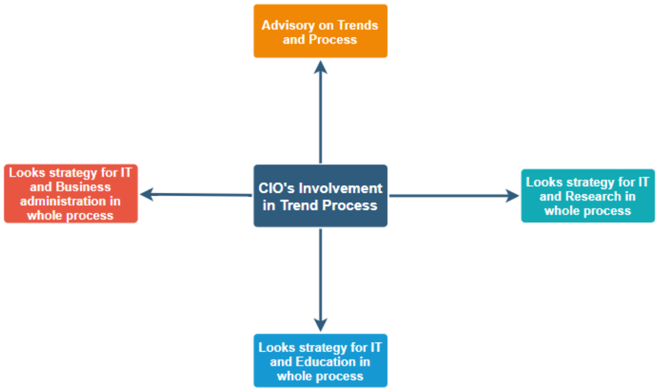


Figure 13: Q1 - Trend research process currently running in the university



Second question was related to interviewee involvement in all of this process as the CIO is the one who looks all the fields related to information technology. Similarly, the CIO of Leiden University does the same. In the *figure 14*, roles of CIO's of LU are shown.

Figure 14: Q2 – What's the interviewee's involvement in this whole process?

Third question was regarding about the things which are working well in the trend process, currently going at the Leiden University. Out of many things which are doing well in this trend process are shown in *figure 15*. These include; collaboration with the related people on personal basis and having informal chats and meetings which means discussions related to the process during lunch or drinks or some other time. Quoting from the interview, where interviewee said "*… we (he and previous director of CFI) chatted informal every month or six weeks on these innovations, (like) would it be nice for us to do as well (these innovations) or some funding. So, it's on the personal basis, and that's good…".*

Figure 15: Q3 – What's going well in this process?

Fourth question was comprised on what improvements can be made in this whole process. This was answered by the interviewee as to have "better structure" and "independency of people" who are involved in the process. Money and time realization are also one factor which is a hurdle. So, this is also added in the improvement to utilize these factors efficiently to remove this hurdle.

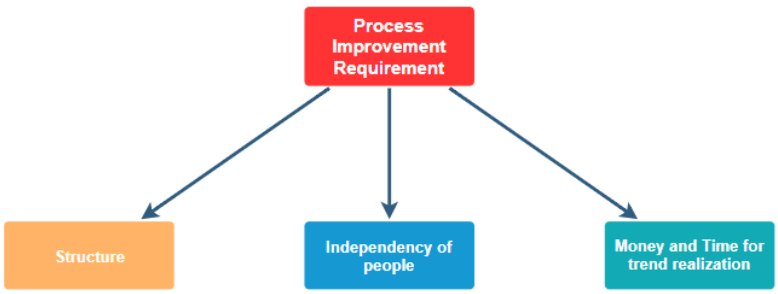


Figure 16: Q4 – What improvement can be made in this whole process?

The last question was about how a text-mining techniques or tools can help trend process. According to the interviewee, it is really good to use text mining tool as it helps to extract the information easily and then we don't have to go through "*rubbish*" data which means un-needed data. But on the other hand, there are complexities involving around this which includes that what type of sources are being used in the text-mining tool. According to interviewee, in case text-mining tool is used for the trend process of Leiden University, Leiden University's strategies and ambitions should be considered as its sources. But again, there are some possible issues in this. According to interviewee, as these strategies and other university-related documents, which should be consider as a source for Leiden University trend process in text-mining, are in Dutch language, and text-mining on Dutch language text is not as great as it is for English language. The other problem which is really important that which information is valuable? What are the criteria for the information to be valuable so that it can be extracted through text-mining tool because many people have different opinion on the information being valuable? And the last part to check the text-mining tool is effective or not is by testing on different groups. One group should be provided by the information extracted through text-mining tool, and other one is the simple group which follow daily news. Check, what are the differences between the outcomes of both groups to see the effectiveness of the text-mining tool.

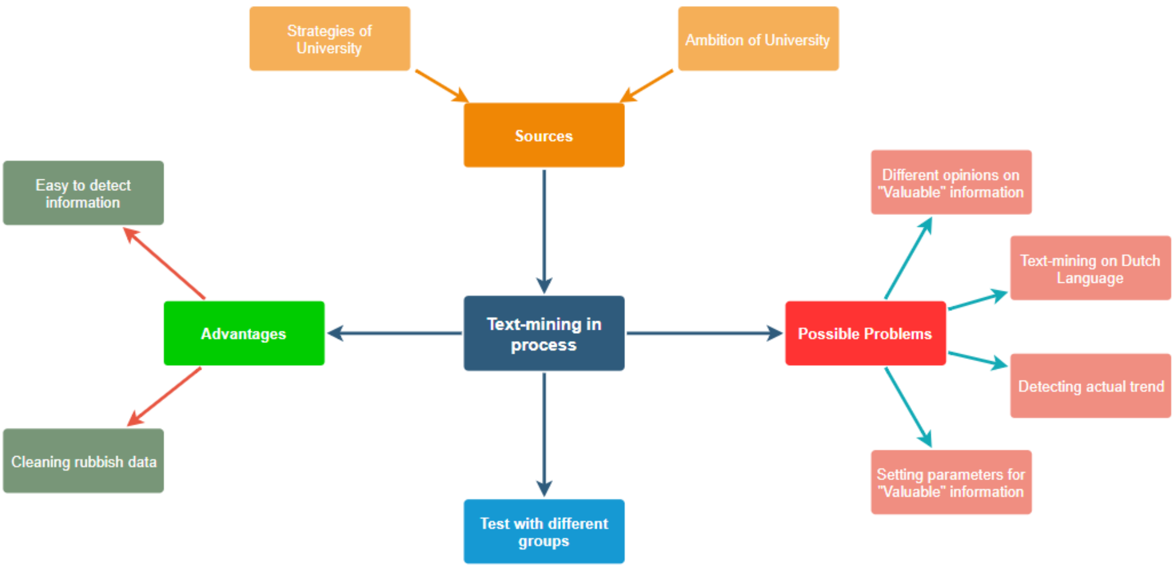
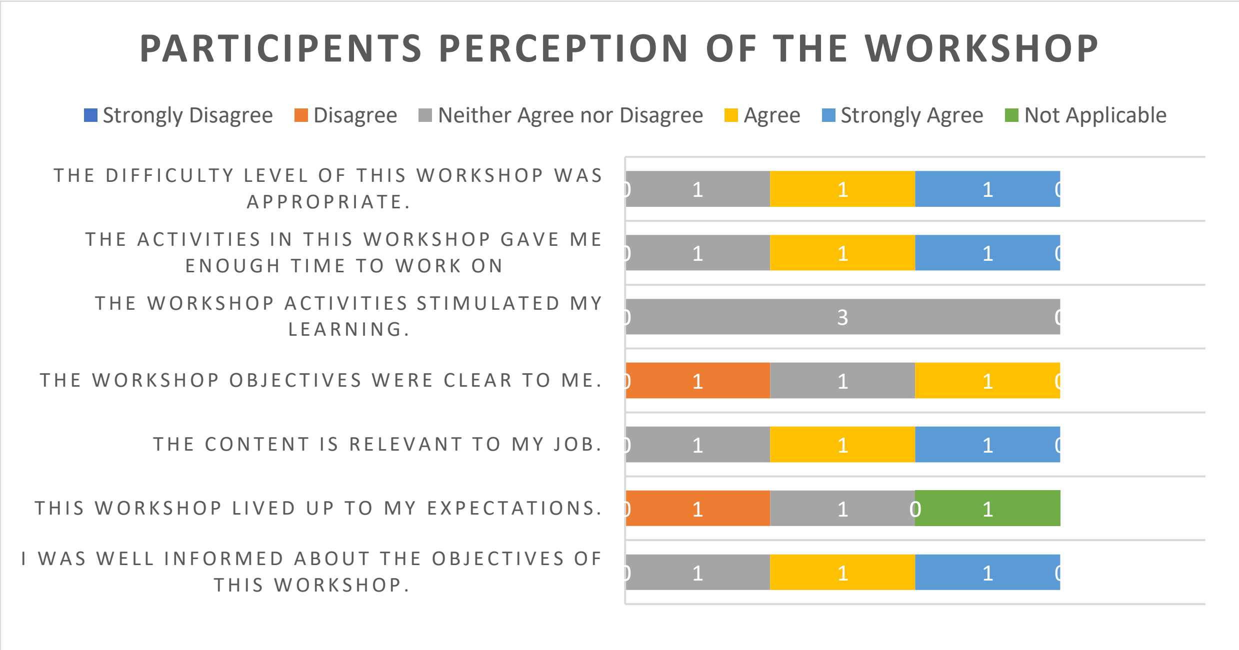


Figure 17: Q5 – How text-mining tool can help in this process?

4.3.3.3 1st workshop results (Online Learning Lab)

The first workshop was an expert discussion. Three experts of Online Learning Lab (OLL) participated in the workshop. Several questions were part of the workshop. By the end of the workshop, questionnaire was a data collection method. Questionnaire was also separated in three parts. First part consists of participants perception about the workshop; second was about the method (reports + Twitter + experts) and the last part consist of feedback and suggestions for the next workshop.

In the graph below, this was the perception of the workshop from the participants. The responses for difficulty of the workshop, its content, objectives and time requirement for this workshop to work on, we got answer in the positive (agree) direction. Whereas, workshop activities were neither in negative direction nor positive for stimulating the participants learning. The detailed result of the first section of questionnaire is shown in the *Graph 6*.



Graph 6: Participant's perception of the workshop (OLL)

Second part of the questionnaire was related to the method which was about the extraction of information through trend reports and Twitter and provide that information to experts for planning and making strategies related to their Online Learning Lab in coming future.

The first question was about the positive side of this method. The common responses, which we got, were that it is a good theoretical framework which includes discussion with other experts. All the participants liked the discussion part where they were discussing with other experts while making strategies and plans. One of the participants mentioned the content "SWOT Analysis", which were used during workshop, as useful.

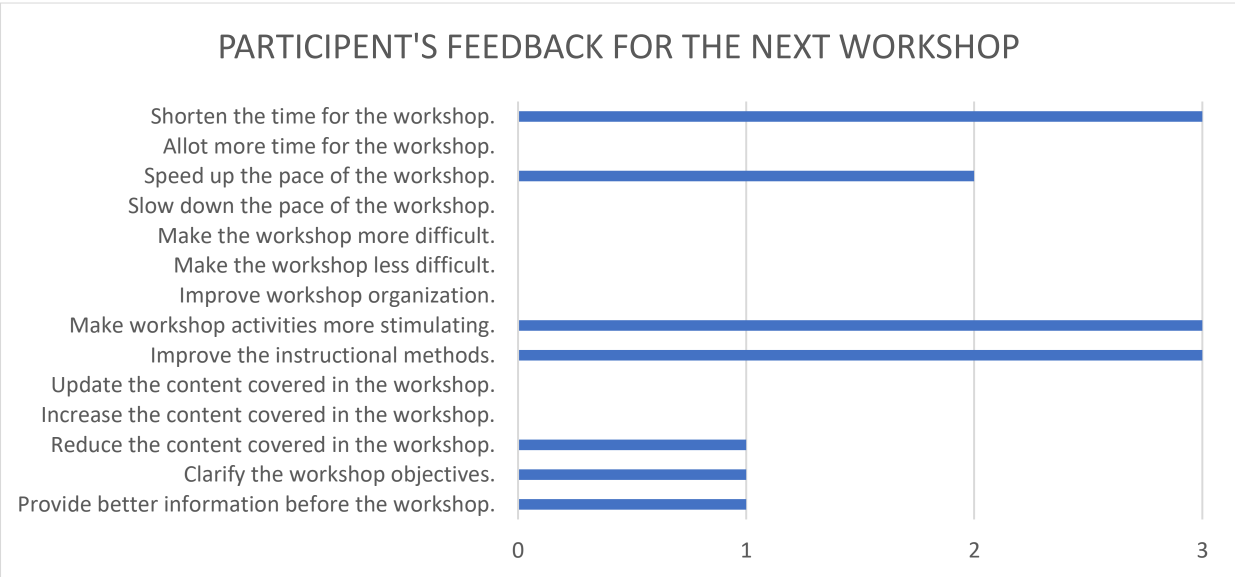
Second question was what improvements participants recommend in this method. The common answer by the participants was to activate or stimulate the discussion between the participant and one participant said that they already know the trends.

Third question was about the least valuable thing about this method. Out of the responses, common answer of two participants were that, the method was not 100% clear to them. They were unable to understand the connection between experts, reports and social media. And the other response was about the one part of content of the workshop, which was "PESTEL" analysis that was least valuable. As according to the participants, not all categories of PESTEL were applicable for Online Learning Lab.

Fourth question was what is the most valuable about this method. The common response to this was – "discussion". Other than discussion part, it was mentioned by the experts that they can see through this method that *"where is technology moving and which side people are moving as an early adopter."*

The last question was, what adjustments in the workshop are required to make this trendsetting workshop suitable for Online Learning Lab in future. The common response for this was to make the workshop more active and inspiring. Reduce the forecasting window to 1-2 years as several questions were asked in the forecasting window of 5 – 10 years. According to participants, due to the speed of technology developments, forecasting window of Online Learning Lab is not 5-10 years, but more about 1-2 years.

The last part of the questionnaire was comprised of participants feedback for the next workshop. Some common responses were to cut short the time for workshop (which was 2 hours), make the workshop more stimulating and improvement in the instructional methods. *Graph 7* shows the detail answer of the last part of the questionnaire. As left side describes the statements which were asked from them. Right side of the graph shows the votes for that specific statement.

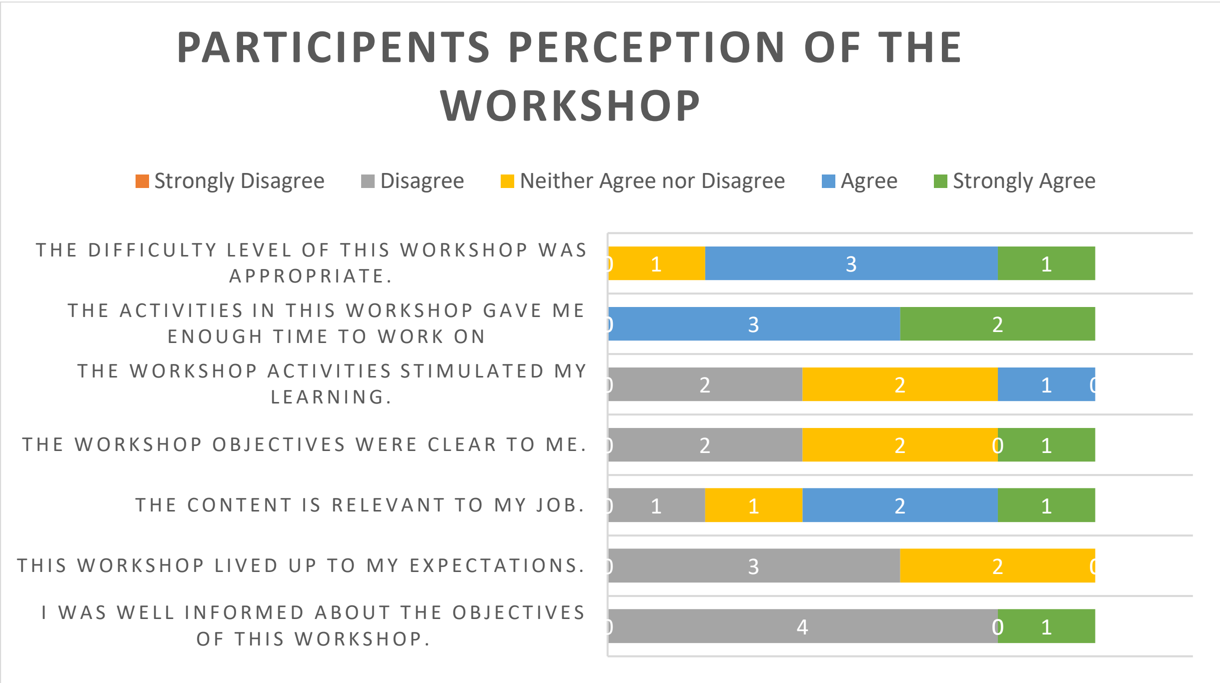


Graph 7: Participant's feedback for the next workshop (OLL)

4.3.3.4 2nd workshop results (Future Work Lab)

The second workshop was with the experts of Future Work Lab of Centre for Innovation. Five experts participated in the workshop. Questionnaires were used as a data collection method. For the Future Work Lab workshop, same questions were used in questionnaires, as it was for Online Learning Lab workshop, to get their perception on the workshop, method and feedback/suggestions for the future workshop.

In the graph below, this is the participants perception of the Future Work Lab workshop. On the left side. On the right side, numbers are considered as votes, and the colour shows their decision on the statement. *Graph 8* shows the result of participants perception of the workshop.



Graph 8: Participant's perception of the workshop (FWL)

The second part of the questionnaires was about method. This section included five questions. The first question was what participant thinks about this method (Reports + Twitter + Experts)? The responses for this question were little different to some experts. One participant described it as a "fact-based method", and the one described this as a "cool method" as this method is using different mediums of information (reports, Twitter) and can use multiple reports in one go. And other described the method as "good".

Second question was what improvements they would recommend in this method. For this question, several recommendations were made by the experts. Like, the first recommendation was to let experts fill the ideas about the topic first, before presenting extracted information to them. Other recommendation was to go "one level deep" in extracting information, to have more knowledge. Apart from this, clarity on concepts was asked to be made during presentation and asked for explaining the steps:

word → multi-words/ sentence + relevance.

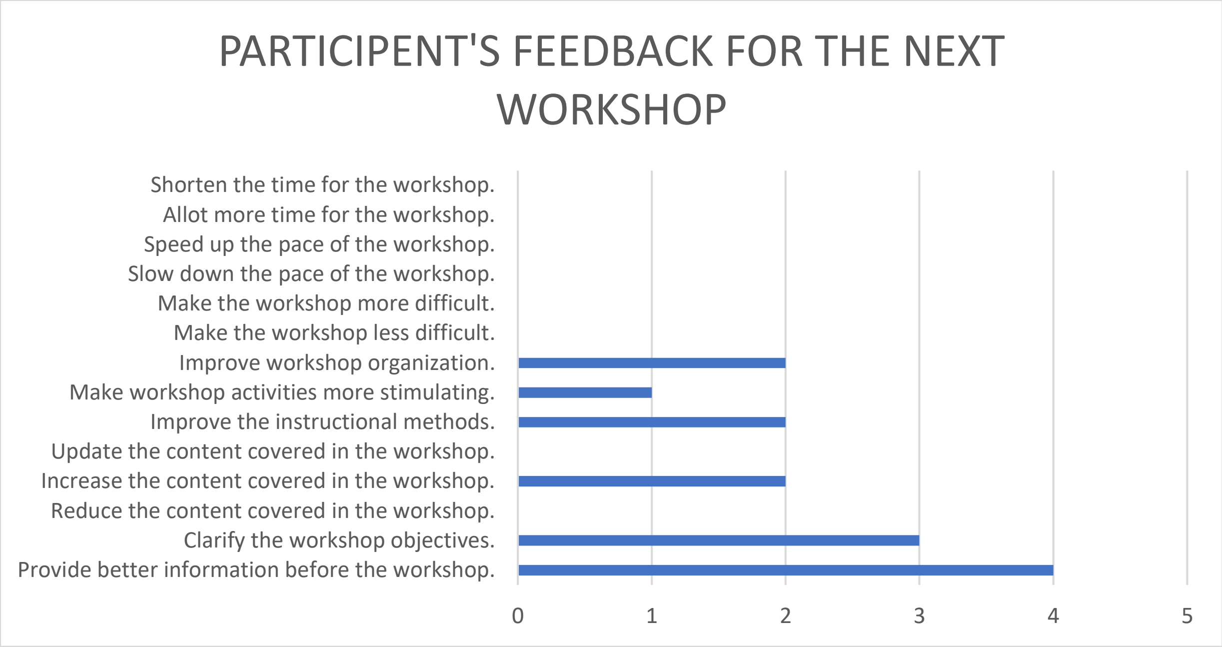
And the last recommendation which is somehow also an answer to the sub-problem of the fifth question of the interview *Figure 17 section – 4.3.3.2 – Q5*, "*how one can know that this information is valuable that is extracted from the reports and Twitter?*" The recommendations were to ask the experts about the trend/word analysis beforehand to have better information, so the information should be related to their required topics.

Third question was what is the least valuable about this method. The common response was contextualising by SWOT analysis for Future Work Lab. As SWOT analysis was in the content of workshop, but expert of 2nd workshop (Future Work Lab) didn't find SWOT much relating to the trends for Future Work Lab. One response was using social media information in the method was least valuable, and one response was "*that the exploration of these trends was less"* (means: little amount of sentence extraction where shown related to the multi-words and trends).

Fourth question was, what is the most valuable thing about this method? This question resulted into different responses including that it's a fact-based method, discussion with other experts was on the top and method sparked the curiosity in the participants when they saw visualization of data which was extracted through the text mining tool. Other than that, using the reports and experts at the same time was most valuable to one participant.

The last question was what need to be adjusted, to make this trendsetting workshop suitable for Future Work Lab? Several suggestions were made in order to adjust the second workshop to be more suitable for the Future Work Lab on a longer term. In these responses, one was, "*discussion should be more focused*". One suggested that, "*I would love this, to be part of the CFI (Centre for Innovation) team process and team meetings each quarter/ half year because when approached only on Future Work level, it is too broad*". One suggested that "*Backcasting exercise*" should be included in the content of this workshop and use the sentences (which are based on multi-words) better.

The last part of questionnaire was comprised on the overall feedback/suggestions for the next workshop in which majority of votes were cast on providing the better information related to the workshop beforehand. The result of last section is shown in the *Graph 9* below.



Graph 9: Participant's Feedback for the next workshop (FWL)

4.4 Evaluating

This section describes the explanation and interpretation of the results and linking them to the existing literature. The results which are derived from the text-mining tool, interview, meetings, and the workshops are also interpreted and linked with the existing literature.

As described in (Kayser & Blind, 2017) that data sources which are available for forecasting, these sources are hardly combined with other sources. As the paper specifically mentioned "web-contents" which is often ignored by researchers as a data source. So, for that, Twitter data was also used as a data source other than trend reports.

Text-mining was applied almost in the same structure as described in (Fan et al., 2006), (Yu et al., 2005) and (Kayser & Blind, 2017). Information was extracted by association with other terms (co-occurrence) to analyse the data. While doing text preprocessing, stemming was used in the start but later, it was skipped as there were number of entries/terms which were not making sense (because stemming function of KNIME tool was cutting some important words). To avoid that, stemming was omitted later in this process. Information was extracted on the basis of TF (Term Frequency), TF-IDF (Term Frequency-Inverse Document Frequency), keywords, multi-words and sentences.

As described in *section – Literature Gap –* that there are several methods and techniques to do forecasting but rarely combined (Kayser & Blind, 2017). And to get better result for forecasting, it is better to use multiple mediums (Database → e.g. reports, Social Media → e.g. Twitter and People → e.g. experts) as described in (Porter, 2009). To get better results for forecasting, all these mediums were combined in this research to see how good and helpful this method is for experts to make strategies and plan.

The most interesting findings after implementing this method are explained below:

4.4.1 Considering multiple sources for forecasting is useful

The question raised in literature review that to what extent can text mining from different data sources (reports and Twitter) support experts in trend forecasting? So, participants (experts) found this method useful as described in the *section – 4.4.2.3 and 4.4.2.4 –*. As (Kayser & Blind, 2017) describe in their research,

> *"Data sources like social media or user-generated content are rarely considered in foresight for capturing societal points of view."*

They believe that the user-generated content, which is mostly present on social media platforms especially Twitter, is hardly used in forecasting. Keeping this thing in mind, we have used social media content in this research. We had focused on one platform, which was Twitter.

(Porter, 2009) has distinguished the information resources into six main types (as shown in *Table 11*).

Medium \ Message	**1) Technology**	**2) Context**
A) Databases	Research funding, publication & patent abstracts, citations	Business, market, policy, popular opinion
B) Internet	Technical content sites	Company sites, blogs, etc.
C) People	Technical experts	Business experts

Table 11: Six Information Type - (Porter, 2009)

As (Porter, 2009) describes,

> *"…Many technology managers fail to take advantage of rows A and B; they rely almost exclusively on tacit judgment. This is folly. Those who obtain empirically based knowledge, in addition to using human expertise, will dominate those who don't do so. Others equate electronic information resources with the internet (row B). That also misses a vital resource – row A."*

This shows that all these mediums (Databases, Internet and People) are important to use together. (Kayser & Blind, 2017) also describe to use other mediums apart from social media (Twitter), like scientific publications, newspapers, patents etc. Considering multiple sources of information is useful and as described by Porter, the one who use all these mediums dominate the ones who don't.

So, as literature suggested to use different mediums for text-mining to extract the information, we have used different mediums of information in our research to see to what extent these mediums can help experts in forecasting. Experts found this method (using multiple sources into consideration and text-mining tool to extract information out of these sources) beneficial and useful. As text-mining tool is taking multiple reports in one go and extracts the meaningful information out of all the "rubbish data" (as described by interviewee). The common thing which experts praised is the "discussion section" in this process. As in the discussion section, first, every expert described and wrote their answer. They were free to write and describe what they think about that question. Mostly, in the sessions like these (workshops or seminars) most people get influenced by other answers. They just agree to other people answers without coming up with their own opinion. So, the Delphi method has solved that problem which we have used as an expert discussion in this method. As for Delphi technique, participant identity is kept secret from others. And no one knows who has answered what and they have to come up with their own answer or opinion first.

In the workshops, experts answered the questions first. Then those answers were shown to all experts. By keeping participants' identity a secret, only answers were shown to all experts without mentioning their name like which expert has answered what. Then the researcher/coordinator who was handling the workshop or session extracted the common answers from all the answers and separated the unique answers. In other words, the coordinator/researcher did the census (which is the core part of Delphi method) and then came up with answers. Some of the answers were common in between all answers, and some were uncommon. Every answer was discussed among the experts – like what you think about this answer? How much this answer relates to the question? etc. When most of experts answered the same and had agreed on one answer as a majority, then we had considered that answer. And also, we questioned in the workshop on uncommon answers like why this answer is unique? Does it relate to our question? Or might other experts have missed something related to that question, and only one expert has caught that idea/point and mentioned that thing in answer. So, we discussed on unique answers to see another experts' perspective. Experts from both labs favoured this section of process and supported the whole process. They described this method (after updating with the feedbacks) a perfect way for the trend selection process.

4.4.2 Improvements for future workshop

Although the method still needs improvements like in organizing the workshops in a better way and putting more and related content in the workshop. To make the workshops more stimulating, we can improve by including content according to the feedback which was given by the experts of both the labs (as explained in *section – 4.4.2.3 and 4.4.2.4 –*). In this research, we have improved the second workshop through the feedback from the first workshop. The feedback which was taken from the first workshop, that feedback was taken into account for the second workshop, and that made the second workshop better than the 1st one. Similarly, feedback was asked from the experts in the second workshop for the next (future) workshop, and those things will be considered for the next workshop to improve it and so on.

Like this, this process will keep on improving and will be more helpful for forecasting. As, concerning to content, SWOT analysis was appreciated by the experts of first workshop but experts of second workshop didn't find SWOT analysis related to their field. So, such kind of content needs to be improved for future and only use that content which is related to the workshop. There is a need to improve the content (which is used during workshop) before every workshop because it is not necessary that if the content or methods (like SWOT, PESTEL analysis etc.) are working in one workshop, can work on others too. So, it is required to ask the expert(s) beforehand, that which content is related to their field so that you can only use that content.

The common feedback which we got from both workshops as a most valuable and for improvement, is the "discussion". Discussion between researcher and participants needs to be improved for future workshop(s). Workshops should have more active content, and the coordinator should activate the participants during the workshop. From a second workshop, we got a feedback, which will be taken into consideration for the future workshops, shown in *figure 18*.

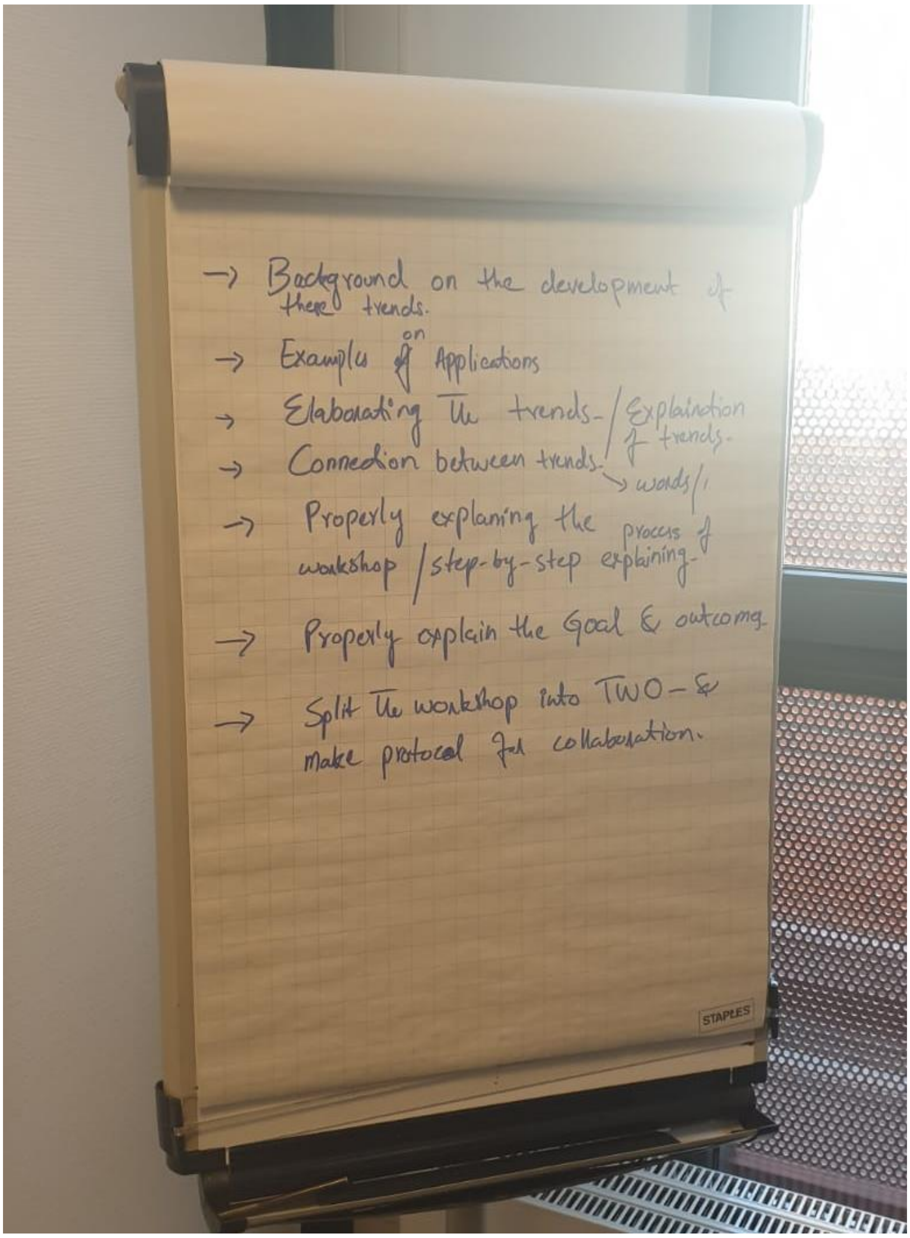


Figure 18: Things need to improve/add for next workshop – (Experts feedback during 2nd workshop)

This feedback (mentioned in *Figure 18*) can help us to make the future workshops more effective. The mentioned things can be improved by giving experts more content and information for which they have asked. Other than that, we can improve the workshop when the coordinator/ researcher who is extracting the data through text-mining has the same background and knows about the subject which will be discussed during the workshop. In this way, co-ordinator can make the workshop more stimulating and active because he/she then exactly knows what the requirements of the experts are and which information is more valuable to them.

When using experts' opinion in the method, for example in Delphi method, there are several things which need to keep in mind. (Rowe & Wright, 2001) describe that there are several principles to keep in mind while going for Delphi method. In order to get better result, it is required to use experts of different domain (around 5 – 20). And to get the census; it (Delphi technique) requires to have multiple rounds (which of course requires a lot of time) (Rowe & Wright, 2001). These things should be kept in mind while using expert methods in the forecasting that "availability of experts of different domain" and "time" is required for them, either its Delphi method or Scenario planning method. Like for scenarios, you need group of experts working on each scenario, which takes time and depends on how many scenarios are you considering or planning.

Apart from experts and workshops, there are some other things which need to improve in text-mining to extract more clean information out of it. During the pre-processing, stemming was first used to reduce the word to its basic form. But this process was reducing several words into that form which was not making any sense, and it was difficult to understand the actual words. This process was later discarded from pre-processing. But it is better to use "lemmatization" as it replaces the word with the synonyms. As described by (Kayser & Blind, 2017), techniques like,

> *"…lemmatization (which reduces word to root form based on dictionary) are applied in pre-processing."*

So, it is better to use lemmatization, as it provides better results (after having detailed dictionaries)[5].

4.4.3 Trends are not issue, but allocation of resources

One of the main finding in this research was that it is not a problem to see which trends will be there in the future for Centre for Innovation to work on. And this was said by a participant in Question 2 of Online Learning Lab's workshop, but the main problem is allocation of the resources as they (CFI/LU) have limited amount of resources: "Budget and Time", as answered by the interviewee (described in the *figure 13 of section 4.3.3.2*). There is a limited amount of budget for realization of the trends, and also CFI/LU cannot handle multiple trends into consideration due to limited time. To overcome this problem, a better distribution is required across observing, experimenting and implementing these trends. The reason for having a better distribution is that, when you have little amount of resources, you make conscious choices like where to spend and where to invest more efficiently and effectively. This way you can extract maximum out of the little resources.

(Langen & Kammergruber, 2013) have used radar visualisation to analyse non-technical trends or to have better view on critical issues. They call it "Future Radar". In their research, they have used "Radar Visualization" to analyse "PESTEL analysis" which they have used to identify trends or issues in the external environment to look over the substantial changes for the business. We have taken the "Radar Visualization" concept from (Langen & Kammergruber, 2013) and altered in our research in order to analyse the trends. In our "Radar Visualization", we have included three phases which includes observing, implementing and

[5] https://blog.bitext.com/what-is-the-difference-between-stemming-and-lemmatization/

experimenting phase. In observing phase, we have included those trends which needs to be observe. In experimenting phase, we have included those trends on which we should experiment. And in the phase, implementing, we have included those trends which we should implement.

So, if one has better "Radar Plan", they will know exactly what they have got, what they are going to do and how much can they achieve.

An example of such a radar is shown in a *Figure 19*:

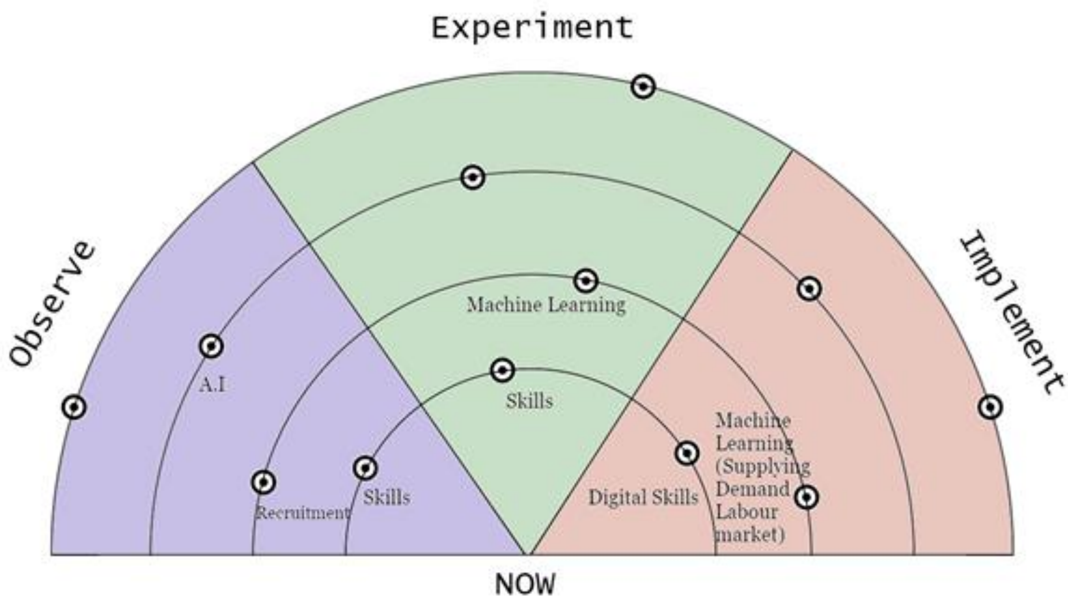


Figure 19: Radar Plan for Future Work Lab

When you have better allocation of your resources, and you know exactly what your demand is and what you are going to do, then you can use your resources very efficiently. This radar in *Figure 19* shows the trends, which were resulted during the workshop with Future Work Lab.

So, having such a radar plan and knowing the trends and the distribution, you can divide the trends into three categories.

Here is a brief explanation of these three categories:

For Observing:	Yes! We will look these trends into next year.
For Experimenting:	We have limited resources, but maybe we can spend month or two on this.
For Implementing:	This is time for this trend now! We can set a project and so on.

4.4.4 It is not about the "tool" only, but discussion among people

Other main finding in this research is that it is not about the tool only, which is extracting information out of trend reports and social media, but it's the discussion which plays the main part. No doubt that text mining tool can extract meaningful information out of these mediums (Educational, Work and Technological reports and Social Media) and play a major role in whole process but until the people who are connected to this process are not discussing on the given information or exchanging information between them, this process can't work perfectly.

As described by (Porter, 2009), it is very much important to add experts opinion into forecasting along with other mediums like database (reports) and social media. Even though, these mediums are important but if we combine expert's opinion with these mediums, it will make the forecasting much more accurate. Using experts' opinion is also backed by (Rowe & Wright, 2001) as they mentioned,

> *"Expert opinion is often necessary in forecasting tasks because of a lack of appropriate or available information for using statistical procedures."*

During the research, it was mentioned by the interviewee that people follow news and social media in their daily life, they mostly know what the emerging trends are, and they don't require a tool to know emerging trends. According to interview, *"… of course, your process makes it easier, but people follow news, read articles. Right now, are they able with limited information to select the good and best trends already? Or do they need already this tool? My hypothesis is that they already know the trends …"*. On the other hand, interviewee mentioned "discussion" as an important part in the whole trend selection process. As during the question related to how to improve the trend selection process, interviewee mentioned, *"make it more structural and independent of the persons. So, do meeting every twice a year to discuss this trend."* This shows that, in order to make the trend selection better for future, it is required to have meetings and discussions among people, who are related to trend selection process.

Other than interviewee, experts also mentioned discussion as a key factor in this whole process during the workshops. As these discussions are not occurring more often (one such kind of discussion/meetings in a 2 and half years – as mentioned by interviewee), there is a need of such kind of workshops and meetings for the people to have at least one such workshop/meeting in a half or a year or more often.

4.4.5 Criteria for "valuable" information is a question

As this question was raised during the interview by the interviewee that what is the basic algorithm behind information being valuable or important. As asked by the interviewee, *"How you can determine what is valuable and what is rubbish and is valuable for me is same as for you? So, is there a one algorithm for me, for people at my position or for every person on this earth?"*

As the tool is extracting information, that information for some people might be valuable but for others maybe not. This is tough because it may involve debate among the experts or the people involving around the information. Information may be contrary to some's interest, but for others, it may be not. So, criteria for being an information being "valuable" is required, and it should be there so that everyone agrees to that.

One expert described the solution for this and other experts agreed to that solution. It was, to have an expert's opinion before extracting the information out of multi-words and sentences. Expert mentioned that, "Ask the experts beforehand that what information do you require from the most often term's and of what type? So that when you extract the information, it should be the one which is interested to the experts."

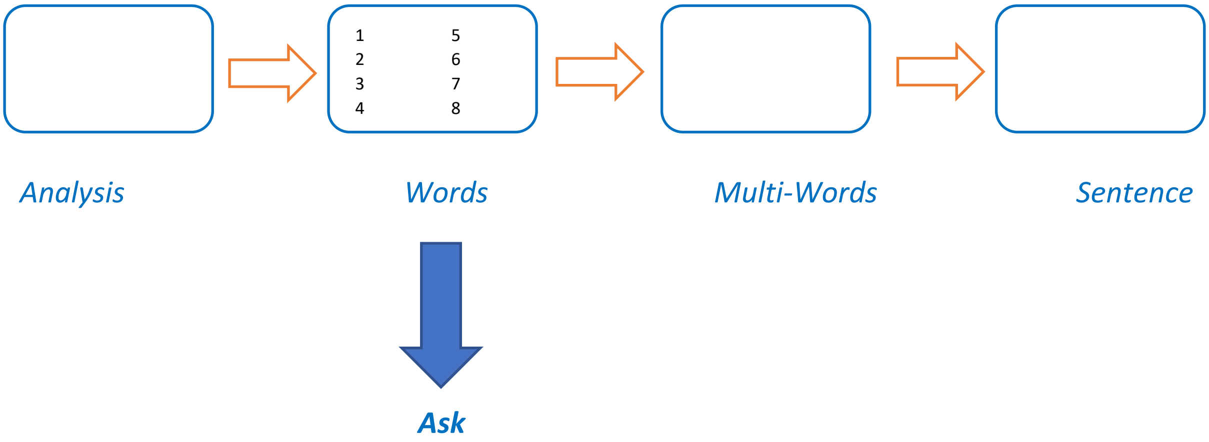


Figure 20: Participants feedback on making the process better to extract better information

This figure shows that at what point we should ask for experts' feedback so that the information which will be extracted from this process, will only base on what experts have asked for. For example, out of 10 trends (words), ask the experts first that which trends are related to you or for which trend(s) you require more information? After asking them and having their feedback on the term/trends, in the end, experts will get only that information which they have asked for. So, they will only get the information which they have asked for, and this can save their time and also save the time for the workshop. Then there will be no need to discuss that information in which experts are not interested or which is not important to them.

Future Workshop Layout:

Keeping the feedbacks from both workshops in mind that how the possible future workshop should be, these points should be kept in mind for future workshop:

1. Workshop shouldn’t exceed 2 hours as it creates tiredness among participants and there is a chance that they might lose interest.
2. Possibly have a 5-10 minutes break during the workshop.
3. Provide the better information regarding the content before the workshop, so they come prepared.
4. Present the content in stimulating way to gain participant’s interest.
5. Discuss the content with the one of the experts (or their leader) to see the content is related to that specific field.
6. Discuss on the background on the development of the trends.
7. Examples on applications of trends should be more provided.
8. Explain the trends better.
9. Step-by-step explaining the process of the workshop.
10. Explain the goal and outcomes properly.
11. Split the workshop into two half and make protocol for collaboration.
12. KNIME is a good tool for text-mining, but you can use some other better tools which can give you more clean and accurate results due to their better and extra built-in functions for text-mining (in case if you want to go for some GUI analytical platform).

5. Conclusion

In this research, forecasting practice was tested, and the previous practice, which was used at the Centre for Innovation (CFI), was improved. Several techniques and mediums were combined in the light of different recommendations by (Porter, 2009) and (Kayser & Blind, 2017). Outcomes of this method were discussed with experts, and the results were gotten by combining different mediums of information and different techniques for forecasting.

The struggle of 9 months and facing several challenges, i.e. usage of right and appropriate functions for text-mining tools to extract clean data, making sure experts were available to participate in the workshop due to their busy schedule and thereafter, arranging another workshop to see the changes from the prior workshop which was a method was highly appreciated by the experts.

To see how far text-mining can help experts for forecasting, different mediums of information were used to extract the information which was helpful for experts to forecast and have better understanding on the distribution of resources. Data was collected from different sorts of reports, Twitter and people who are connected to CFI. Experts of different labs of CFI participated in the workshops to test the methodology and to see how far this methodology can help them in forecasting and to have better understanding on how to distribute the resources. Experts appreciated this method, especially the “expert discussion” in the workshops. Several recommendation were also made to make the future workshop more stimulating and better. They would like to see this method (after applying their recommendations) for Centre for Innovation’s team process and team meetings in half a year or so.

So, this research concludes that it is not about the problem of identifying the trends, but allocation of resources to explore the trends that matter the most. And secondly, it is not about the tool only, which is extracting the information, but the “discussion among experts” play a key role.

5.1 Limitations and Recommendations

There are some limitations to this research due to several time and resource constraints. Results might be slightly different if we would have more experts in the workshops. Due to busy and hectic schedule of experts, limited number of experts were able to participate in the workshops. Similarly, experts’ methods, like Delphi technique requires several sessions to have census. It was impossible to do other sessions in the available time of this research. Otherwise, that would have exceeded the time for this research. So, there was one session of discussion with each workshop. During the research, it was planned to do multiple experts’-based methods like scenario planning along with Delphi method but again, it was unable to do so due to unavailability of experts. Because getting results from different methods, like from Delphi technique and scenario planning, results would have different (by comparing the results of Delphi and Scenario Planning). It was also planned to only have one workshop with the same lab and then dividing the experts into two groups. We would have provided the extracted information to only one group, and other group would have used their own knowledge (as asked by the interviewee). That would have provided the clear results like how far our method is helpful, by comparing the results of both groups. In this case, result would have been different. But again, due to unavailability of experts, it was unable to do so because of the busy schedule of experts. So, that is why, another workshop was organized with different lab. Other restriction was of the availability of Twitter data. Twitter data was used through Twitter API, but that data was of limited amount – from last 7 days (as mentioned by Twitter on their website), not like 10-15 years old data. For accessing large amount of old Twitter data, it was costing a lot of money (by buying premium Twitter API). If the data from past several years was available for this research, result would have been slightly different from the present result for Twitter search. There was also a little hurdle to clean the data (while pre-processing) because some filters of KNIME are partly broken, and some are unavailable (as told by KNIME Team Members), so some glitches were present in the data (as shown in "word cloud"). But the data presented to the experts was clean and without these glitches like "HTTP" and ".com" etc. as these small glitches were removed/ignored manually while presenting the data to the experts.

This method also needs several changes to make it better for future. Techniques and mediums are good which are using reports, Twitter data and adding expert's knowledge in it. Required changes are in the "process" of this method and the outcome information:

(1) Before extracting the information, get expert's feedback on exactly what they require so that you don't have to spend your time to deal with the information that is not relevant to them because they are the one who will make strategies and plans.

(2) Ask the experts, how much information do they require when you initially provide them "trends". Before going for "multi-words" and "sentence extraction", ask from experts what they require and how much samples or examples do they need.

(3) It is recommended to have the expert with you when you are extracting the information. Because knowledge about related field is required when you extract the information. A person with no knowledge about the field for which he/she is extracting information from the mediums, likely to miss some important information which he/she may think as not important or he/she could give you only "rubbish" information which he/she may think as important. A person who has knowledge related to the field, for which they are extracting the information, will help to identify the useful information.

(4) Text-mining machines usually take a lot of processing power. So, it is better to use the machine with large amount of RAM, storage and processor. When sentences and multi-words are extracted, the machine store those words and sentences into some temporary tables – like excel, but those words can cross millions of lines (e.g. during phrases extraction), which exceeds the limit of those tables and resulted into memory

shortage. So, it is required to either split the inputs, like process 1st half of documents or any input, and then the other one, or divide your input into different categories and perform text-mining separately or use a powerful machine for text mining.

5.2 Future Work

In future, the method and workshops can be improved from the above recommendations for the process and for arranging future workshop. Using "Backcasting exercise" will be helpful and feedback from 2nd workshop is highly recommended. It would be amazing to see the outcome of the results when these sources; trend reports and Twitter are combined with other sources like news web pages (to see what is happening right now or any future news about trend or anything this method is applied for) and on patents. As there are several cases where text-mining has been applied on patents separately, but it would be of great importance if it is combined with the other sources to see what actual innovations are in process, which are likely to be there in the future.

6. References

Albert, T., Moehrle, M. G., & Meyer, S. (2015). Technology maturity assessment based on blog analysis. *Technological Forecasting and Social Change*, *92*, 196–209. https://doi.org/10.1016/j.techfore.2014.08.011

Ayers, R. (1969). Technological Forecasting and Long-Range Planning.

Balachandra, R. (1980). Technological Forecasting: Who does it and how useful is it? *Technological Forecasting and Social Change*. https://doi.org/10.1016/0040-1625(80)90075-X

Banu, G. R., & Chitra, V. (2015). A Survey of Text Mining Concepts. *Internacional Jornal and Technology (IJIET)*, *2*(2319–1058).

Berry, M. W. (2004). Survey of Text Mining : Clustering, Classification, and Retrieval. *New York*, 262. https://doi.org/10.1007/978-1-84800-046-9

Cachia, R., Compañó, R., & Da Costa, O. (2007). Grasping the potential of online social networks for foresight. *Technological Forecasting and Social Change*, *74*(8), 1179–1203. https://doi.org/10.1016/j.techfore.2007.05.006

Coates, Vary, F. (2001). On the future of technological forecasting. *International Journal of Foresight and Innovation Policy*, *1*(3/4), 270. https://doi.org/10.1504/IJFIP.2004.004964

Comin, D., & Hobijn, B. (2004). Cross-country technology adoption: Making the theories face the facts. *Journal of Monetary Economics*, *51*(1), 39–83. https://doi.org/10.1016/j.jmoneco.2003.07.003

Crawford, K., & Boyd, D. (2012). CRITICAL QUESTIONS FOR BIG DATA, (2012), 662–679. https://doi.org/10.1080/1369118X.2012.678878

Cunningham, S. W., Porter, A. L., & Newman, N. C. (2006). Special issue on tech mining. *Technological Forecasting and Social Change*, *73*(8), 915–922. https://doi.org/10.1016/j.techfore.2006.06.004

de Miranda Santo, M., Coelho, G. M., dos Santos, D. M., & Filho, L. F. (2006). Text mining as a valuable tool in foresight exercises: A study on nanotechnology BT - Tech Mining: Exploiting Science and Technology Information Resources. *Technological Forecasting and Social Change*, *73*(8), 1013–1027. https://doi.org/10.1016/j.techfore.2006.05.020

Delen, D., & Crossland, M. D. (2008). Seeding the survey and analysis of research literature with text mining. *Expert Systems with Applications*, *34*(3), 1707–1720. https://doi.org/10.1016/j.eswa.2007.01.035

Fan, W., Wallace, L., Rich, S., & Zhang, Z. (2006). Tapping the power of text mining. *Communications of the ACM*, *49*(9), 76–82. https://doi.org/10.1145/1151030.1151032

Feldman, R., & Dagan, I. (1995). Knowledge Discovery in Textual Databases (KDT). *International Conference on Knowledge Discovery and Data Mining (KDD)*, 112–117. https://doi.org/10.1.1.47.7462

Gardner, E. S., Jr, & McKenzie, E. (1985). Forecasting Trends in Time Series, *31*(10).

Glassey, O. (2012). Folksonomies: Spontaneous crowd sourcing with online early detection potential? *Futures*, *44*(3), 257–264. https://doi.org/10.1016/j.futures.2011.10.008

Glenisson, P., Glänzel, W., Janssens, F., & De Moor, B. (2005). Combining full text and bibliometric information in mapping scientific disciplines. *Information Processing and Management*, *41*(6), 1548–1572. https://doi.org/10.1016/j.ipm.2005.03.021

Goodwin, P., Ord, J. K., Öller, L.-E., Sniezek, J. A., & Leonard, M. (2002). *Principles of Forecasting: A Handbook for Researchers and Practitioners*. (J. S. Armstrong, Ed.), *International Journal of Forecasting* (Vol. 18). https://doi.org/10.1016/S0169-2070(02)00034-1

Greenwood, D. J., & Levin, M. (2007). *INTRODUCTION to ACTION RESEARCH* (2nd ed.). SAGE Publications.

Gupta, V., & Lehal, G. S. (2009). A Survey of Text Mining Techniques and Applications. *Journal of Emerging Technologies in Web Intelligence*, *1*(1), 17. https://doi.org/10.4304/jetwi.1.1.60-76

Hallowell, M. R., & Gambatese, J. A. (2010). Qualitative Research: Application of the Delphi Method to CEM Research. *Journal of Construction Engineering and Management*, *136*(1), 99–107. https://doi.org/10.1061/(ASCE)CO.1943-7862.0000137

Hearst, M. A. (1997). Text data mining: Issues, techniques, and the relationship to information access.

Hyndman, R. J., & Athanasopoulos, G. (2018). *Forecasting: Principles and Practice* (2nd ed.).

Jun, S. (2014). A Technology Forecasting Method using Text Mining and Visual Apriori Algorithm, *40*(1), 35–40.

Kaplan, A. M., & Haenlein, M. (2010). Users of the world, unite! The challenges and opportunities of Social Media. *Business Horizons*, *53*(1), 59–68. https://doi.org/10.1016/j.bushor.2009.09.003

Kayser, V., & Blind, K. (2017). Extending the knowledge base of foresight: The contribution of text mining. *Technological Forecasting and Social Change*, *116*, 208–215. https://doi.org/10.1016/j.techfore.2016.10.017

Kietzmann, J. H., Hermkens, K., McCarthy, I. P., & Silvestre, B. S. (2011). Social media? Get serious! Understanding the functional building blocks of social media. *Business Horizons*, *54*(3), 241–251. https://doi.org/10.1016/j.bushor.2011.01.005

Kitchin, R. (2014). Big Data , new epistemologies and paradigm shifts, (June), 1–12. https://doi.org/10.1177/2053951714528481

Langen, M., & Kammergruber, W. C. (2013). Collaborative Trend Scouting and Roadmapping.

Martino, J. P. (1980). Technology forecasting—An overview, *26*, 28–33.

Mcniff, J., & Whitehead, J. (2002). *Action Research : Principles and Practice* (Second).

Mcniff, J., & Whitehead, J. (2010). *You and Your Action Research Project* (Third).

Meredith, J. R., Mantel, S. J., & Jr. (1995). Technological Forecasting.

Meystre, S. M., Savova, G. K., Kipper-Schuler, K. C., & Hurdle, J. F. (2008). Extracting Information from Textual Documents in the Electronic Health Record: A Review of Recent Research. *IMIA Yearbook of Medical Informatics Methods Inf Med*, *47*(1), 128–44. https://doi.org/me08010128

Mitchell, V. W. (1992). Using Delphi to Forecast in New Technology Industries. *Marketing Intelligence & Planning*, *10*(2), 4–9. https://doi.org/10.1108/02634509210012069

Mittermayer, M. -a. (2004). Forecasting Intraday stock price trends with text mining techniques. *37th Annual Hawaii International Conference on System Sciences, 2004. Proceedings of The*, *00*(C), 1–10. https://doi.org/10.1109/HICSS.2004.1265201

Modrak, V., & Bosun, P. (2014). Using the Delphi method in forecasting tourism activity. *International Letters of Social and Humanistic …*, *25*, 66–72. https://doi.org/10.18052/www.scipress.com/ILSHS.25.66

Navathe, Shamkant, B., & Ramez, E. (2000). Data Warehousing And Data Mining. In *Fundamentals of Database Systems* (pp. 841–872). Pearson Education pvt Inc, Singapore.

Ortner, H., Pfurtscheller, D., Rizzolli, M., & Wiesinger, A. (2014). *Datenflut und Informationskanäle*. Retrieved from http://deposit.d-nb.de/cgi-bin/dokserv?id=4774084&prov=M&dok_var=1&dok_ext=htm%5Cnhttp://media.obvsg.at/AC11979287-1001%5Cnhttps://resolver.obvsg.at/urn:nbn:at:at-ubi:3-385%5Cnhttp://media.obvsg.at/AC11979287-2001%5Cnhttp://ubdata.univie.ac.at/AC11979287

Pang, A. S. K. (2010). Social scanning: Improving futures through Web 2.0; or, finally a use for twitter. *Futures*, *42*(10), 1222–1230. https://doi.org/10.1016/j.futures.2010.09.003

Popper, R. (2008). *How are foresight methods selected? Foresight* (Vol. 10). https://doi.org/10.1108/14636680810918586

Porter, A. L. (2009). Tech mining for future-oriented technology analyses. *Text Mining*, 1–21. Retrieved from http://www.thevantagepoint.com/resources/articles/FRM chapter-TextMining-2009mar3.pdf

Porter, A. L. (2010). Technology foresight: types and methods. *International Journal of Foresight and Innovation Policy*, *6*(1), 36–45. https://doi.org/10.1504/IJFIP.2010.032664

Roper, A. T., Cunningham, S. W., Porter, A. L., Mason, T. W., Rossini, F. A., & Banks, J. (2011). *Forecasting and Management of Technology*. John Wiley & Sons, Inc.

Rowe, G., & Wright, G. (2001). Expert Opinions in Forecasting: The Role of the Delphi Technique, 125–144. https://doi.org/10.1007/978-0-306-47630-3_7

Sanders, N. R., & Reid, R. D. (2012). *Operations Management: An Integrated Approach* (5th ed.). John Wiley & Sons, Inc.

Saunders, M., Lewis, P., & Thornhill, A. (2007). *Research Methods for Business Students* (Fourth).

Stettina, C. J., & Smit, M. N. W. (2016). Team Portfolio Scrum : An Action Research on Multitasking in Multi-project Scrum Teams, *2*, 79–91. https://doi.org/10.1007/978-3-319-33515-5

Tan, A.-H. (1999). Text Mining: The state of the art and the challenges. *Proceedings of the PAKDD 1999 Workshop on Knowledge Disocovery from Advanced Databases*, *8*, 65–70. https://doi.org/10.1.1.38.7672

Thomas, G. (2017). Action Research. In *How to Do Your Research Project* (3rd ed., pp. 154–158).

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

van Eck, N. J., & Waltman, L. (2011). Text mining and visualization using VOSviewer. *ISSI Newsletter*, *7*(3), 50–54. https://doi.org/10.1371/journal.pone.0054847

Wang, B., Huang, H., & Wang, X. (2012). A novel text mining approach to financial time series forecasting. *Neurocomputing*, *83*, 136–145. https://doi.org/10.1016/j.neucom.2011.12.013

Yau, C. K., Porter, A., Newman, N., & Suominen, A. (2014). Clustering scientific documents with topic modeling. *Scientometrics*, *100*(3), 767–786. https://doi.org/10.1007/s11192-014-1321-8

Yoon, B., & Park, Y. (2007). Development of new technology forecasting algorithm: Hybrid approach for morphology analysis and conjoint analysis of patent information. *IEEE Transactions on Engineering Management*, *54*(3), 588–599. https://doi.org/10.1109/TEM.2007.900796

Yu, L., Wang, S., & Lai, K. (2005). A rough-set-refined text mining approach for crude oil market tendency forecasting. *International Journal of Knowledge and Systems Sciences*, *2*(1), 33–46. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.115.9384&rep=rep1&type=pdf

Yun, Y. C., Jeong, G. H., & Kim, S. H. (1991). A Delphi technology forecasting approach using a semi-Markov concept, *40*, 273.

Zhai, C., & Massung, S. (2016). *Text Data Management and Analysis: A Practical Introduction to Information Retrieval and Text Mining*.

Zhu, D., & Porter, A. L. (2002). Automated extraction and visualization of information for technological intelligence and forecasting. *Technological Forecasting and Social Change*, *69*(5), 495–506. https://doi.org/10.1016/S0040-1625(01)00157-3

7. Appendix A – Interview Transcript

Interview Transcription with CIO of Leiden University

Rashid:
Can you tell me how this trend selection process is currently done in university?

Jan Willem:
Bit ad-hoc, we made a new strategy plan and it was a process of about two or 1 and half two years, so very long to make new strategy. This is because of research. We have three domains. 1) business administration 2) education and 3) research. And for the research, it was very difficult to who to ask what to do. So, who determines this organization what is our strategy on research. So, we needed to find it out to figure this process first before filling it in, testing it with research at themselves. We started with trends and workshops with the colleague. Ok, what do we see, and what is happening in our university and outside. We have Gartner contract, they make trends analysis. I visited IT Expo few years ago, one of the other colleagues visited two years later. So, this what you collect. We have big network of all universities in this country. We meet each other every eight weeks. This is where we SURF, do you know SURF?

Rashid:
Yes, they make trend reports I think

Jan Willem:
yes, they make reports on higher educations with trends. So, this is what we take in our strategy

Rashid:
So, you meet with the people all around the country in a meeting and you discuss several things which are in report of Gartner and Surf?

Jan Willem:
Yeah, but it's a both, local process and national process, next to each other.

Rashid:
So, then you go to a meeting with Centre of Innovation to discuss those things over there?

Jan Willem:
Yeah, we did this as well.

Rashid:
So, what you normally do? First you discuss in between centre for innovation?

Jan Willem:
So, these processes you do them once every five years or four years. There is no real process, how you do this. Ad-hoc and when we do with in four years, of course we do kind of desk research like Gartner which we collect with our colleague, universities, we go at Centre for

Innovation as well. It's not like we have written out this process and we do very time in a same way.

Rashid:
How is that process, like when you discuss those kinds of things if you want to implement at Leiden University, so how do you do that thing? I have heard that they have been working on VR and AR. So, how those process got selected like this is the time we should experiment with them?

Jan Willem:
Ah… Interesting question! The centre for innovation is there for our own university but not alone. So, they do research on other topics and for other customers as well. So, that everything they investigate is practiced inside our university. So, the cutting-edge innovation which they do research, we are not implementing this cutting-edge thing in our policies. So, we say its ok to do so. Its ok to do single course can do something with this with this VR for example, but this VR is not policy of our university to implement in all courses. So, there is difference. Some trends, they evolve very quickly they are for the total of university. But these cutting-edge trends like VR, we say to Centre for Innovation to innovate, we monitored it.
Centre for Innovation has a model "Maturity of Innovation". Orienting, Experimenting, Scaling and Centre for Innovation in this part (1st two). Before we go to establishing, then me and our broader board, then we the architecture board, we go to look for a several level agreements. If this experiment goes wrong, its ok. We don't say you have to do all the policies and loss because then you dive the innovation (In picture) So over here there is freedom, in this part. In scaling part, we say if we want to continue, use this in our university, then we have to select the suppliers by certain process, you can choose them by yourself. We have to think about archiving things, we have to think about connection with all applications. So, in this part, they are free to choose what they think, if it gets bigger, and get more successful, then we say, now tell us, we will come and look. We have rules from which we go from here to establishing (in the picture – Appendix F – Maturity model of CFI).

Rashid:
Interesting! The thing is, in this process, how they are selecting it?

Jan Willem: They do, I don't. They are looking of course, I can suggest things over here as well and then I do by Gartner as an input, different paper as a input and then we write it down. By the time we write it down, our strategy now, we ask ourselves a question, should we write anything in it? It's not innovation anymore. So, its new for university, but it is not cutting edge. What happened about the subject, I hear about it and we talk about it by letteral things, like did you hear, or did you read this. Would it be interesting for us?
or not? We have quite all as well. We have information architect and I tell him today, that we got an app developed by the professor in Nijmegen. I reviewed by attributes. It's a new way for instead of identification, you only review the attribute you want to, so when you go to rent a video online, it needs to check you over 16. But it doesn't need your identity only to know that you are over 16. So, this app doesn't show you identity, not your name, but only that you are over 16. It's very nice and cutting-edge technology. And I tell him, Wow can we use this, we talk about this in the Internet of Things, it's good for communities with a light pole and streets and that kind of things. Where do we need it over here? Maybe we can see if all student places in library all are occupied by putting sensor in chairs, that kind of things. Should we

develop it my ourselves? No, just wait if suppliers of furniture make this thing then we buy them, but we are not going to innovate. We were just talking and then we decide that we don't do anything with Internet of Things, very big trend in the world, but we don't do anything, but maybe one year somebody find something which makes it very interesting for us besides again to do or not, but there is no real process other than discussing with each other, would it be interesting or not.

Rashid:
What's your role in all of this process? Like you are making policies? Or you advise them?

Jan Willem:
Strategy for IT and research, IT and education, IT and Business administration so what are we going to do in these domains. What can help us in that.

Rashid:
Role with respect to selecting the trend process at Centre for Innovation? Do you make policies, advise them or like that?

Jan Willem:
They advise me more than I advise them. It was centre, real research centre. Its changing now to another place in university, more near to our educational services. And they were free to do what they want, and now they have to be more into the policy of university. So, this is changing right now, these days.

Rashid:
What you think is going well or good in this process?

Jan Willem:
Working together on the personal basis. The previous director, I knew very well and the new director, I shacked hand with her and said we have to meet. You know Gidon?

Rashid: Sorry, No.

Jan Willem: He was previous director and we chatted informal every month or six weeks on these innovations, would it be nice for us to do as well or some funding. So, it's on the personal basis and that's good. But when one-person leaves, Gidon left, then it difficult to get back on same level again. So, you have to build this relation again with the new director. So, the personal thing is good. But the risk is it goes away when somebody leaves.

Rashid:
How you think, this process can be improved?

Jan Willem:
Make it more structural and independent of the persons. So, do meeting every twice a year to discuss this trend.

Rashid:
How you think, a text-mining approach can help this process?

Jan Willem:
There is one more, which is important to me, the strategies from other universities. So, we share quite easily with Nijmegen, Groningen and so the other universities. So, as a source for text-mining, SURF but also the other universities what they are doing. Can it help text-mining? I think it can, but I think it's difficult. It can because if you know our strategy and our ambitions and you can connect trends to our ambitions, it's very interesting. So, for me take our strategy as a source and say Ok, we got an ambition on research data management. So, all the trends, which are, which is data management, FAIR data (Findable, Accessible, …., ….) So, these are different terms, but they mean same research data management FAIR or open-science. So, quite easily with text-mining, filter all the trends which has to do with these ambitions. Then we can say that oh we didn't know that one it could trend to full our ambition. That would be nice and easy.

Then there are two problems: One is our language, there are most trend reports in English, but most strategies are in Dutch. So, it's quite difficult to combine two languages and specially text-mining on a Dutch text is difficult still. So that's the one problem.

So, the other problem is, most of what is said about the research data management, open science is our new ambitions. So, people right a lot of reports about the thing but not really trends. We should invest more money on this. Or it should be more important, or it should be more agenda of board. So, they do a lot of suggestions to make it more important. That's not interesting if it is in your agenda, so you have to filter the real trend. So, there is a lot of being said about research data management and 99% isn't interesting because it should be more important, that's very good, but what's the trend. How to specify what is trend in a lot of communication. When something is trending on a twitter, it's the subject that is trending, isn't a trend by itself.

Rashid:
To cover the problem you just mentioned, if I'm getting through twitter, so I am not checking the only trending topics (hashtags) but extracting actual information related to the subject. If there is discussion about MOOCS, if they are talking about some new application which is related to education purposes. So how they are using it and how many people are reacting on it. So, it includes the all information regarding the subject. So that's information is then transferred into number of questions and statements which will be a topic for experts to discuss.

Jan Willem:
That would be interesting if that's possible. Because you filter from 99% rubbish information, very selective and valuable information because for myself I have to filter the rubbish 99% before getting the real trend so than text-mining is very interesting. How you can determine what is valuable and what is rubbish and is valuable for me is same as for you? So, is there a one algorithm for me, for people at my position or for every person on this earth?

Rashid:
So, the thing is, the idea of this research is to first go to the experts to get their recommendations like you mentioned SURF, Marja mentioned Horizon reports and Gartner

reports, so after getting the information (from experts), these are our sources. Then I am doing text-mining on those reports and then making informational statements (from the result of text mining) for experts. Then doing scenario planning or Delphi technique to have discussion among the experts (to see and plan for the trends in future).

Jan Willem:
I am not sure; the biggest problem is knowing the right trends. Our biggest problem is we have limited capacity, money to realize them. We know what the trends are, which we have to act upon and there is limited time of money. There are more trends but have limited money, so we have to select the top one or two. So, I am not interested in more than two I know. Because the other side is problem. So, on one side it is very interesting if your method and tools helps from lot of information to make it easier and accessible and to select the good trends out of it. It's very interesting. On the other hand, I don't know that it's my biggest problem.

Rashid:
For that side, expert will do the rest of the part, they will discuss what is important and what is not.

Jan Willem:
Do we need the text mining thing if we have the expert, without them, reading before the meeting this much papers (a lot of papers) of course they can get from twitter and reports they get trends. So, if we put them together right know without your method and tool, would or wouldn't the results be the same? So, that's the questions. Of course, your process makes it easier, but people follow news, read articles. Right now, are they able with limited information to select the good and best trends already? Or do they need already this tool? My hypothesis is that they already know the trends, but you can fail that hypothesis. So, think about it. You have two groups of experts that just follow news like all their life and the other is applied with your information and method, how much would the result be better from your group than from peer group.

Rashid:
Thank you so much for your time.

Jan Willem:
If you got more question, send me an email or you can get me on call.

Rashid:
Thank you.

8. Appendix B – Term Frequency (TF) Table

Term Frequency (TF)

1	row ID	Term	TF rel
2	Row1902	automation[NN(POS)]	0.014090344
3	Row1907	digital[JJ(POS)]	0.012349772
4	Row5	netherlands[NNP(POS)]	0.012102874
5	Row0	future[JJ(POS)]	0.011934779
6	Row149	workers[NNS(POS)]	0.011934779
7	Row2227	percent[NN(POS)]	0.011686697
8	Row1975	jobs[NNS(POS)]	0.011438044
9	Row5720	occupations[NNS(POS)]	0.011213401
10	Row16815	skills[NNS(POS)]	0.011054395
11	Row79	market[NN(POS)]	0.010590015
12	Row127	technology[NNP(POS)]	0.010085729
13	Row5648	skills[NNS(POS)]	0.009552156
14	Row47	technology[NN(POS)]	0.009413347
15	Row1901	pagina[NNP(POS)]	0.00890906
16	Row1990	technology[NN(POS)]	0.00870286
17	Row2014	employment[NN(POS)]	0.008537091
18	Row196	key[NNP(POS)]	0.008404774
19	Row3	people[NNS(POS)]	0.008068583
20	Row6362	occupations[NNP(POS)]	0.007994739
21	Row1920	productivity[NN(POS)]	0.0079569
22	Row1945	ai[NNP(POS)]	0.007874016
23	Row16787	future[NNP(POS)]	0.007609683
24	Row14	future[NN(POS)]	0.007564297
25	Row77	dutch[NNP(POS)]	0.007564297
26	Row143	solutions[NNS(POS)]	0.007564297
27	Row16771	industry[NN(POS)]	0.007547052
28	Row24	skills[NNS(POS)]	0.007396201
29	Row16413	report[NNP(POS)]	0.007359158

9. Appendix C – (TF-IDF Table)

Term Frequency – Inverse Document Frequency (TF-IDF)

1	row ID	Term	TF rel	IDF	tf-idf
2	Row120	lexibility[NN(POS)]	0.007228106	0.698970004	0.005052229
3	Row1902	automation[NN(POS)	0.014090344	0.301029996	0.004241616
4	Row1907	digital[JJ(POS)]	0.012349772	0.301029996	0.003717652
5	Row0	future[JJ(POS)]	0.011934779	0.301029996	0.003592726
6	Row149	workers[NNS(POS)]	0.011934779	0.301029996	0.003592726
7	Row93	echnology[NN(POS)]	0.005042864	0.698970004	0.003524811
8	Row14619	abilities[NNP(POS)]	0.007302554	0.477121255	0.003484204
9	Row1975	jobs[NNS(POS)]	0.011438044	0.301029996	0.003443194
10	Row9	index[NN(POS)]	0.004874769	0.698970004	0.003407317
11	Row5720	occupations[NNS(PO	0.011213401	0.301029996	0.00337557
12	Row16815	skills[NNS(POS)]	0.011054395	0.301029996	0.003327705
13	Row165	roadmap[NNP(POS)]	0.004706673	0.698970004	0.003289824
14	Row79	market[NN(POS)]	0.010590015	0.301029996	0.003187912
15	Row127	technology[NNP(POS	0.010085729	0.301029996	0.003036107
16	Row8354	knowledge[NNP(POS	0.006333495	0.477121255	0.003021845
17	Row6362	occupations[NNP(PO	0.007994739	0.367976785	0.002941879
18	Row241	opportunities[NNPS(I	0.004202387	0.698970004	0.002937342
19	Row564	pension[NN(POS)]	0.004202387	0.698970004	0.002937342
20	Row1945	ai[NNP(POS)]	0.007874016	0.367976785	0.002897455
21	Row5648	skills[NNS(POS)]	0.009552156	0.301029996	0.002875486
22	Row47	technology[NN(POS)]	0.009413347	0.301029996	0.0028337
23	Row14223	workers[NNPS(POS)]	0.005883574	0.477121255	0.002807178
24	Row97	inequality[NNP(POS)]	0.007228106	0.367976785	0.002659775
25	Row1990	technology[NN(POS)]	0.00870286	0.301029996	0.002619822
26	Row2014	employment[NN(POS	0.008537091	0.301029996	0.00256992
27	Row110	threats[NNS(POS)]	0.006723819	0.367976785	0.002474209
28	Row132	nequality[NN(POS)]	0.003530005	0.698970004	0.002467368
29	Row3	people[NNS(POS)]	0.008068583	0.301029996	0.002428885

10. Appendix D – (Co-Occurrence Table)

<u>Co-Occurrence Table</u>

Co-occurrence table - 0:52 - Term Co-Occurrence Counter

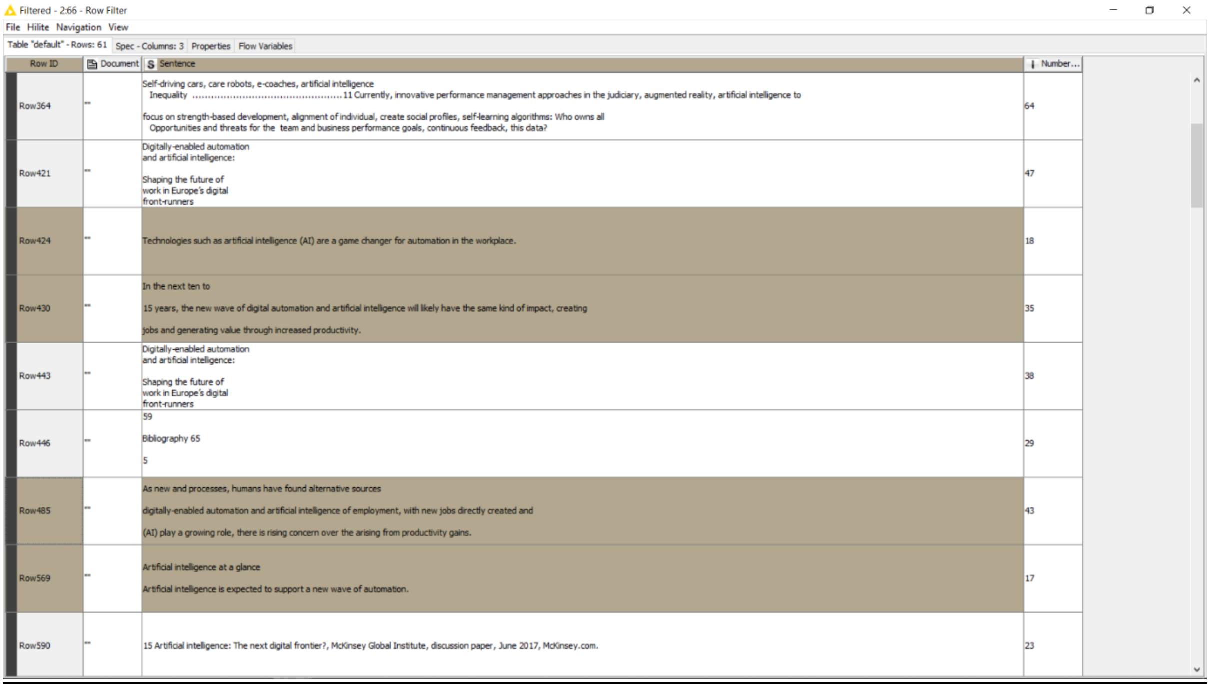
File Hilite Navigation View

Table "default" - Rows: 668790 | Spec - Columns: 6 | Properties | Flow Variables

Row ID	Preproc...	Term1	Term2	Sentence cooccurrence	Neighbor count
Row379822	""	future[NNP(POS)]	jobs[NNP(POS)]	188	182
Row374446	""	jobs[NNP(POS)]	report[NNP(POS)]	159	149
Row154592	""	employment[NNP(POS)]	skills[NNS(POS)]	79	77
Row374453	""	economic[NNP(POS)]	world[NNP(POS)]	82	77
Row154591	""	future[NNP(POS)]	skills[NNS(POS)]	77	75
Row374454	""	economic[NNP(POS)]	forum[NNP(POS)]	70	69
Row395845	""	emerging[VBG(POS)]	markets[NNS(POS)]	70	63
Row386527	""	basic[NNP(POS)]	infrastructure[NNP(POS)]	64	61
Row386577	""	communication[NNP(POS)]	technology[NNP(POS)]	65	60
Row392223	""	families[NNS(POS)]	job[NN(POS)]	74	60
Row397897	""	changing[NNP(POS)]	nature[NN(POS)]	56	56
Row542063	""	customers[NNS(POS)]	total[JJ(POS)]	30	50
Row424182	""	overall[JJ(POS)]	specialists[NNS(POS)]	33	48
Row610336	""	ease[VB(POS)]	relative[NNP(POS)]	32	48
Row386576	""	communication[NNP(POS)]	information[NNP(POS)]	68	47
Row153743	""	employment[NNP(POS)]	skills[NNP(POS)]	49	46
Row376368	""	gap[NNP(POS)]	gender[NNP(POS)]	47	46
Row197656	""	personal[NNP(POS)]	service[NNP(POS)]	67	45
Row394967	""	class[NN(POS)]	emerging[VBG(POS)]	47	45
Row396401	""	cloud[NN(POS)]	technology[NN(POS)]	52	45
Row397898	""	flexible[JJ(POS)]	nature[NN(POS)]	51	45
Row433191	""	demand[NN(POS)]	skills[NNS(POS)]	61	45
Row197655	""	customer[NNP(POS)]	personal[NNP(POS)]	47	44
Row397900	""	class[NN(POS)]	middle[NNP(POS)]	43	43
Row408328	""	related[NNP(POS)]	sales[NNS(POS)]	52	42
Row468288	""	roles[NNS(POS)]	staff[NNP(POS)]	21	42
Row153742	""	future[NNP(POS)]	skills[NNP(POS)]	44	41
Row380970	""	current[JJ(POS)]	recruitment[NN(POS)]	72	41
Row396395	""	internet[NN(POS)]	mobile[NNP(POS)]	41	41
Row408107	""	computer[NNP(POS)]	mathematical[NNP(POS)]	47	41
Row515846	""	job[NNP(POS)]	main[NNP(POS)]	44	41
Row381074	""	family[NN(POS)]	job[NN(POS)]	53	40
Row260334	""	occupations[NNP(POS)]	service[NNP(POS)]	83	39
Row374434	""	employment[NNP(POS)]	skills[NNP(POS)]	60	39
Row395019	""	change[NN(POS)]	climate[NNP(POS)]	39	39
Row396621	""	cloud[NN(POS)]	internet[NN(POS)]	40	39
Row408184	""	architecture[NNP(POS)]	engineering[NNP(POS)]	43	39
Row386585	""	entertainment[NNP(POS)]	media[NNP(POS)]	66	38
Row390744	""	family[NNP(POS)]	job[NNP(POS)]	39	38
Row386598	""	professional[NNP(POS)]	services[NNP(POS)]	36	37
Row395234	""	natural[JJ(POS)]	resources[NNS(POS)]	40	37
Row397904	""	geopolitical[NNP(POS)]	volatility[NN(POS)]	35	37
Row397918	""	power[NN(POS)]	processing[NNP(POS)]	42	37
Row466490	""	gap[NN(POS)]	wage[NN(POS)]	37	37
Row545312	""	aallrreeaaddyy[JJ(POS)]	ffeelltt[NN(POS)]	19	37
Row197510	""	comprehension[NNP(POS)]	oral[NNP(POS)]	47	36

11. Appendix E – Sentence Extraction Example

<u>Sentence Extraction (based on words "Artificial Intelligence")</u>

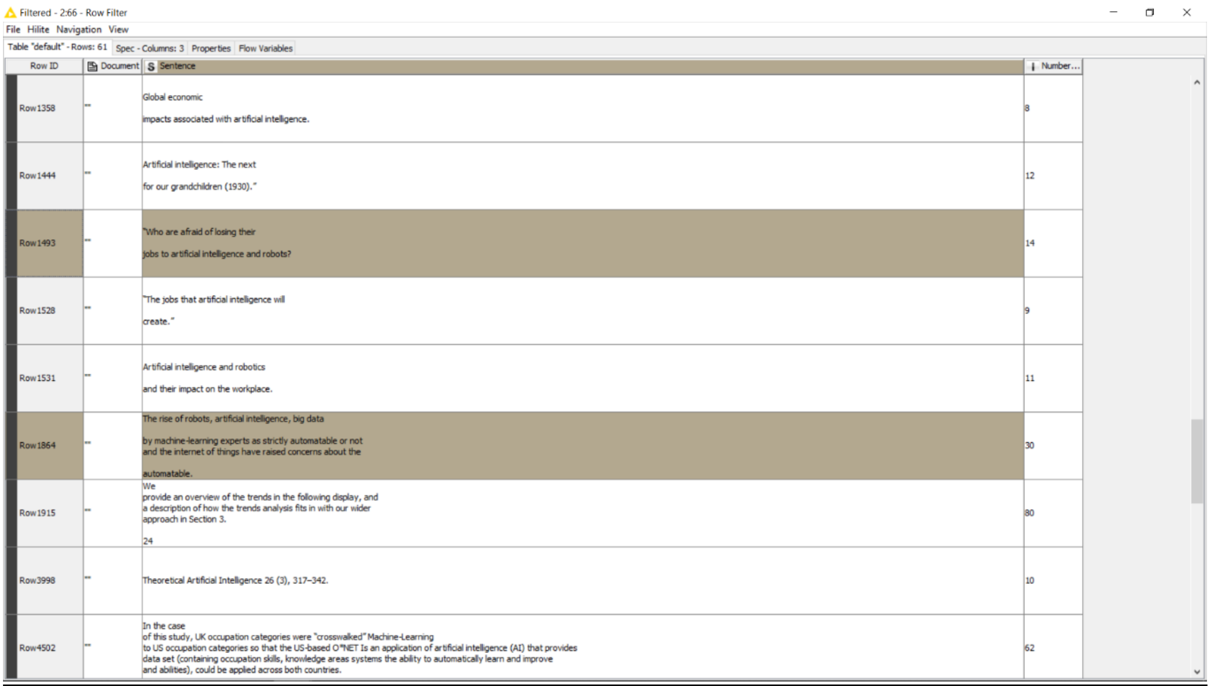


Filtered - 2:66 - Row Filter

File Hilite Navigation View

Table "default" - Rows: 61 | Spec - Columns: 3 | Properties | Flow Variables

Row ID	Document	Sentence	Number...
Row364	""	Self-driving cars, care robots, e-coaches, artificial intelligence Inequality ....11 Currently, innovative performance management approaches in the judiciary, augmented reality, artificial intelligence to focus on strength-based development, alignment of individual, create social profiles, self-learning algorithms: Who owns all Opportunities and threats for the team and business performance goals, continuous feedback, this data?	64
Row421	""	Digitally-enabled automation and artificial intelligence: Shaping the future of work in Europe's digital front-runners	47
Row424	""	Technologies such as artificial intelligence (AI) are a game changer for automation in the workplace.	18
Row430	""	In the next ten to 15 years, the new wave of digital automation and artificial intelligence will likely have the same kind of impact, creating jobs and generating value through increased productivity.	35
Row443	""	Digitally-enabled automation and artificial intelligence: Shaping the future of work in Europe's digital front-runners	38
Row446	""	59 Bibliography 65 5	29
Row485	""	As new and processes, humans have found alternative sources digitally-enabled automation and artificial intelligence of employment, with new jobs directly created and (AI) play a growing role, there is rising concern over the arising from productivity gains.	43
Row569	""	Artificial intelligence at a glance Artificial intelligence is expected to support a new wave of automation.	17
Row590	""	15 Artificial intelligence: The next digital frontier?, McKinsey Global Institute, discussion paper, June 2017, McKinsey.com.	23



Filtered - 2:66 - Row Filter

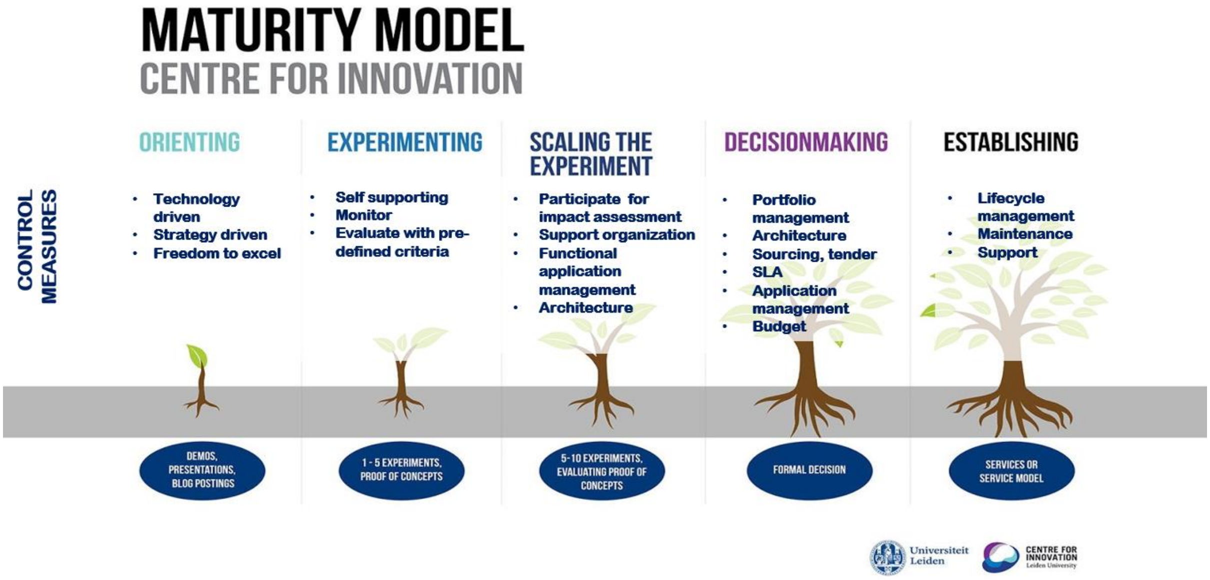
File Hilite Navigation View

Table "default" - Rows: 61 | Spec - Columns: 3 | Properties | Flow Variables

Row ID	Document	Sentence	Number...
Row1358	""	Global economic impacts associated with artificial intelligence.	8
Row1444	""	Artificial intelligence: The next for our grandchildren (1930)."	12
Row1493	""	"Who are afraid of losing their jobs to artificial intelligence and robots?	14
Row1528	""	"The jobs that artificial intelligence will create."	9
Row1531	""	Artificial intelligence and robotics and their impact on the workplace.	11
Row1864	""	The rise of robots, artificial intelligence, big data by machine-learning experts as strictly automatable or not and the internet of things have raised concerns about the automatable.	30
Row1915	""	We provide an overview of the trends in the following display, and a description of how the trends analysis fits in with our wider approach in Section 3. 24	80
Row3998	""	Theoretical Artificial Intelligence 26 (3), 317–342.	10
Row4502	""	In the case of this study, UK occupation categories were "crosswalked" Machine-Learning to US occupation categories so that the US-based O*NET Is an application of artificial intelligence (AI) that provides data set (containing occupation skills, knowledge areas systems the ability to automatically learn and improve and abilities), could be applied across both countries.	62

12. Appendix F – Maturity Model of CFI

Maturity Model of Centre for Innovation



13. Appendix G – Playbook (Online Learning Lab Workshop)

Playbook for Online Learning Lab Workshop (September 27th, 2018)

Yellow highlight	Start, break, end
Grey highlight	3 minutes max
Normal	5 minutes max

TIME	ACTIVITY
15:00 – 15:15	Welcome and Introduction
15:15 – 15:20	Login in Socrative and begin with Q1 – characteristics of OOL's learner
15:20 – 15:25	Q2 – Write technological trends related to OLL in next 5-10 years
15:25 – 15:28	Q3 – Identifying *Top 10 Technological Trends*
15:28 – 15:31	*Q4 – Identify trends from Q3, which need to be observed*
15:31 – 15:34	*Q5 – Identify trends from Q3, which need to be experiment*
15:34 – 15:37	*Q6 – Identify trends from Q3, which need to be implemented*
15:37 – 15:42	Q7 – Write educational trends related to OLL in next 5-10 years
15:42 – 15:45	Q8 – Identify *Top 10 Educational trends*
15:45 – 15:48	*Q9 – Identify trends from Q8, which need to be observed*
15:48 – 15:51	*Q10 – Identify trends from Q8, which need to be experiment*
15:51 – 15:53	*Q11 – Identify trends from Q8, which need to be implemented*
15:53 – 16:00	**BREAK**
16:00 – 16:05	Q12 – What are the strengths of OLL
16:05 – 16:10	Q13 – What are the weaknesses of OLL
16:10 – 16:15	Q14 – What are the opportunities for OLL
16:15 – 16:20	Q15 – What are the threats for OLL
16:20 – 16:25	Q16 – What Political trends/factors can affect OLL
16:25 – 16:30	Q17 – What Economic trends/factors can affect OLL
16:30 – 16:35	Q18 – What Social trends/factors can affect OLL
16:35 – 16:40	Q19 – What Legal trends/factors can affect OLL
16:40 – 16:43	***In next year***, *which trend(s) OLL should* ***Observe****?*
16:43 – 16:46	***In next year***, *which trend(s) OLL should* ***Experiment****?*
16:46 – 16:49	***In next year***, *which trend(s) OLL should* ***Implement****?*
16:49 – 17:00	Closing and Feedback Questionnaires

14. Appendix H – Playbook (Future Work Lab Workshop)

Playbook for Future Lab Workshop (October 19th, 2018)

Yellow highlight	Start, break, end
Grey highlight	3 minutes max
Normal	5 minutes max

TIME	ACTIVITY
10:00 – 10:15	Welcome and Introduction
10:15 – 10:20	Login in Socrative and begin with Q1
10:20 – 10:25	Q2 – Write trends related to FUTURE WORK LAB in next 5-10 years
10:25 – 10:28	Q3 – Identifying *Top 10 Trends*
10:28 – 10:31	*Q4 – Identify trends from Q3, which need to be observed*
10:31 – 10:34	*Q5 – Identify trends from Q3, which need to be experiment*
10:34 – 10:37	*Q6 – Identify trends from Q3, which need to be implemented*
~~**10:37 – 10:42**~~	~~Q7 – Write educational trends related to FUTURE WORK LAB in next 5-10 years~~
~~**10:42 – 10:45**~~	~~Q8 – Identify *Top 10 Educational trends*~~
~~**10:45 – 10:48**~~	~~*Q9 – Identify trends from Q8, which need to be observed*~~
~~**10:48 – 10:51**~~	~~*Q10 – Identify trends from Q8, which need to be experiment*~~
~~**10:51 – 10:53**~~	~~*Q11 – Identify trends from Q8, which need to be implemented*~~
10:53 – 11:00	**BREAK**
11:00 – 11:05	Q12 – What are the strengths of FUTURE WORK LAB
11:05 – 11:10	Q13 – What are the weaknesses of FUTURE WORK LAB
11:10 – 11:15	Q14 – What are the opportunities for FUTURE WORK LAB
11:15 – 11:20	Q10 – What are the threats for FUTURE WORK LAB
11:20 – 11:25	Q11 – What Political trends/factors can affect FUTURE WORK LAB
11:25 – 11:30	Q17 – What Economic trends/factors can affect FUTURE WORK LAB
11:30 – 11:35	Q18 – What Social trends/factors can affect FUTURE WORK LAB
11:35 – 11:40	Q19 – What Legal trends/factors can affect FUTURE WORK LAB
11:40 – 11:43	***In next year****, which trend(s) FUTURE WORK LAB should* ***Observe****?*
11:43 – 11:46	***In next year****, which trend(s) FUTURE WORK LAB should* ***Experiment****?*
11:46 – 11:49	***In next year****, which trend(s) FUTURE WORK LAB should* ***Implement****?*
11:49 – 12:00	Closing and Feedback Questionnaires

15. Appendix I – Workshop Questionnaires (OLL)

Workshop Questionnaires (Online Learning Lab)

Please circle your response to the items. Rate aspects of the workshop on a 1 to 5 scale. Your feedback is sincerely appreciated. Thank you.

WORKSHOP CONTENT (Circle your response to each item.)
1=Strongly disagree 2=Disagree 3=Neither agree nor disagree 4=Agree 5=Strongly agree N/A=Not applicable

1. **I was well informed about the objectives of this workshop.**

1	2	3	4	5	N/A

2. **This workshop lived up to my expectations.**

1	2	3	4	5	N/A

3. **The content is relevant to my job.**

1	2	3	4	5	N/A

WORKSHOP DESIGN (Circle your response to each item.)
1=Strongly disagree 2=Disagree 3=Neither agree nor disagree 4=Agree 5=Strongly agree N/A=Not applicable

4. **The workshop objectives were clear to me.**

1	2	3	4	5	N/A

5. **The workshop activities stimulated my learning.**

1	2	3	4	5	N/A

6. **The activities in this workshop gave me enough time to work on (while doing analysis i.e., SWOT and PESTEL)**

1	2	3	4	5	N/A

7. **The difficulty level of this workshop was appropriate.**

1	2	3	4	5	N/A

8. **What you think is good about this method? (Reports + social media + Experts)**

____

____

9. **What improvements would you recommend in this method?**

____

____

10. What improvements would you recommend in this method?

____

____

11. What is least valuable about this method?

____

____

12. What is most valuable about this method?

____

____

13. What would need to be adjusted, to make this trend setting workshop suitable for OLL on a long term?

____

____

____

14. Remarks about the methodology (Reports + social media + Experts)

____

____

____

____

15. How would you improve this workshop? (Check all that apply.)

___Provide better information before the workshop.
___Clarify the workshop objectives.
___Reduce the content covered in the workshop.
___Increase the content covered in the workshop.
___Update the content covered in the workshop.
___Improve the instructional methods.
___Make workshop activities more stimulating.
___Improve workshop organization.
___Make the workshop less difficult.
___Make the workshop more difficult.
___Slow down the pace of the workshop.
___Speed up the pace of the workshop.
___Allot more time for the workshop.
___Shorten the time for the workshop.

16. Any general remarks for workshop, which could be helpful for future?

16. Appendix J – Workshop Questionnaires (FWL)

Workshop Questionnaires (Future Work Lab)

Please circle your response to the items. Rate aspects of the workshop on a 1 to 5 scale. Your feedback is sincerely appreciated. Thank you.

WORKSHOP CONTENT (Circle your response to each item.)

1=Strongly disagree 2=Disagree 3=Neither agree nor disagree 4=Agree 5=Strongly agree

1. **I was well informed about the objectives of this workshop.**

1	2	3	4	5

2. **This workshop lived up to my expectations.**

1	2	3	4	5

3. **The content is relevant to my job.**

1	2	3	4	5

WORKSHOP DESIGN (Circle your response to each item.)

1=Strongly disagree 2=Disagree 3=Neither agree nor disagree 4=Agree 5=Strongly agree N/A=Not applicable

4. **The workshop objectives were clear to me.**

1	2	3	4	5

5. **The workshop activities stimulated my learning.**

1	2	3	4	5

6. **The activities in this workshop gave me enough time to work on (while doing analysis i.e., SWOT)**

1	2	3	4	5

7. **The difficulty level of this workshop was appropriate.**

1	2	3	4	5

8. **What you think is good about this method? (Reports + social media + Experts)**

____

____

9. **What improvements would you recommend in this method? (Reports + social media + Experts)**

____

____

10. **What is least valuable about this method? (Reports + social media + Experts)**

____

____

11. **What is most valuable about this method? (Reports + social media + Experts)**

____

____

12. **What would need to be adjusted, to make this trend setting workshop suitable for Future Work Lab on a long term?**

____

____

____

13. **Remarks about the methodology (Reports + social media + Experts)**

____

____

____

____

14. **How would you improve this workshop? (Check all that apply.)**

___Provide better information before the workshop.
___Clarify the workshop objectives.
___Reduce the content covered in the workshop.
___Increase the content covered in the workshop.
___Update the content covered in the workshop.
___Improve the instructional methods.
___Make workshop activities more stimulating.
___Improve workshop organization.
___Make the workshop less difficult.
___Make the workshop more difficult.
___Slow down the pace of the workshop.
___Speed up the pace of the workshop.
___Allot more time for the workshop.
___Shorten the time for the workshop.

15. **Any general remarks for workshop, which could be helpful for the future.**