How to predict the development of breakthrough technologies with the help of electronic databases?

An explorative study

L.C.G. Verkruijsse

December 2010

How to predict the development of breakthrough technologies with the help of electronic databases?

An explorative study

MASTER THESIS

Author: L.C.G. Verkruijsse (1541609)

MSc Management of Technology

Graduation Committee:

Chairman: Prof. Dr. C.P. van Beers
Professor, Head of Section Technology, Strategy & Entrepreneurship

First supervisor: Dr. J.R. Ortt
Associate Professor, Section Technology, Strategy & Entrepreneurship

Second supervisor: Dr. S.W. Cunningham
Assistant Professor, Section Policy Analysis

MSc Management of Technology

Faculty of Technology, Policy and Management

Delft University of Technology

December, 2010

ABSTRACT

Breakthrough technologies can be defined by ‘new-to-the-world’ or ‘radical (improved)’ technologies which have the capacity to change the behaviour of end-users. The journey these technologies practise towards the mainstream market can be regarded as a dynamic process with lots of uncertainties. Companies investing in the development of these technologies face some serious risks. For the managers of these companies it would be of tremendous value if they could, even in the slightest way, make strategic decisions supported by reliable forecasts.

This research aims to investigate the added value of electronic databases in determining the chances of succeeding in the market. Different kinds of electronic databases can measure the activity on a specific topic, which subsequently can be used in forecasting whether the activity will increase or not. This information, in combination with current forecasting methods, can be applied in a business intelligence tool; a tool supporting the decision making process of managers. One of these databases, besides the scientific and patent databases, is offered by Google News and includes business press and news articles from many different sources. This database indicates the activity and popularity on a particular topic among future consumers. Because of its potential, this database is included in this research as well.

To answer this challenging question about the added value of electronic databases, two analyses were performed using data from 14 breakthrough technologies in the material- and pharmaceutical industry. The *first* analysis included different viewpoints in literature on scientific-, technological-, and market activity and when the databases appear to show the highest activity over the life-cycle of a technology. Then, the analysis based on these 14 cases, is used as verification. As a result, it became clear that scientific and market activity increases over time in parallel.

The *second* analysis focused on a completely different aspect. A further dive was made into the history of these technologies, looking for a correlation between the patterns generated by databases and the historical patterns. Remarkably, about 50% of the cases showed a correlation with the patterns generated by Google News. Although this result seems initially not significant, future research is proposed where even higher results might be found. Then, this database might be of added value for future forecasting tools.

This explorative study adds new and improved perspectives on scientific and managerial aspects. It contributes to the concept of forecasting the development of breakthrough technologies. Also, it clearly shows the added value of electronic databases and what they could mean for future research. Nevertheless, this study bears with some limitations. The small sample size, the focus on only two industries, noise in the data, and the lack of more effective queries during the search ensure an inevitably bias in the results. However, the explorative nature of this study does supply the first large building block on this topic, which will be used in future research.

Keywords: *technology life-cycle, large-scale production and diffusion, breakthrough technologies, forecasting methods, business intelligence, prediction of technological diffusion.*

ACKNOWLEDGEMENTS

Initially, this thesis addressed a clear and intriguing question about electronic databases and *how* these could be used in forecasting tools. It took some time before I realised that this question raised more and more question marks about undiscovered areas. The proposed question was simply too ambitious to be answered in this Master thesis alone after I muddled through hundreds of literature pages and several discussions with my supervisors and fellow students. I realised it was important to start from the beginning, build a clear building block first on this topic, and focus on the core question about electronic databases and whether they could deliver any added value at all for forecasting tools. In a way, the research shifted completely to a new goal, however I enjoyed this dynamic and instructive process and I am satisfied and pleased with this unique learning experience.

For the pleasure I experienced in this process, I am indebted to a large amount of people I would like to express my gratitude. My first special gratitude goes to Dr. Roland Ortt for supporting me in this journey, guiding me in a pleasant way and allowing me to exploit my personal interests. Of course I am also grateful for Dr. Scott Cunningham, who always had some very interesting viewpoints and useful comments on the topic I was working on.

One person in particular I would like to thank is Toos, who always supported and stimulated me in difficult times. Furthermore, I have to thank two dear friends of mine, Nick and Suprapto, with whom I worked with for the last couple of months in the same workspace in the ‘corner’ of the faculty and who always had inspiring ideas and great inputs during this process. Thank you both for leaving your fingerprints in this thesis work.

I also would like to thank some fellow students from my Master study. Robert, thank you for your time and reflection on my research and the enjoyable lunches we had on a weekly basis. Sjoerd, thank you for your insights, your ways to keep me focused on my work and your reflection during my final weeks. Also, David and Trung, thanks for the nice discussions and for motivating me when this was necessary.

At last, but not the least, my special thanks goes to Joyce who made a marvellous cover of this Master thesis, and to my dear parents who made it possible for me to complete this Master study.

Delft, 18th of November 2010

L.C.G. Verkruijsse

TABLE OF CONTENTS

TABLE OF FIGURES

TABLE OF TABLES

1. INTRODUCTION

1.1 Research problem

Before we dive into the research problem we would like to explain shortly the *vision* we have in mind where we would like to end up at the end of this paragraph. This vision entails that market research does not fulfil the requirements managers need nowadays to support their strategic decisions on the development of breakthrough technologies. As a result, alternatives are needed to create new opportunities to stay ahead of competitors. One of these alternatives is with the use of new forecasting tools, but this will be explained in a later stage of this introduction.

1.1.1 The development of breakthrough technologies

In the past there have been many breakthrough technologies that were accepted in the market after commercialisation. All of them walked their own path of diffusion before they became successful. This path can be drawn in a pattern of diffusion (also known as a life-cycle pattern) which is divided in three phases; the innovation phase, the adaptation phase, and the market stabilisation phase. The pattern looks like an S-shaped curve in which it represents the breakthrough technology and how the cumulative adoption looks like over time. Figure 1 illustrates how the phases are divided (Ortt, Shah et al. 2007). The phases are separated by three *hallmarks*: invention, market introduction, and large-scale production and diffusion. This technology life-cycle will be further elaborated in the literature study of chapter 2; however, it will form the basis in this thesis.

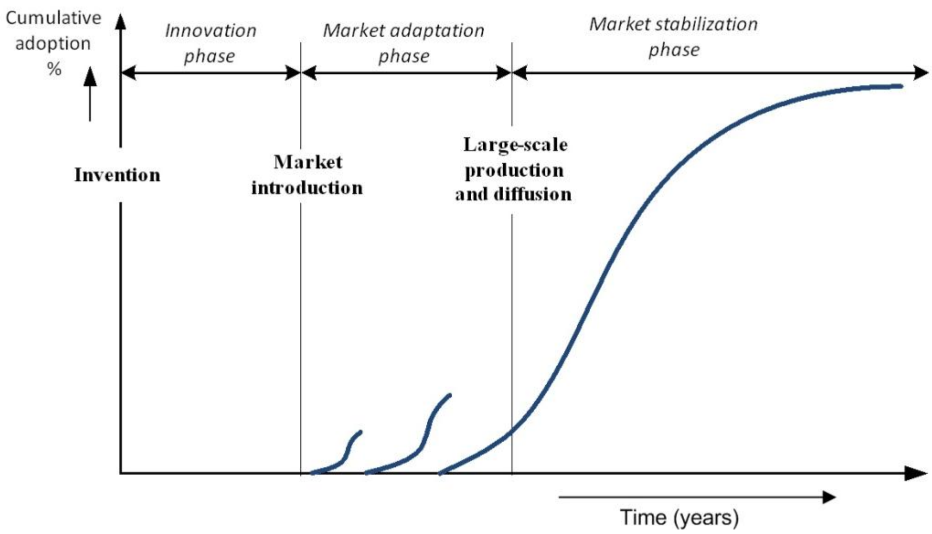

Figure 1 – The technology life-cycle pattern with the three phases (Ortt *et al.* 2004, 2007, 2009).

1.1.2 Reasons for knowing the moment of occurrence of the hallmarks

Whatever the strategy is for launching a breakthrough innovation in the mainstream market, firms often lack the knowledge about the type of pattern that emerges after the invention. Therefore it is quite hard to predict the exact moment of market introduction (hallmark 2) and when large-scale production and diffusion is about to start (hallmark 3). Knowing when these hallmarks occur would be preferable for them, so that they can prepare themselves in many different aspects. Companies usually cannot control the technology life-cycle; however, they can react to it. Only companies that have a monopolistic position sometimes have the power to create the pattern, instead of reacting to it.

Once a company knows what it can expect about the diffusion of a new breakthrough technology, they can pursue their opportunities and diminish certain risks by reacting fast and accurate. Lots of companies pursue the technology life-cycle pattern, trying to stay in front of their competitors. They are challenged to remain innovative in a fast-changing competitive environment (Trumbach, Payne et al. 2006). Once a company realizes that they are behind schedule and cannot react in a purposefully way then they will often cancel the project. Especially in a dynamic and competitive environment it is key to have knowledge about when to enter the market in order to balance the risks of premature entry against the missed opportunity of late entry (Lilien and Yoon 1990). Preparations that are needed for mass-production consist of many different steps.

First, a production plant is needed for production which requires several months before this plant is installed. Installing this plant too soon is not desirable due to the high-investments that are required upfront while these investments are also needed for other actions (e.g. development costs and marketing costs), especially in the innovation phase. More risks emerge when a production plant is installed too early, because changes in the technology might still occur. The production plant might need to be adjusted then, which would result in even more costs. Moreover, there is always the possibility that a project will be cancelled which would result in an unusable plant which is already installed. *Second*, a good marketing campaign is of paramount importance in order to make the consumers aware of this new technology. Several months, or even years are often planned for this. It becomes clear that the hallmark 'market introduction' should be well-known once the marketing campaign starts. Any delays of the market introduction are not acceptable. The risk of investing in marketing on the wrong moment will be reduced when more information is known about the occurrence of the hallmarks. *Third*, having the knowledge of when your product is going to be launched in the market and when mass-diffusion is going to take-off brings other advantages with it, although this is not a guarantee. A company might gain a competitive advantage by being the first on the market with this new technology. They might create more credibility with it compared to their competitors or a large pool of customers can be created upfront. The same holds for complementary products. Many high-tech products only succeed in the market due to its large variety of complementary products which offers the customer many possibilities. The complementary products support the strong position of the main product. Complementary products sometimes lead to lock-in, which eventually may lead to a leading position of the firm. However, it takes time to found out what the customer requires in complementary products. Also, it takes time to prepare production facilities of these products. Complementary products can often be integrated in the production plant of the main technology, so it would be far more preferable if the complementary product is worked out in detail as soon as possible. *Fourth*, a company needs to prepare in logistic issues. They need to build a logistic network consisting of buffers and connections. Especially when the hallmark-dates are well-known the chance of a smooth distribution without any delays of the mass-product can be guaranteed.

All these factors show that it is more than relevant to know when the hallmarks will occur. Again, singular companies cannot control the technology life-cycle, but they do have the possibility to react to it. This is exactly why it would be beneficial to have more knowledge about where you are at the technology life-cycle. However, there are some cases known where the technology life-cycle influenced by companies with sufficient market power. The multinational manufacturer of a wide range of electronic products for both consumer and business markets, Philips, is one of those companies with a lot of market power. Philips has many different independent business units which are operating in distinct markets. Philips sometimes operates during the developments of new products with business alliances. By corporate venturing, Philips is able to align with leading start-up companies and to benefit from massive industry R&D. Due to the decentralized business units and by corporate venturing, Philips sometimes manages to determine the technology life-cycle pattern itself. They have the power to wipe away competitive markets which consist of single companies without sufficient market power. One example of this phenomenon is the alliance between Philips and Douwe Egberts and the development of a new coffee machine called Senseo. It may be clear that these examples will not benefit from any prediction methods, due to the fact that they can influence the market themselves. Basically, they can create the technology life-cycle pattern themselves. These situations will therefore be excluded in this research.

1.1.3 Market research and its limitations

Unfortunately, there are only a limited amount of models that are capable to make any predictions for breakthrough innovations. Market research seems to be the best grip for a company to make any predictions for the future. A complete analysis of projected life-cycle costs and development time is only partly reliable due to the fact that market research is mainly based on the future customer interests. The problem here is that the customer often says contradicting things during market research than they would do in real life. For instance, if a customer says that he/she is willing to buy a new high-tech product that will be launched in the market, this does not actually mean that he/she would actually buy the product once it is launched in the market (Zeithaml 1988). Also, the information in high-technology markets is time sensitive and therefore quickly loses its value. A company cannot make reliable predictions only based on market research due to these issues. There are several barriers that limit market research in general (Mohr, Slater et al. 2006). *First*, there are strategic barriers like the lack of resources, the lack of strategic direction, the lack of marketing expertise, etc. Good personnel and a reasonably large amount of money are necessary in order to achieve reliable results. *Second*, there are functional barriers which involve the lack of coordination between different departments inside the company. *Third*, there is often a lack of market-based information which makes the results even more unreliable. *Fourth*, there is a lack of customer focus and *last* there are tactical barriers which interfere. Moreover, market research is a time-consuming method which is complex and costly.

However, market research for breakthrough innovations shows an even lower validity. There are three main methods which are proposed to predict the adoption of innovations: *consumer, expert,* and *data analysis* (Taschner 1999; Armstrong 2001a; Meade and Islam 2006). Each of these categories has certain difficulties which might result in the unsuccessful diffusion of a breakthrough technology (Tauber 1974; Hippel 1986; Langley, Pals et al. 2009). They will be discussed shortly below.

Consumer research

First, consumer research has limitations concerning breakthrough technologies and the fact that initially they seem to plead for very specific customer segments. It often occurs that these segments have not even been acknowledged (yet); consumer researchers often use a general sample of customers in order to determine the future success of the breakthrough technology. However, this sample results in a negative evaluation of the breakthrough technology.

The second difficulty in consumer research is that breakthrough innovations sometimes create a change in behaviour that is needed in order to launch this innovation successfully. However, this change in behaviour tends to evoke consumer resistance, which slowly dissolves when the breakthrough innovation is actually used. When large groups of consumers show resistance towards a certain breakthrough technology, it might constrain the further development of the technology.

Last, for potential consumers it is difficult to recognize the benefits of a breakthrough technology, especially in the early stage of the development. Only descriptions and prototypes are available for consumer research, however, for a valid analysis it is required to actually use the technology in order to experience the effects of the breakthrough technology in daily life.

Expert research

There are also difficulties when applying the opinions of experts with regard to breakthrough innovations. Often, breakthrough innovations change the consumption pattern. Similarly, a change in consumer segments occurs in the market. Therefore, experts simply cannot make reliable forecasts based on their knowledge and current information they possess. Extrapolating the existing segments and their preferences towards a new situation is not sufficient to make reliable forecasts.

Experts often tend to be easily influenced to bias and inconsistency. As a result, the accuracy of this forecasting method is doubtful, because experts seem to emphasize too much on their own predictions. Therefore, this expert research is not well known for its practical results (Brenner, Koehler et al. 1996; Arkes 2001).

Data analysis

Data analysis also meets some vital difficulties when it is applied to breakthrough innovations. The past data that is collected from predecessors of innovations should not be extrapolated in order to make forecasts for the latest breakthrough innovation, due to the shift in customer behaviour. Of course, comparison with similar products (e.g. in other countries) can be helpful, however, each dissimilarity in consumer patterns will diminish the quality of the forecasts (Armstrong 2001b).

1.1.4 Conclusion

From an overall viewpoint it also becomes clear that market research has a validity problem due to the fact that this research is not useful in every phase of the technology life-cycle. When we look at the technology life-cycle pattern it becomes clear that market research cannot predict much about the developments of a new technology in the innovation phase, because consumers are not even aware of this technology (yet). The methods assume that the potential consumers which are invited in market research understand the product and its use and are able to understand the consequences. It is often also assumed that the conditions in the pre-diffusion phase will remain essentially the same in market circumstances; however, this is often not the case. Therefore, these predicting methods do no longer hold. The consequences of the limited added value of market research can be devastating for companies. Pioneers (further explained in chapter 2) try to be the first in the market with their breakthrough technology, but face many risks due to high investments, uncertainly (about the length of the phases), and the acceptance of the consumers. Market research does not diminish these risks significantly (Ortt, Langley et al. 2007). Therefore, it is time to discover new and useful methodologies which could replace the obsolete methods of forecasting.

For a company it would be far more beneficial to have a simple business intelligence tool (further explained in chapter 2), which can be used continuously during every life-cycle phase and while still offering reliable results for only a small price. This kind of tool uses many different available sources and tries to forecast the future developments and the environment. The percentage of the success rate of the predictions of this business intelligence tool should be higher than it is at the moment with current methods. The tool should also be capable making predictions as soon as possible in order to respond directly as a company. This is exactly what business intelligence is about. Success depends on the how fast and in what manner a company can adapt to changing market conditions when they have limited resources to gain more insights in the developments of the market and processes. They need to respond fast and accurately to new knowledge, changing technological platforms, and new products (Cunningham, Porter et al. 2006). Thus, the *two main goals* for the tool should be to assess and predict the stage of the technology life-cycle in a quick and reliable manner *and* to predict any changes in the technology life-cycle. Business intelligence focuses on making better and faster decisions with the information that is available within the company by collecting and structuring the data in such a way that it provides new information.

1.2 Research objective

1.2.1 Background information

Having the knowledge when a hallmark might occur would be of major added value for a company. It enhances many different opportunities and reduces certain risks. The goal of this research is to analyse methods and new sources which could add value to current business intelligence tools by improving the reliability of current forecasting tools. This will be done for high-tech products and their relation to search engines (and databases) in order to improve the predictive percentage of the start of the three hallmarks. However, in order to achieve this, the focus of this research will mainly be on finding a correlation between the technology life-cycle and databases. Basically, when the added value of databases is regarded as positive, it can be compared to a kind of dashboard with indicators on it. Once all these indicators turn into a “green light” it is time to prepare for the crossing of one of these hallmarks. If a company knows that the large-scale industrial production and diffusion of their brand-new product will occur, for instance within two years, they have the opportunity to prepare themselves in every possible way. However, this is only applicable for companies who have the potential to react fast and accurate based on these predictions, otherwise it will not be as beneficial as expected.

1.2.2 Different types of databases

In general there are many different databases where information is captured from many sources. Some databases collect image information, others collect numeric information, audio information, full-text information or bibliographic information[1]. *Content*-databases operate on a large range of contents such as: general (e.g. Google, Yahoo), geographical, business (e.g. Nexis), enterprise, scientific, ethnic, job, legal, medical, entertainment (e.g. YouTube or IMDB), news, people, technology, real property, etc. Besides content databases there are even more databases referring to

[1] *http://www.usg.edu/galileo/skills/unit04/primer04_02.phtml*

information-databases (e.g. blogs, multimedia, source codes, maps, email, and price) and *model*-databases (e.g. open source, social search engines)[2]. Nowadays, the internet is seen as the gateway to these databases. The internet is a global system of interconnected computer networks to serve billions of users worldwide[3]. Actually, it is seen as one of the prime databases for finding information in general. However, there are many sub-databases (as explained above) which have specific functions that the internet itself cannot offer. Specific databases that are available look deliberately at certain fields and can be regarded as a sub-database within the internet. As a result, the internet in general cannot offer the information that is looked for. Thus, specific databases are required to fulfil specific needs. In this research, we believe that there are basically three specific types of databases which can be used in this research. They are specifically closely related to the development of breakthrough technologies and therefore we will focus on these three in this thesis. Although an overview is given in chapter 2 about the differences of these databases, they are first explained below:

- *Scientific literature databases:*
  - *Scientific literature in fundamental;*
  - *Scientific literature in applied research;*
- *Patent databases;*
- *(Google's) Business press and news databases.*

Scientific databases

With these three areas there is a broad overlap of possible correlations. Scientific literature represents information that emerged with the idea of developing new theories and exploring undiscovered areas about a specific topic. Therefore, scientific literature databases can be regarded as useful sources due to the fact that literature was developed with a certain goal. Hence, it is a source without much noise (e.g. advertisements), though it includes a broad spectrum of subjects concerning the same technology.

Patent databases

Once a breakthrough technology is launched in the market it will often gain more literature, but often also more patents due to the competitive nature between firms. The number of patent applications increases to a maximum after the invention and slowly decreases thereafter (Schmoch 2007). Patents sometimes follow a second growth after about 15 years of the first growth, depending on the type of technology. A possible correlation might be found between the amount of scientific articles and the technology life-cycle pattern. However, patent databases are also clear measurement tool due to the fact that they only have information about patents and the technology itself. They do not account any other irrelevant information (e.g. advertisements), so there is hardly any noise. Patents can be a good measurement tool in some cases and often used for determining the innovativeness of the firm (Kleinknecht, Montfort et al. 2002). Patents tell something about the technological development itself and the technological know-how. Possibly, they tell something about the commercial potential of this new technology as well, because companies often prefer to patent only technologies which have the capacity to commercialise and become a success. Also, patents give an indication about the technology life-cycle, even before the diffusion or

[2] *http://en.wikipedia.org/wiki/List_of_search_engines*

[3] *http://en.wikipedia.org/wiki/Internet*

commercialisation takes off. Finally, patents are easy to be found and measured (Haubt, Kloyer et al. 2007). When a new product is developed and eventually launched in the market, the amount of patents is increasing as well. Moreover, patents can be a good indicator for assessing the first hallmarks because firms often invest in patents during the development of a new technology. The number of patent applications reflects the level of corresponding technology activities, though it mentions nothing about turnover. Patent activities should therefore be visible already before market introduction. Thus it can tell something about the technology activities for long periods, which make them extremely useful. As a result, it can say something about the maturity of the technology and how far the technology it is developed. Overall, about 40-60% of the patents are actually used for new products or processes in the market. Still, some companies have reasons not to invest in patents (e.g. secrecy reasons), but once they do invest in patents it can be a good measuring tool for assessing the progress in the technology life-cycle pattern (Schmoch 2007). The downside of using patents is the fact that not all industries are that strict in filing patents. Especially the less competitive industries file fewer patents than strong competitive industries. Also, technologies are often patent filed by subsystems, which creates a complex system to investigate the patents. The measurement of patent applications requires the complete statistical survey of all patent applications and applicants of the considered technological field. Due to the limited amount of time in this thesis, we will put less emphasis on patent databases (Haubt, Kloyer et al. 2007).

News databases

Still, there is a third area; business press and news databases, also referred as just 'news databases' in this thesis. Nowadays, search engines provide clear information on the amount of publications that were launched on this breakthrough technology and on specific parts of this breakthrough. These publications even go back to the late 19th century and look at a broad spectrum of information. They focus on business press and news articles and primarily search in news papers and magazines like the New York Times, Pittsburgh Press, Fox News and BBC News. Google offers such a search engine and even expresses their hits in a visual graph (see Figure 2).

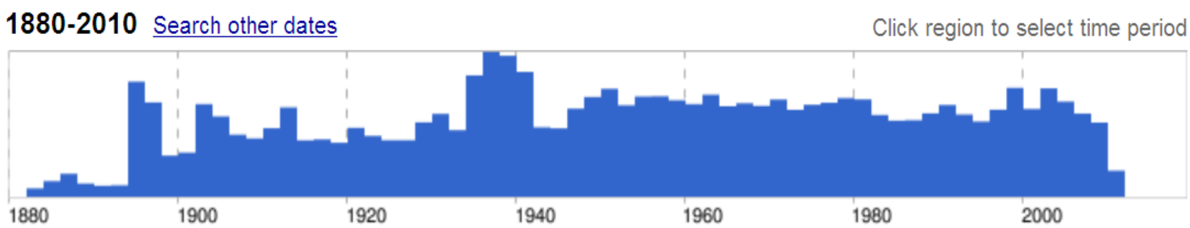

Figure 2 – The pattern of news publications of X-ray.

Thus, these databases offer great insights in future developments of breakthrough technologies. They have a much higher level of social significance compared to random blogs or web-pages. Analysis in this thesis will focus on finding a correlation between these patterns of publications and the life-cycle patterns. This can also be regarded as the main objective of this research in order to make it possible to find more about where a company is positioned in the technology life-cycle pattern.

1.2.3 Overview of the objectives

Capturing information from technological literature can be seen as a challenge due to the immense amount of useful databases. A field of "bibliometrics" emerge when information is captured which is closely connected to the content and patterning (Watts and Porter 1997; Watts, Porter et al. 1999).

Bibliometrics can be regarded as a method where counts of publications, patents, or citations are measured and interpreted in such a way to make any forecasts for the future. These counts represent the amount of R&D activity and innovations, depending on the source. Chapter 2 will further elaborate on bibliometrics.

In a nutshell, there are three main databases which can be helpful determining a possible correlation with the technology life-cycle by looking at bibliometrics. However, these databases can be classified into the technology life-cycle where they are most efficient. It is believed that scientific databases are most useful in the beginning of the technology life-cycle, because breakthrough technologies and science are well intertwined (Watts and Porter 1997; Trumbach, Payne et al. 2006). It is assumed that literature citations will develop and grow the fastest from market invention to sales take-off, because in those phases of the technology life-cycle there is much fundamental new knowledge published in literature (Haubt, Kloyer et al. 2007). Scientists share their knowledge by publishing scientific articles. Especially in the innovation phase when new developments occur on a new technology it is expected that more scientific articles are published than in other phases of the technology life-cycle. Patent databases become more important in the adaptation phase because it is believed they reflect the technology activity and new applications. Companies are slowly launching their new technologies in the market and therefore protecting their technologies by patent filing their innovations. Databases about news, in order to measure market activity, are expected to be most useful in the market stabilisation phase. However, thus far this had never been checked. Figure 3 shows which databases can be used best in the different phases of the technology life-cycle pattern. However, now we encounter our first clear objective.

First objective

As has been explained there is a strong tendency to believe that each database will be most useful in only a specific part of the technology life-cycle. However, different views attracted this topic by scientists and therefore we will give our viewpoint and analysis in this thesis based on our own analysis. This *first objective* is to investigate how certain databases develop over the technology life-cycle and at what moment are they most useful and active. In our analysis we will explain the different viewpoints from scientists and analyse whether it is indeed true that certain databases show a higher activity in only a specific part of the technology life-cycle. Basically, we investigate at what moment of the life-cycle there is most activity on news and science. The objective is to find an answer about how databases, measuring activity, develop over time. We will not focus on patent databases, since these have been investigated in previous research already (Schmoch, Grupp et al. 1991; Schmoch 2007) and because this does not fit in the time span of this thesis.

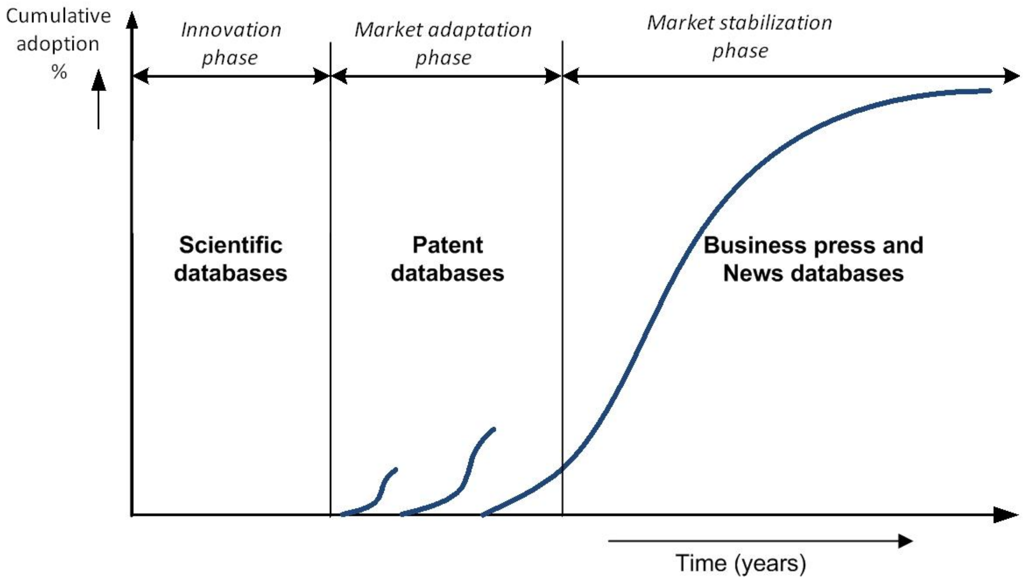

Figure 3 – The technology life-cycle and when databases are expected to show most activity.

Second objective

The *second objective* focuses on another topic, namely to investigate whether there is a correlation between databases (scientific and news) and the technology life-cycle of breakthrough technologies. If this would be the case, then it might become possible to use a new kind of analysis in business intelligence tools. Basically, there are two strategies that can be executed in this second objective for finding a first correlation between the technology life-cycle and the pattern of publications for different databases.

Strategy 1

By looking at different databases it might become possible to reflect to the phases of the technology life-cycle. This strategy would involve scientific- and news databases. Subsequently, these sources would be used in the innovation phase and the market stabilisation phase. Figure 4 shows how this strategy in an overview. Thus, this strategy only looks for a correlation between the patterns in a specific part of the technology life-cycle like many forecasting tools do.

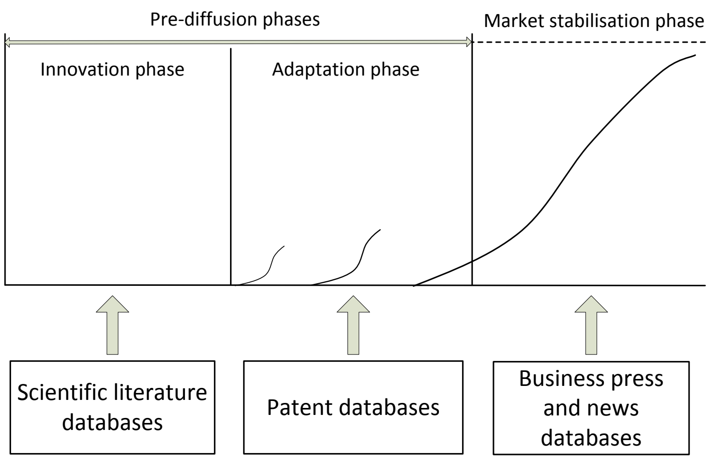

Figure 4 – Strategy 1: different databases are used in the three phases.

Strategy 2

By focussing on one database at a time over the complete technology life-cycle and finding a possible correlation over all the three phases, it might become possible to reflect where the highest degree of reliable correlations is located. Figure 5 shows this strategy in an overview.

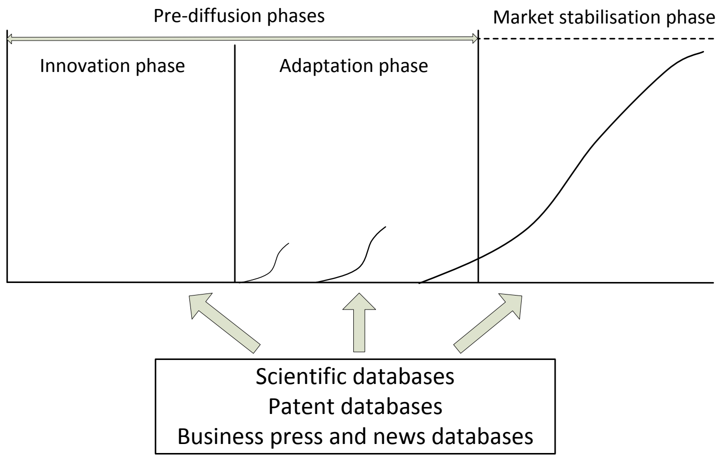

Figure 5 – Strategy 2: one database is used in all the three phases.

The *second strategy* seems to be very promising due to several reasons; therefore this research will focus on this strategy. *First*, news and business press articles are a strong indicator due to the fact that it relates to the interests of the consumers instead to scientists and companies. More news articles make consumers aware of a new technology. Scientific literature and patent databases are not regarded as a tool to inform consumers. The large-scale diffusion of the technology is partly depending on the amount of consumers. Therefore, news articles might relate to the popularity of a technology among future consumers. *Second*, scientific databases and patent databases are already been partly discovered, however, scientific databases still deserve some attention. News articles databases on the other hand are relatively new and undiscovered. Although using the news database might fit best in the market stabilisation phase for finding a correlation, other phases like the adaptation might also be useful. New technologies are often implemented in niche markets first. It represents the first application of a new technology available for the consumers. Therefore, it will grab the attention of the consumer and more news articles become available. *Last*, news databases are hardly used in science so far. Therefore it is interesting to see in what way they create added value in the three phases of the technology life-cycle. With this strategy it becomes possible to find a possible fit and the added value in this explorative research.

1.2.4 Conclusion

This explorative research mainly focuses on scientific- and business press and news applications and will put less emphasize on patent databases. So the main focus lies on finding a correlation between the technology life-cycle pattern and the patterns of news and scientific publications. However, we will also elaborate on the moment these two databases show most activity over the technology life-cycle. This research is one of the first attempts that focuses on this topic and includes this business and news application. The choice on this particular area was deliberately taken due to the undiscovered and possibly potential outcomes. Furthermore, it includes many aspects which are not involved in the areas of patents and scientific literature.

Whatever the result will be of this research it can be regarded as useful for both managerial and scientific aspects. When a correlation is found between both patterns it will contribute for further research on different scientific levels. It enhances the perception there is about what can be used for indicators for determining large-scale production and diffusion. Also, due to the fact that this research is broadly orientated by involving different cases of breakthrough technologies it will be highly useful, also for companies. Companies can fully benefit on having the knowledge of when a hallmark will occur.

1.3 Research questions

The proposed research intends to answer the following main research question:

- *How to predict the development of breakthrough technologies (i.e. the technology life-cycle pattern with three hallmarks and phases) with the help of electronic databases?*

In order to find answers for this research question, a set of independent sub-questions are constructed as well. These can be divided into content- and method questions.

- *How do scientific and news activity develop over the technology life-cycle?*
    - *At what moment of the technology life-cycle are news publications most active?*
    - *At what moment of the technology life-cycle are scientific publications most active?*

- *How can news databases and scientific databases add value to a business intelligence tool?*
    - *Is there a correlation between scientific literature patterns and the life-cycle patterns (using the name of the breakthrough technology only)?*
    - *Is there a correlation between publication patterns and life-cycle patterns (using the name of the breakthrough technology only)?*
    - *What is the added value of (news and scientific) databases for forecasting the development of a breakthrough technology?*

Additional remark 1

An additional note about the term correlation is justified. In this thesis we will refer many times to the term correlation. Although we use this term regularly it might cause discussions whether it is justified to do so. The correlations between the patterns (chapter 5) will be of subjective nature and not tested statistically. As a result, the term correlation might be too ambitious. Although we will use the term correlation in this thesis, we want to make clear for once that this term can be better explained as a fit or match instead of a clear-cut correlation.

Hypotheses

In order to find an answer to these research questions, we have developed some hypotheses. The hypotheses are divided into two areas; phases and hallmarks. First the focus lies on the phases of the technology life-cycle, subsequently the focus shifts towards the hallmarks.

Additional remark 2

Another important note concerns whether the term hypothesis is justified. The first two hypothesis will be tested statistically, the third and fourth will not be tested. However, for the convenience we will keep using this term.

Part I - Phases

1. *Scientific databases show most activity and collect most publications in the beginning of the development of breakthrough technologies (i.e. around the innovation phase of the technology life-cycle).*
2. *News databases show most activity and collect most publications at the end of the development of breakthrough technologies (i.e. around the market stabilisation phase of the technology life-cycle).*

Part II - Hallmarks

3. *There is a correlation (i.e. match/fit) between the pattern (i.e. hallmarks) of technology life-cycles of breakthrough technologies and the pattern of publications from scientific databases.*
4. *There is a correlation (i.e. match/fit) between the pattern (i.e. hallmarks) of technology life-cycles of breakthrough technologies and the pattern of publications from news databases.*

1.4 Research methodology

This explorative study will be executed in several steps. First a literature study will be conducted in order to gain more insights in breakthrough innovations, technology life-cycles, electronic databases, forecasting methods, business intelligence tools, etc. Also, more insights will be gained for future research like indicators and factors which could be used in search combinations. A framework will be proposed of useful indicators which are most relevant in the different phases of the technology life-cycle.

After diving into the theoretical background of this topic, we will start with our first analysis. Different viewpoint from different scientists will be explained and subsequently an analysis will be done which will be executed on scientific- and news databases. Data will be collected for 14 cases from two industries, materials and pharmaceutical. The data will consist of the amount of publications (both news and scientific) over the length of the technology life-cycle and data on the hallmarks of the life-cycles of the different breakthrough technologies. The analysis will focus on whether it is true if certain databases show most activity in only a specific part of the life-cycle or whether all these databases show equal activity in a parallel form.

Thereafter, different methods will be proposed on how to fulfil a complete analysis for finding a correlation between the pattern of the technology life-cycle and the patterns of publications. Although only one method will be used, other methods might gain more attention in future research and will therefore be included as well. However, the most appropriate method will be used for a second analysis for finding this correlation which will be fulfilled subsequently with the same data that was collected prior. As a result, it becomes possible to say something about the databases and their added value in business intelligence tools. Figure 6 shows how the research methodology conceptually looks like in different steps.

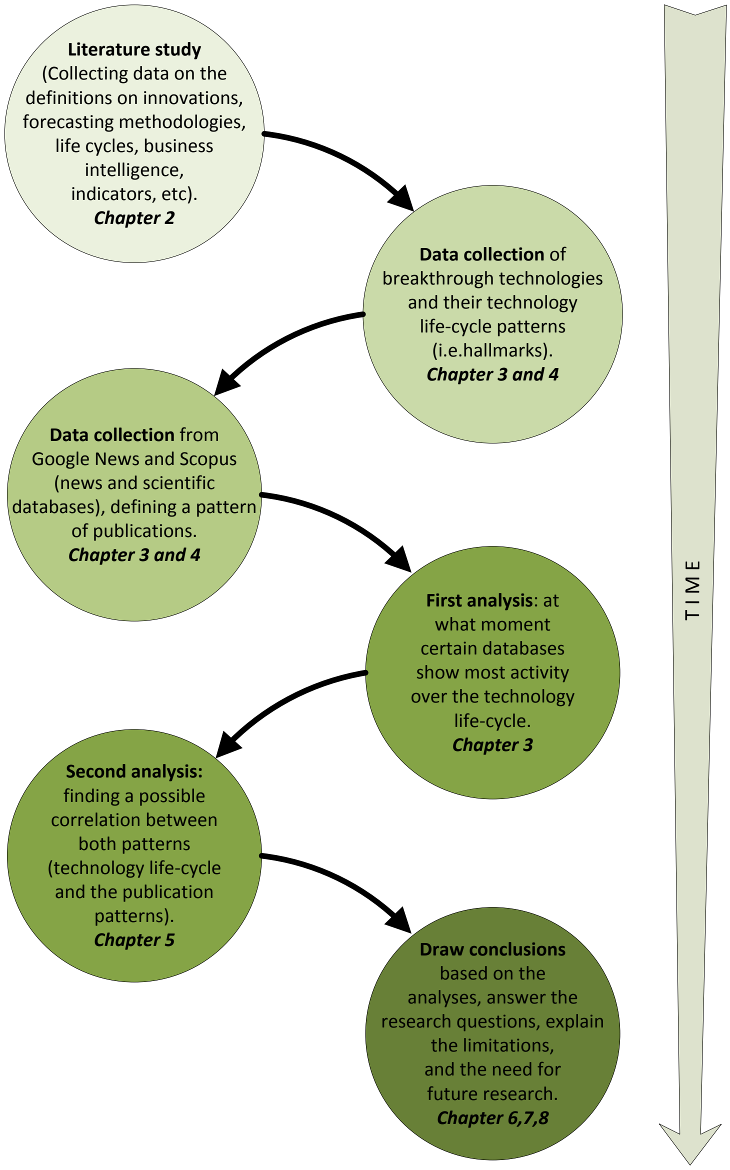

Figure 6 – The theoretical framework of the research divided in several steps.

1.5 Limitations

There are a few possible limitations and difficulties that might occur during this research. It is important that these possible drawbacks are discussed upfront in order to reassure the best results possible. First of all, there is a possibility that no correlation is found between both patterns. There are many reasons to believe that this might occur, like too much noise in the publications which cannot be filtered out.

A second possible limitation is the database which includes many breakthrough technologies. Due to the fact that this research focuses on breakthrough innovations it is of paramount importance to filter out the incremental innovations. As a result the sample size might diminish.

Also, choosing the right industries is important. Telecommunications for instance often show a breakthrough technology in daily products. Bluetooth, for instance, is a breakthrough technology which is implemented in many different products (e.g. mobile phones, computers, cars, etc). When data is collected on Bluetooth, chances are high that the amount of hits is extraordinary high due to the many products that this application is implemented in. The occurrence of this phenomenon should be taken into account.

Thirdly, there are many ways to investigate a possible correlation between the patterns. However, due to limitations in the data it might become quite difficult to assess in a quantitative manner. The data might have too much noise. Although the patterns of publications give a clear overview of the amount of publications in a particular period, it does not show the exact amount of publications. Also, some patterns show many different peaks, instead two clear peaks that can be linked to the hallmarks. This phenomenon might even occur when the search terms were made very specific. In these cases it will be very hard to generalize based on these cases. The most appropriate method for assessing a correlation will be determined in a later stage.

Fourth, the strategy focuses on two industries; materials and pharmaceutical. As a result, there is a possibility that the industry with the best result is not tested in this research. However, the two industries that are chosen will be tested quite elaborately.

Last, this research is limited, because it includes only the databases of science and news. Patent databases are not included in this research, because this topic has been studied in earlier research; however, it can still be of major added value combined with scientific and news databases. Due to the fact that this area of business press and news publications is expected to be the most fluctuant one, it possibly also brings less reliability with it.

1.6 Report outline

This thesis consists of 8 chapters and will have a similar order as the research methodology as has been explained in the previous paragraph. In chapter 2 we will review all relevant literature concerning this topic; technological life-cycles, definitions on radical and breakthrough technologies, forecast methodologies, different definitions on intelligence, business intelligence tools, and factors which could be used in these types of research.

In chapter 3, the added value of different databases for technological forecasting will be analysed. Different viewpoints based on literature will be explained. Next, an analysis will be done based on our own data collection on scientific and news databases. Also, the first two hypotheses will be answered.

In chapter 4 we will propose different methodologies for the second analysis in this thesis. Both quantitative and qualitative methodologies will be explained and we will elaborate on the data collection and the difficulties we encountered. Parts of this chapter will be extremely useful for future research as well.

Chapter 5 contains the full analysis on finding a correlation between the pattern of the technology life-cycle and the pattern of publications. Also, two extreme exhibits of cases will be explained in detail. The final two hypotheses will be assessed in this chapter as well. Chapter 6 further explains how our results can be further implemented into a future business tool and how we believe it can add value to these forecasting tools.

The final two chapters, 7 and 8, focus only on limitations of this research, on future research and the final conclusions and discussions.

2. LITERATURE STUDY

This literature study can be divided into two main parts. The first part is about the background of this topic. It further explains the technology life-cycles and the applications, but also what kinds of forecasting methods are available, which are used nowadays and why these are not sufficient in many cases. Also, a closer look at business intelligence is given.

The second part focuses on factors that can be useful in the business intelligence tools. The factors and associate search terms will be assessed on their reliability and whether they give useful results when they are used in Google's business press and news database, called Google News.

2.1 The development and diffusion of new technologies

2.1.1 Defining breakthrough and radical innovations

There are different definitions for defining a new technology that is far more than just an incremental improvement of an existing product. Most of them will be discussed in order to provide a clear context and understanding. Technological innovation can be explained by the process of introducing something new or a new technological idea or product. Basically, technological innovation can roughly be divided into radical innovations and incremental innovations (Veryzer 1998a; Mohr, Sengupta et al. 2005). Incremental innovations can be seen as innovations which focus on exploitation of existing technologies (Veryzer 1998a; Mohr, Sengupta et al. 2005). Radical innovations are often referred as "really new" (Song and Montoya-Weiss 1998; Garcia and Calantone 2002), "discontinuous" (Lynn, Morone et al. 1996; Veryzer 1998a; Brentani 2001), or "disruptive" innovations (Christensen 1997). These innovations focus on the exploration of new technologies or new applications of technologies (Veryzer 1998a; Mohr, Sengupta et al. 2005).

Thus, many researchers define these innovations differently; as a radical high-tech product, a breakthrough technology or a disruptive innovation. Although they all seem to be the same initially, there are some differences to be found. So how are these definitions defined to distinguish them from one another? Overall, the level of innovativeness seem be make the difference. First, a product which emerges from a radical innovation can be defined as follows: "a radical new product is a product built upon highly advance technological developments that is perceived as offering substantially enhanced benefits to customers". These products transform existing markets, create entirely new markets, or change the pattern of customer behaviour. Also, a radical innovation shows a significant advancement in technological capability (over the "old" solutions), because it is built upon new technological capabilities. It is perceived as new by the firm and/or customer. As a result, the price/performance ratio and product benefits are greatly improved which causes an enormous change in the consumption pattern of the customers. Therefore, these radical products can revolutionize existing markets or industries or create new ones. Due to the fact that radical innovations emerged from R&D inventions, they are typically a technology push (Tushman and Anderson 1986; McIntyre 1998; Veryzer 1998b; Mohr, Sengupta et al. 2005). Not all radical innovations are necessarily disruptive, and vice versa. Some products can even be both radical and disruptive (e.g. cellular phones) (Slater and Mohr 2006). Others state that firms which are highly market orientated cannot innovate disruptively, because then they would lose their industry leadership position by listening too much to their customers (Christensen and Bower 1996).

Discontinuous innovations are innovations in which the commercial impact of the application of existing technological know-how to a new domain of application really matters (Levintal 1998).

Incremental innovations are innovations based on existing technologies, products or processes and improved. The focus of these types of innovations lies on low-cost advantages and fulfilling the needs of the demand side of the market. Therefore, they can be considered as a customer pull (Veryzer 1998a; Mohr, Sengupta et al. 2005).

Breakthrough technologies on the other hand are defined a bit different. They are considered "new-to-the-world or radical (improved) products" which can be considered as a breakthrough in the market. Radical innovations do not automatically become a success in the market. Breakthrough technologies can be seen as a radical innovation which also diffuses successfully into the market and represent either an advance in technology that is so significant that attainable price/performance ratios are altered dramatically or enable an entirely new kind of application that changes the behaviour pattern of end users (Ortt, Langley et al. 2007). These products are new to the market and new for the firms (Booz, Allen et al. 1982; Gaynor 1993). These types of innovations are quite important, or can even be considered as crucial, for the future growth and long term survival of high technology companies. Especially because they provide the foundation on which further future generations of products or services are developed (Kleinschmidt and Cooper 1991; Lynn, Morone et al. 1996; Veryzer 1998b). From now on, we will use the term *breakthrough technology* in this thesis. Table 1 below shows an overview of how different authors interpret the definition about radical innovations.

Author(s)	Term	Definition of the term
(Booz, Allen et al. 1982)	New-to-the-world products	"Products that are new to the firm and that create an entirely new market"
(Tushman and Anderson 1986)	Product discontinuities	"Fundamentally different product forms that command a decisive cost, performance, or quality advantage over prior product forms"
(Moore 1991)	Discontinuous products	"Products that require customers to change their current mode of behaviour or modify other products and services they rely on"
(Green, Gavin et al. 1995)	Radical innovation	"Radical innovation can be measured according to four dimensions: technological uncertainty, technical inexperience, business inexperience, and technology cost"
(Christensen 1997)	Disruptive innovation	"An innovation is disruptive if they create discontinuities in the market"
(O'Connor 1998)	Discontinuous innovation	"The creation of a new (new for both the firm and the market) line of business (product or process) that either has unprecedented performance features or familiar features that offer the potential for fivefold to tenfold improvements in performance or cost"
(Veryzer 1998a)	Radical or discontinuous new products	A radical or discontinuous new product can be "technologically discontinuous" (new technological capabilities), "commercially discontinuous" (new product capabilities), and

		"technologically and commercially discontinuous".
(Veryzer 1998b)	Really new or discontinuous new products	"Product innovation may be viewed as lying along dimensions reflecting changes in three dimensions: product benefits, technological capabilities, and consumption or usage patterns".
(Chandy and Tellis 2000)	Radical product innovation	Radical product innovation is "a new product that incorporates a substantially different core technology and provides substantially higher customer benefits relative to previous product in the industry"
(Moerloose 2000)	Breakthrough products	"Products that are technologically new (e.g. include new raw material, a new production process, technical change, input replacement), commercially new (i.e. create originality by commercialization), functionally new (i.e. provide new benefits/new solutions for the customers), and induce new customer groups"
(Leifer, O'Connor et al. 2001)	Radical innovation	" A product, process, or service with either unprecedented performance features or familiar features that offer significant improvements in performance or cost that transform existing markets or create new ones"
(Ortt, Langley et al. 2007)	Breakthrough technology	"Breakthrough technologies represent either an advance in technology that is so significant that attainable price/performance ratios are altered dramatically or enable an entirely new kind of application that changes the behaviour pattern of end users"

Table 1 – Overview of different definitions of innovations based on literature.

To conclude, high-tech radical technologies are of paramount importance for the market and companies. However, not all new innovations become a success. Many different criteria have an influence whether the technology fails or succeeds. There are many understandings of the characteristics of high-tech products, but there is no easy way to explain this by one clear definition (Mohr, Sengupta et al. 2005). However, from now on we will define these types of technologies as a breakthrough technology. Despite the potential of many breakthrough technologies, there are associated costs and risks. The technically more complex products need to be developed faster and in and highly dynamic environment. Resources need to be allocated to research and development projects, and the new products need to be introduced in the market (with strong competition) (Maier 1998). As a result, many innovations will not survive. A further emphasis will elaborate on this topic after the explanation of the technology life-cycle. Moreover, often there is some resistance from inside the firm because there are uncertainties about the future success and the risk of no market for the product. Thus, they find it difficult to support managerial decisions. If there would be more certainty about the future market, it would enhance the decision making process of managers (Christensen 1997).

2.1.2 The technology life-cycle

In the past there have been many models developed which represent the evolution of technologies and high-tech products. The main difference they contain is that some patterns focus on the life-cycle of a technology (Abernathy and Utterback 1978), others focus on the life-cycle of (high-tech) products (Hauser, Tellis et al. 2006). In both cases the patterns are based on the S-shape pattern, which represents the cumulative take-off of a technology. Often, this pattern takes off significantly after the introduction of several niche-applications who will succeed or fail. The technology life-cycle can be considered as a long and uncertain journey.

One famous model of product diffusion is the 'Bass' model and can be considered as one of the most influential models in marketing literature history. It was introduced in 1969 (Bass 1969). The S-shape curve represents the cumulative distribution of adoption over time and the related growth. The model is relatively easy; however, it offers great support for forecasting methods (especially prior to product launches). It is particularly helpful when there are no sales data available at that moment of the life-cycle (Bass 2004; Chandrasekaran and Tellis 2007). Although the model is quite accurate with its predicting capability, it contains two start-up problems, i.e. two key turning points in early sales (sales take-off and slowdown). As a result, it needs to be remarked that the S-shape model cannot fully describe the retrospective purpose in order to describe the evolution of new product life-cycle, especially for breakthrough technology (Ortt and Schoormans 2004; Ortt, Shah et al. 2008; Golder, Shacham et al. 2009) .

In order to complete the model, an improved pattern has emerged over time. Although some technologies (e.g. telephone and television technology) have shown a prefect S-shaped pattern, there are also technologies which do not follow this prefect pattern. Research on the chronological dates from their first introduction in the market, the pattern transforms in a more erratic pattern (Ortt and Schoormans 2004; Ortt, Shah et al. 2008). As a result, the new technology life-cycle (TLC) gives a far more detailed overview of the diffusion over time. Figure 7 shows how this looks like. As also has been explained in the introduction of this thesis, this technology life-cycle pattern is famous for its acceptance in technology developments. Many firms (and researchers) use this cycle as a method to determine where they are positioned in the development of a product and where their competitors are positioned.

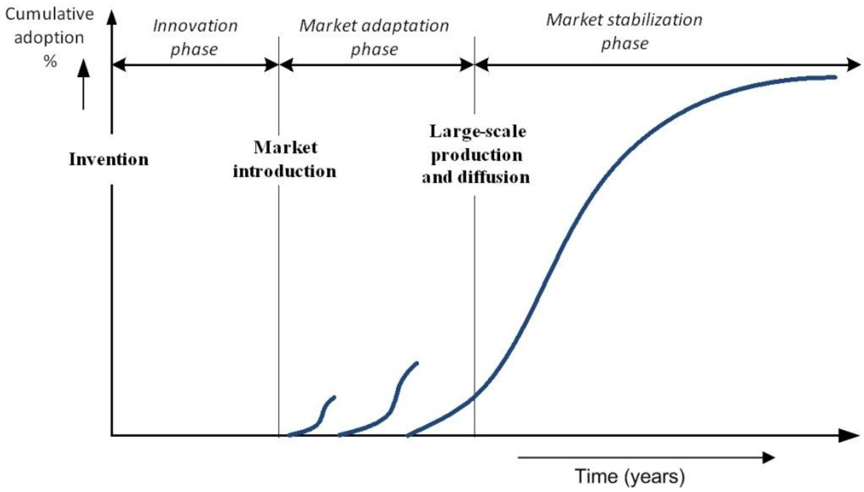

Figure 7 – The technology life-cycle model of breakthrough technologies (Ortt *et al.*, 2004, 2007).

Phases and hallmarks of the technology life-cycle

Hallmarks

The phases are separated by hallmarks; the moment of *invention*, the moment of initial *market introduction* and the moment of *large-scale production and diffusion*. The occurrences of these hallmarks are dependent of many variables. These variables have an influence on the moment of when the hallmark is crossed. So, they all have an effect of the duration of the pre-diffusion phase. The availability of production- and marketing factors, complementary products, network of actors, regulations, select a dominant design over a standard design and this is the moment when the large-scale diffusion will slowly take off (Ortt, Tabatabaie et al. 2009; Ortt 2010).

The hallmarks can be seen as clear moments in the life-cycle of a technology between the different phases. However, many researchers have slightly different viewpoints of what these hallmarks exactly entail. Therefore, we will give an overview of the three hallmarks first as they will be used in this research (Ortt and Schoormans 2004; Ortt 2010):

Invention

The invention of a new high-tech product category can be assessed by the first time that the technical principle (the functionality) of this category is demonstrated and mastered.

Market introduction

The introduction date is defined to be the date at which the product is available for sales or can be transferred to users. In some cases, products are not sold, for example, if a government institute develops a new weapon that is used by military forces.

Large-scale production and diffusion[4]

This hallmark is defined using three elements:

- A standard product that can be reproduced multiple times (or standard product modules that can be combined in many different ways but are based on the same standard platform);
- A (large-scale) production unit with dedicated production lines (industrial production of a standard product; and
- Diffusion of the product.

Thus, the hallmarks can be considered as a milestone, since companies see it as the start of a new phase. Often they need to prepare themselves in many different ways in order to start with the new phase. The technology will be tested and improved over and over again during the innovation phase. Especially in the pharmaceutical world, medicines need to be tested extensively before they can be used in a responsible manner. Only when it is considered mature for market introduction, the technology will be used in a product and launched in the market as a first attempt to commercialise the product. Therefore, the second hallmark "market introduction" can be considered as a clear and

[4] Due to the fact that the "milestone" *sales take-off* is not a clear indicator, it will not be used in this research. Sales take-off means the time in which the products in a category starts to grow in the mainstream market, however it is hard to define, therefore the *large-scale production* will be regarded as the start of the third hallmark.

strict moment of the technology life-cycle, since companies have often an official launch date for market introduction. The third hallmark, industrial production and large-scale diffusion, is on the other hand crossed less clear and strict. The hallmark is dependent on many variables and will be crossed with less attention, partly due to the slow growth in popularity among (future) consumers. When the production capacity increases and more products are diffused into the market, the hallmark will be crossed.

Phases

Companies often introduce a new technology into a niche application so that people can slowly get acquainted with this new technology. However, not all niche markets become a success and it might take a while before more people become aware of the benefits of this technology. Still, sometimes a new technology is pushed into the market by companies (*technology push*) instead that the market demands new technologies and solutions (*market pull*). For a technology push situation, consumers often need to adapt to this technology. The last phase of the life-cycle will start once the adaptation phase is becoming more stable and sometimes when the battle of what will become the new standard is determined if there are similar technologies that were introduced in the adaptation phase. As a result, these phases emerge over time and can vary in length of time. The definition and characteristics of the phases in development and diffusion will be explained below according to the model of Ortt (Ortt and Schoormans 2004; Ortt, Tabatabaie et al. 2009):

Innovation phase

> The innovation phase is the period starting from the invention date or "*the first time that the technical principle of this category is demonstrated and mastered*" to the first market introduction or "*the date at which the product (category) is available for sales or can be transferred to users".*

Market Adaptation phase

> The market adaptation phase is *the period starting from the first market introduction to industrial production and sales takeoff* which is characterized by three elements, the realization of large-scale production of product categories, the emergence of standard products or dominant design, and large-scale diffusion or the start of sales growth. In the market adaptation phase the new high-tech product category is often introduced one or more times by either pioneers or early followers.

Market Stabilisation phase

> The market stabilisation phase is the period after the *sales takeoff*. This phase entails the fastest growth of diffusion (rapid sales growth), maturity, and decline.

From empirical research it emerged that the phases have different time-lengths. In this study (Ortt, Tabatabaie et al. 2009), 50 different product categories were investigated, divided among five industries. As a result, the average duration of the pre-diffusion phase is about 17 years for breakthrough technologies. As we know, the pre-diffusion phase can be divided into two phases; the innovation and adaptation phase. The average innovation phase is about 10 years, while the average adaptation phase is about 7 years before the technology diffused widely in the mainstream market. Other researchers, (Golder and Tellis 1997), did a similar study on 29 radical innovation products. The results of this research were approximately similar. The average duration of the innovation phase is about 15 years, while the adaptation phase is about 8 years. If both studies are compared, we found a small difference. It has to be remarked that they used slightly different terms which distinguish the phases.

Different scenarios of development of breakthrough innovations

Overall, there are three main scenarios of these diffusion patterns. First, an adoption of a new product that immediately diffused after market introduction. The second scenario shows a long adaptation phase before the actual large-scale diffusion took off. Finally, there is a scenario that shows a very large innovation phase before the adaptation phase finally occurs. Figure 8 shows how these scenarios look like. It is interesting to see that different industries also clearly show phases with different lengths. For instance, a breakthrough technology in the pharmaceutical industry often has a very long innovation phase due to the fact that a lot of time is required for testing and improving a medicine. Electronic equipment, on the other hand, often has a significantly longer adaptation phase than technologies that were developed for aerospace and defence purposes which show a short adaptation phase (Ortt, Tabatabaie et al. 2009). These results are quite important for companies to know in order to determine the strategy that is required. There are different strategies to execute; a mass market strategy, niche market strategy and wait-and-see strategy (Ortt, Shah et al. 2007).

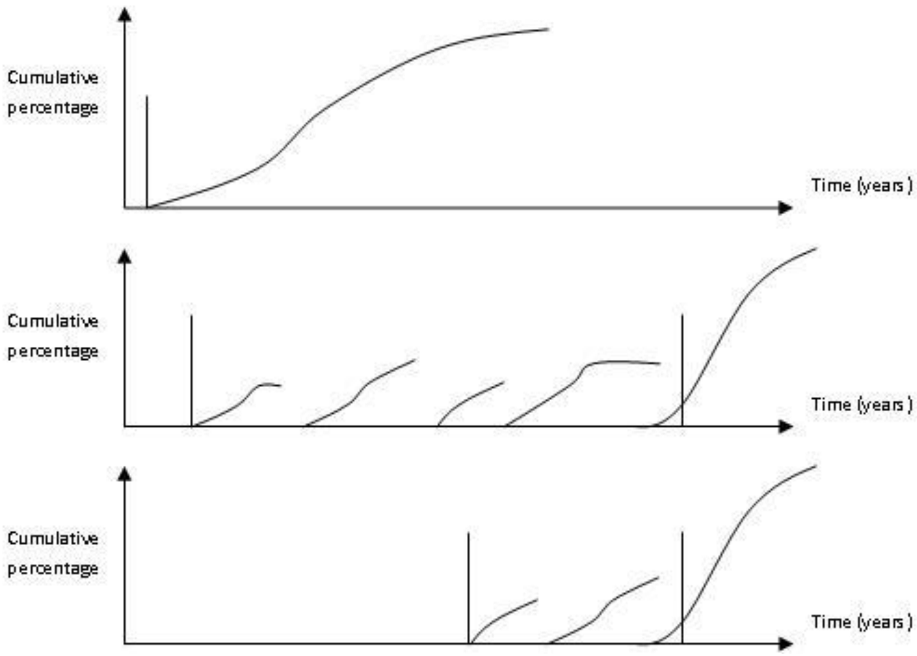

Figure 8 – Three different scenarios that might occur during the diffusion of products

Often companies do not know where they are positioned in the technology life-cycle pattern. They are also unaware when the two hallmarks of market introduction and large-scale production and diffusion will occur. As a result they have difficulties in determining their best strategy. They face difficult questions whether they should prepare for mass-production or wait until the pioneers succeed or fail with their technology first. The first-mover opportunity might be very beneficial for pioneers due to possible learning effects, patents, and buyer switching costs (Lieberman and Montgomery 1988). However, many pioneers fail in their attempt to become a mass producer due to the many risks and costs. Often they end up in fulfilling a smaller niche market. The smart followers can adjust their strategy and sometimes become a mass producer, once the pioneers failed with their strategy (Golder, Shacham et al. 2009).

2.1.3 The adoption and diffusion of new technologies

The diffusion of a breakthrough technology is accepted by consumers in different ways. Sometimes the technology is accepted immediately, sometimes the consumers need to adapt to this new technology. The technology adoption life-cycle (TALC) is a model for classifying the market and its reaction to high-tech products. Customers can be categorized into five main classifications: innovators, early adopters, early majority, late majority, and laggards (Moore 1991; Mohr, Sengupta et al. 2005).

These different categories are of paramount importance for companies in order to understand the differences between customers groups in terms of the probability of (early) adoption by customers. Moreover, it explains the discontinuity between the early market of breakthrough technologies and the lack of enthusiasm of the mainstream market (Mohr, Sengupta et al. 2005). According to Moore (1991), there is a '*chasm*' between the early adopters and the early majority. A chasm is a gap in the technology adoption life-cycle pattern between the early adopters and the early majority. The early adopters are people who can be regarded as trendsetters who want to use the latest gadgets. It takes a while before more people are convinced about this new technology and are also willing to use it. The chasm should be crossed in order to start with mass-production. The early and late majority will then follow the early adopters by using this technology as well. In order to achieve this result, a company should invest in many different aspects; like developing complementary products, gaining credibility, and further improve the technology. It is hard to predict when a chasm is crossed due to many different variables. However, crossing the chasm can suddenly occur and often a company is not yet prepared for this moment. As a result, a competitor might become a market leader due to the slow response of the developing company. Thus, knowing the exact moment of the hallmark would be of tremendous value.

This revised TALC model explains that this chasm should be crossed in order to reach the mainstream market. In order to cross the chasm, a firm needs to understand the dynamics of the market adaptation phase, because this phase includes the early market (innovators and early adopters) and the mainstream market (early majority, late majority, and laggards) (Moore 1991). Below in Table 2 an overview is given for each customer category according to Moore.

High-Tech Adopters	Profile	Needs/Wants
Innovators: Technology enthusiasts	The people who have the interest to learn about a new technology and value the technology for its own sake. They are the first customers who use this new technology just to see if it works.	These people need: 'truth without trick', access to the most technically knowledgeable person; however "price" is not an issue.
Early adopters: the visionaries	The people who have 'executive ranks', 'highly motivated', 'driven by big dream'. They can foresee an 'order-of-magnitude' improvement and willing to take high risks. The least price-sensitive adopters often can	These people want to create a leap, move through fundamental breakthroughs, specific customized solution to fulfil their dreams. They urge project oriented relationships, speed, and time pressure deadlines.

	provide 'up-front money' but are 'very hard to please'.	
Early Majority: the pragmatists	The people who make the bulk of the market volume for any technological product. They believe in evolutionary, not revolutionary products/innovations.	They want economically proven technological products for productivity enhancement.
Late Majority: the conservatives	The majority of people after the pragmatists who are pessimistic about their ability to gain value from any technology investment. They rely on trusted adviser to help them justify the technology.	They want cheapest price and bulletproof solutions. They want to buy new technological products just to stay even with competition.
Laggards: sceptics	The latest people who buy technological products. They tend not to believe the benefits of technology unless they cannot avoid it anymore.	They only want to maintain the status quo. They want the best alternative with solid cost justification.

Table 2 – Overview of the different customer categories of Moore's model (Moore, 1991).

So far, both the technology life-cycle and the technology adaptation life-cycle have been explained. Both patterns can be combined into one overview. This is done in Figure 9. The patterns clearly show that the chasm is located in the adaptation phase. Once the chasm is crossed, the large-scale production and diffusion will start. After combining both models it becomes possible to see why the life-cycles simply do not just start taking off, but rather have some difficulties in the start-up. It also explains why it takes some time before the market stabilisation phase starts.

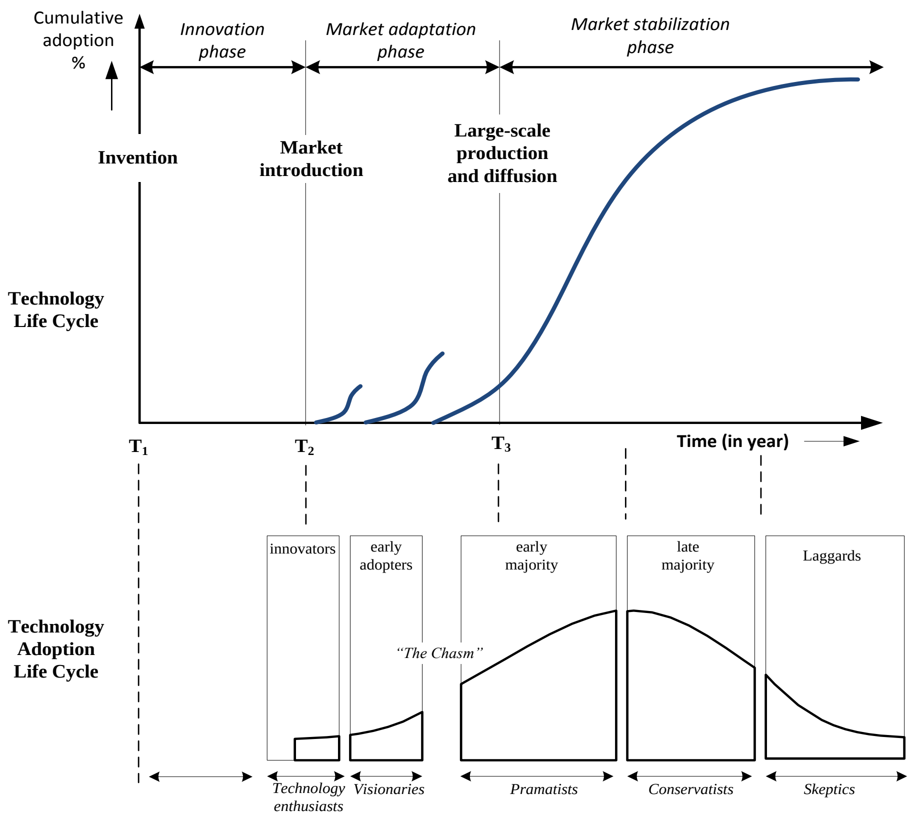

Figure 9 – The technology life-cycle model and the technology adoption life-cycle combined.

Causes of failure

It often happen that breakthrough technologies do not survive at all. This can occur in the invention phase or in the adaptation phase. The chasm often causes the most difficulties for survival. The chasm is caused due to the discontinuity effects of a breakthrough technology. These effects can be caused by:

- Newness to firms and the unknown application of the technology (Veryzer 1998a; Veryzer 1998b);
- Newness to customers (compared to the traditional products) (e.g. technological complexity, product capability, or benefits) (Windrum and Birchenhall 1998; Veryzer 1998a; Veryzer 1998b);
- Limited advantage of price/performance ratio of the breakthrough technology over to the alternative technologies (Windrum and Birchenhall 1998; Agarwal and Bayus 2002);
- Lack of firm's competences, lack of willingness, and limited ability to invest (Chandy and Tellis 2000);
- The dominant design or standard has not emerged yet (Moore 1991; Mohr, Sengupta et al. 2005);
- No final end-to-end solution, no complementary products, service and support (Moore 1991; Mohr, Sengupta et al. 2005);
- Network effects (i.e. the application of a technology requires an extensive system of infrastructure, complementary products or services that are not yet available) (McIntyre 1998). Lock-in (i.e. well established products can hinder the demand for the new products) (Church and Gandal 1993; Lee and O'Connor 2003; Stremersch and Tellis 2007).

As has been explained earlier, the technology life-cycle consists of three phases. Most problems concerning the development occur in the innovation- and adaptation phase; however, innovation itself occurs in all phases of the life-cycle. Still, failure of innovation projects occurs quite regularly. About 60% of innovation research projects fail due to technical failure. From the remaining 40% of technically feasible innovation research projects, only 22% of those projects show a positive feasibility to be incorporated into new products. The rest is stopped due to insufficient economic success potential. From the 22% projects which have the potential to become a success, only 40% ultimately survives and will be successfully introduced in the market (start of the adaptation phase). This indicates that, on average, only 8.8% of all innovation projects have a chance to be successfully diffused into the market (Maier 1998). In other research, the failure rate is around 30% to 40% (Crawford 1977; Crawford 1979; Crawford 1987; Karakaya and Kobu 1994). At firm level, the failure rates of pioneering companies report a lifetime failure rate of 47% for market pioneers. This rate is found by using historical data of the known successful products (Tellis and Golder 1996). However, in case of extremely breakthrough products, it emerged in another research that the failure rate exceeded 77% for market pioneers, while for incremental new products the failure rate would be 39% (Min, Kalwani et al. 2006). This failure rate of 77% for pioneers which try to commercialize radical innovation to the mainstream market is quite similar compared to another research, where is was 76% (Golder, Shacham et al. 2009).

Conclusion

To conclude, the highest failure rates are found by pioneers due to the newness of the products and the acceptance of the market. The unfamiliarity to both innovating firms and the market leads to uncertainties (Kleinschmidt and Cooper 1991; Karakaya and Kobu 1994). Firms need to cope with market uncertainties and have to deal with many risks. Although often failure of new technologies occurs in the first phase of the life-cycle, in the second and third phase (adaptation phase) is where most costs (more than 50%) are made by the firms (Mansfield, Rapoport et al. 1981; Maier 1998). Firms need to invest in logistics, production, marketing, etc. If a new technology fails in these phases it would be worse than when a new technology fails in the first phase. A company would benefit tremendously when it would have more certainty in that second phase about the diffusion of a new technology, especially by knowing when the chasm is crossed. Forecasts and predictions might help managers make wiser decision. The next paragraph will emphasize more on this particular topic.

2.2 Forecasting with the help of intelligence

Breakthrough technologies and their development and commercialisation involve major challenges for firms (Meldrum and Millman 1991; Meldrum 1995; Mohr, Sengupta et al. 2005). During the development of the technology there is the potential risk of technological infeasibility and during the commercialization there is the potential risk of market uncertainty (Veryzer 1998b). Managers would like to reduce the risks, manage the uncertainties and improve their chances in the market, however they often take decisions largely based on intuition (Porter 2007). Unfortunately, it is very difficult and expensive to identify market opportunities and implement the new technology into products that fulfil customer needs. Moreover, due to the long development of breakthrough technologies (about 10 to 20 years), the needs of customers will constantly change. For firms is it important to aim at the right customers needs on the right moment (Christensen 1997; Veryzer 1998a; Ortt, Shah et al. 2007; Golder, Shacham et al. 2009; Ortt, Tabatabaie et al. 2009). The main challenge is how to predict the market demand for an innovation when little or no data is available (Meade and Islam 2006). This paragraph will emphasize on the possibilities that are offered by focusing on technological forecasting. Also, it will explain how forecasting can be implemented into a business intelligence tool, which offers managers support during their daily decisions.

Generally, there are two main types of forecasts areas; *supply* forecasts and *demand* forecasts. Supply forecast focuses on the supply side of the innovation, which means that it tries to make predictions which have a relation to the developments of the innovation itself (often within the company). The demand side focuses more on the external environment, thus it tries to make predictions on the market developments and customer behaviour. Google News Archive particularly focuses on the supply side due to the fact that it represents the popularity within the market.

2.2.1 Forecasting on breakthrough technologies

In the past there have been many attempts to forecast innovation processes, because innovation forecasting gives insights in the prospects of significant technological change. These forecasts can help managers to make better decisions in their daily activities, like strategic corporate planning, R&D management, creating competitive advantages, investments in new technologies, production and marketing and product development (Watts and Porter 1997). But there are more reasons to forecast: to identify competitors and potential customers, discover additional application arenas for the outputs of the R&D department, forecast development paths for emerging technologies, managing risks of technology development (Porter and Cunningham 2005). Basically, there are different angles from which this topic can be studied. Economists have studied market conditions, but sociologists and psychologists have tried to assess the key characteristics of the innovators. Also, there are researchers that focus on the consumers and try to estimate how potential customers react to new innovation (Langley, Pals et al. 2009). As a result, there are multiple methods which can be included in analysis, but they all focus on different disciplines. This research will mainly focus on the popularity of new innovations among consumers in the market.

Different forecasting methods

Watts and Porter mention different methods or sources for gaining information that is required for innovation forecasting. Some of them are highly-quantitative of nature; others are semi- quantitative of nature. Most of these methods are based on "*bibliometrics*". A field of bibliometrics emerge when information is captured which is closely connected to the content and patterning of the innovation. Bibliometrics can be regarded as a method where counts of publications, patents, or citations are measured and interpreted in such a way to make any forecasts for the future. These counts represent the amount of R&D activity and innovations, depending on the source. Three basic types of analysis have emerged in bibliometrics. First, *citation analysis* examines referencing patterns from patents and papers to forecast emerging research areas and detect contributions and interaction in these patterns. The second analysis is called *patent analysis*, which examines the activity of patents to profile trends and the interests of companies. Last, there is a methods which is called; *publication analysis*. This analysis focuses on articles and publications in order to reflect to R&D activities and developments. Still, not all forecasting techniques use bibliometrics. An overview of forecasting techniques is given below in Table 3 (Watts and Porter 1997; Watts, Porter et al. 1999; Vanston and Hodges 2004).

Extrapolator	Pattern analyst	Goal analyst	Counterpuncher	Intuitors
Technical trend analysis	Analogies	Implication analysis	Scanning	Delphi surveys
Substitution analysis	Precursor trends	Content analysis	Monitoring	Nominal group conference
Growth limit analysis	Morphological analysis	Stakeholder analysis	Alternate scenario planning	(Un)structured interviews
Learning curves	Feedback models	Patent analysis	Monte Carlo models	Comprehensive opportunity analysis

Table 3 – Different forecasting analysts and their techniques and methods (Watts and Porter 1997; Vanston 2003).

Basically, five different types of analysts focus on forecasting; extrapolator analysts, pattern analysts, goal analysts, counterpuncher analyst and intuitor analysts. The analysts on the left side often perform quantitative analyses, while the right side performs more qualitative analyses. Each type of analysts will be explained shortly (Vanston 2003).

Extrapolators – believe the future will represent a logical extension of the past

This type of forecasting analyst believes that the future can be related as a logical extension of the past. So, in their vision, the future continues on a large-scale in a reasonably predictable manner. They focus on past trends en extrapolate them in a logical manner.
One limitation of this method is that they use fairly straightforward logic and do not take into account possible changes due to rapid and dramatic trends (e.g. the internet). Therefore, this approach often fails.

Pattern analysts – believe the future will replicate past events

This type of forecasting analysts believes that future trends and events occur in an identifiable cycle as they occurred in the past. So, past patterns predict a replication of past experiences. Analysts believe that due to powerful feedback mechanisms in our society, together with basic human nature. Thus, these analysts believe that history often repeats itself. Examples of this phenomenon are the colour television, which followed the pattern of the black and white television, which followed the pattern of the radio adoption. Therefore, these analysts expect the adoption of the high definition television will show the same pattern of adoption as with their predecessors.

Goal analysts – believe actions and beliefs will determine the future

This type of analyst believes that beliefs and actions from individuals, institutions, and organizations will have a great effect on the future. They think the future will represent the overall aggregate of all these implications. The future becomes projected by the examination of stated and implied goals of various decision makers and trendsetters. Goal analysts include the assessment of the impact of each action by looking at the predicted goal and the assessment of the role of each stakeholder. This method faces the possible treat that certain forces are overlooked.

Counterpunchers - believe the future will result from unpredictable events and actions

Counterpunchers are the type of forecasting analyst who believes that the future will result from a series of events and actions that are essentially unpredictable and random. By identifying and monitoring a broad range of trends, events, and developments in various different environments it becomes possible to deal with the future. However, this is only possible when there is a high degree of flexibility in the planning process.

Intuitors – see the future shaped by inexorable forces, random events and actions of individuals and institutions

This type of forecasting analysts believe that the future will be shaped by a complex mix of inexorable driving forces, random events, and the key actions from institutions and individuals. They believe that there is no rational method can be used to make to predict the future, because modern society is too complex for that. Intuitors think that by collecting data as much as possible and subsequently depend on personal intuition is the best method to make predictions.

Quality of forecasting

One of the criteria what makes innovation forecasting successful is the many variables that should be included. Variables like technology characteristics, market forces, knowledge of the firm, the fit between the technology and the firm, socioeconomic factors, etc, all play a vital role. In order to create a reliable forecasting method, it is impossible to focus only on one or two variables (Porter 2007). On the other hand, it also looks almost impossible to focus on all variables that might play a role. As a result, some researchers have assessed several variables to characterize significant factors and recommend managerial practiced that explain the key successes in the development of breakthrough technologies. One of the most famous concepts is Michael Porter's four-factor framework which explains the importance of different competitive forces (Porter 1985). Another famous concept is the innovation funnel which is a broadly used concept (Dunphy, Herbig et al. 1996).

In bibliometrics it is exceptionally interesting to link certain keywords or citations. There are methods which focus on key-words, whole texts or words that appear together (Watts and Porter 1997). Co-word analysis or co-citations analysis makes sense due to the fact that it involves more possibilities to find useful results. Others call it "tech mining"; exploiting searches of electronic science and technological databases to inform science, technology and innovation management (Nasakawa and Nagano 2001; Cunningham, Porter et al. 2006). The main difference between "*bibliometric analysis*" and "*tech mining*" is in essence that bibliometric analysis lists and counts publications, while tech mining analysis is an analysis focussing more on the content (Watts, Porter et al. 1999). It allows an analyst to identify who is doing what research, identify key sources of technical information, and identify emerging or unfamiliar research that may intersect functional interests using lists, matrices, maps, or trends obtained from publication databases. Tech mining involves extracting information from texts and mining the text to discover rules and patterns. The text being analysed are primarily science and technology databases (Trumbach, Payne et al. 2006) In order to do so, it is often necessary to have a closer look at the technology itself. Porter and Watts for instance, explain an example of ceramic engines. They define keywords of the ceramic engines which have a close relation to the technology itself (e.g. durability, ceramic materials, and silicon nitride). They even used up to 201 different keywords, divided into two subgroups; material types and a combined group of material properties and applications. By using these keywords in multiple databases they were able to extent their results (Porter and Cunningham 2005).

Unfortunately, bibliometrics has limitations as well. The quality cannot be completely valued by counting the publications, citations or patents. Also, not all innovation activities are reflected into publications or patents. The high costs for patent filing or secrecy reasons are often responsible for this. Subsequently, not all companies publish equally in amounts and there are also limitations due to the secrecy of some R&D organizations. One who uses bibliometrics should also include a *time-lag*. It takes time before articles and patents are published and even more time to get included into databases. Then it should be included that institutions have different intellectual objectives in property and knowledge management. On the other hand, bibliometrics provide good accessible information, which is cost-effective as well. It offers critical innovation measures that can be collected in a relatively short time period and still looks at large amount of data. Results can be found rapidly and understandably (Zhu and Porter 2002). So, there are clear advantages and disadvantages about the use of bibliometrics. However, the few limitations can be diminished by searching for general trends instead of looking for specific events (Porter and Detampel 1995; Watts and Porter 1997; Cunningham, Porter et al. 2006).

Different characteristics of databases

Bibliometrics can be collected from different databases. Chapter one already emphasized on the potential of several databases in the fields of science, patents, and news. The analyses of this thesis will focus only on scientific and new databases, but the three databases together are particularly helpful for this type of study due to different criteria and the advantages they offer (compared to the internet):

- Scientific database (Scopus) offers the ability to search on a specific moments in time, which subsequently finds hits which are related to a certain year;
- News database (Google News) offers the ability to search on specific moments in time, which subsequently finds hits which are related to a certain year;
- These databases search on a specific topics (e.g. patents or news articles), which as a result diminishes noise;
- These databases include many reliable sources which are selected carefully.

The internet itself cannot offer these advantages that specific databases offer. Using a random search engine for a query, for instance Google, results in hits which are not repeatable due to the fact that the sources and websites constantly change or even disappear from the internet. Also, noise is one of the largest issues, because the data is not categorised. From this perspective it makes sense to focus on the specific databases alone, having the capacity to search specifically on profitable hits and being able count the hits within a specific timeframe. Second, the first application of internet was successfully implemented in December 1990[5], which results in the conclusion that the use of internet in this research is basically meaningless due to the fact that the breakthrough cases used in this research originated before 1990 (with the exception of Viagra). More information on the cases will be explained in paragraph 4.2.1. At least, the databases used in this research have the capacity to distinguish their hits over a timeframe which as a result makes them more useful for this research. The databases still differ in characteristics; therefore an overview of the main differences is given below in Table 4.

	Scientific database 'Scopus'	Patent databases in general	Business and News database 'Google News Archive'
Content	The content found concerns scientific articles, academic journals, abstracts, and archives on research literature.	The collected content can be described as visual and textual information on specific technologies filed in patents.	The content found concerns news articles, business press, magazine articles, historical archives, scanned archives, online news archives, events, demonstrations, free accessible and fee-required content[6].
Source types	The content is collected	The patents are	The content is collected

[5] *http://wiki.answers.com/Q/Who_is_the_inventor_of_the_internet*

[6] *http://en.wikipedia.org/wiki/Google_News_Archive*

	from scientific journals, trade journals, book series, conference material, conference papers, list of "conference proceedings", and abstracts, all on quality web sources on scientific, technical, medical, arts & humanities, and social sciences fields[7] [8].	collected from different databases like: World Intellectual Property Organization (WIPO), European Patent Office, US Patent Office, Japanese Patent Office, and UK Intellectual Property Office.	from sources such as: national and international newspapers, business-, entertainment-, and sport magazines[9] [10].
Includes	Includes many publishers like for instance Elsevier, Springer, Cambridge University Press, Royal Society of Chemistry, AAAS Science, etc, covering archives prior to 1995 and archives after 1996[11].	n/a	Includes many different news papers such as The New York Times, Times Magazine, etc.
Launched	November 2004 [12]	n/a	June 6, 2006[13]
Who are the creators of the sources of the hits	Scientists, researchers, technologists, academics, innovators, research orientated firms.	Industrial companies, (applied) research enterprises, entrepreneurs, innovators, universities, federal government (e.g. military, NASA), scientists.	Reporters (of news papers), trend watchers, authors of magazines, business reporters, journalists, all in combination with companies who would like to gain publicity about their new technology.
Who are the users/seekers	Scientists, researchers, academics, companies, (business) economics, business analysts, innovators, students.	(Competitive- and innovative orientated) firms, entrepreneurs, business analysts, universities.	Consumers (from early adopters to the laggards), business analysts, trend watchers, companies, scientists.
Differences in results (regarding the goal)	The ability to search over a time line in order to find content published in a certain year and with a limited level of noise.	Clear cut technical information over a time line with a limited level of noise.	The ability to search over a time line in order to find content published in a certain year and with a limited level of noise.

---

[7] *http://www.info.sciverse.com/scopus/scopus-in-detail/content-coverage-guide*

[8] *http://www.info.sciverse.com/scopus/scopus-in-detail/content-coverage-guide/sourcetypes*

[9] *http://www.google.com/support/news/bin/answer.py?hl=en&answer=1005653*

[10] For the complete list of sources:
http://www.google.com/support/news/bin/answer.py?answer=148418#rights

[11] *http://www.info.sciverse.com/scopus/scopus-in-detail/content-coverage-guide/pre1996*

[12] *http://www.info.sciverse.com/scopus/scopus-in-detail/content-coverage-guide*

[13] *http://en.wikipedia.org/wiki/Google_News*

	More limited amount of hits (compared to News hits) due to a *relatively* smaller group of interested people. Therefore, due to this smaller amount of people, fewer hits will relatively occur.	n/a	Extensive amount of hits due to the large amount of creators and sources. News papers and magazines cover a larger group of people compared to scientific literature.

Table 4 – Overview of characteristics between three databases.

Different areas within forecasting

Innovation forecasting (which also includes technical or technology forecasting) depends on three main pillars; *theory, data, and methods*. The theory explains the conceptualization of the innovation process in different areas. Often, the linkage between theory and innovation forecasting practice is weak. In technological forecasting the data is usually weak, because they often contain a lot of noise. Second, data is most useful in particular time series. Third, industries often try to protect specific and personal information. The methods used for technological forecasting try to end up with the best results with limited theory and data. In order to hold these three pillars together it would be beneficial if another area is included as well for forecasting; *monitoring*. Monitoring can even be regarded as the most vital element in technological forecasting. It enhances the perception about technological change by current developments and changes in technological and socioeconomic variables. Monitoring of related technologies, contextual influences and the target technology are most essential to complete the effectiveness of forecasting (Porter, Roper et al. 1991; Porter and Detampel 1995; Watts and Porter 1997).

As has been explained earlier, a complete innovation forecasting method includes different areas and variables in order make the predictions more reliable. A complete innovation forecasting method focuses on three areas:

1. Technology life-cycle status;
2. Innovation context receptivity;
3. Product value chain and market prospects.

The technology life-cycle status can tell something about how far the development of a new technology is progressing, the growth rate and the status of technologies on which it is dependent. The second area, innovation context receptivity, includes economic factors and other influences on the development of the technology. Product value chain and market prospects are the third area and looks at the potential payoffs and what is needed to fulfil these payoffs.

Added value of forecasting

It is clear that there are multiple ways to make forecasts about innovations of certain technologies. However, an undiscovered topic in this research is why companies would put so much time and effort in these predictions. Besides the fact that forecasting methods can help managers make better (strategic) decisions about investments, product development, production and marketing, there are more reasons why innovation forecasting is beneficial. First of all, not all technologies are the same.

They differ by complexity and sector. Therefore, there are multiple ways to successfully launch a new technology in the market. Second, technology cycles and trends also notice a significant change on a regular basis (Schmoch 2007). Science-based technologies with a high-level of complexity are a growing trend in industrialized countries for the last decades. It would be useful if the evolution of technology cycles gets more reflection. Many different cycles have been explained in the past, for instance market and technology cycles. What sometimes occurs is a hype cycle, which is a peak before the S-shape of the technology life-cycle reaches maturity. It reflects the pattern with respect to hype and time. For a short time there is a over-enthusiasm by customers about the new technology (Linden and Fenn 2002). Another cycle is suggested by Utterback and Abernathy. They explain the concept of double cycles of technology. The first cycle occurs when the focus is on innovation activities on products in early stage. However, in later stages when the product group is mature the focus is more on price competition and process innovation. This is when the second cycle emerges (Utterback and Abernathy 1975).

Thus, there are many different cycles explained in the past. One is about the shift between long-term and short-term activities; another is about a shift from product to process innovation. Still, cycles and trends can be found and should be included in forecasting methods. The question remains in what way innovation forecasting methods or tools are useful for companies. To begin with, it helps in planning for technology development and formulation of technology strategies. It is used as a method to diminish the probability of failure and improve performance of the technology (Cooper and Schendel 1976). Companies use technological intelligence to prioritize R&D, new product development planning, but also because it helps making decisions about licensing and joint ventures (Zhu and Porter 2002). The focus lies then on technology assessment, technology forecasting and technology monitoring. An overall term is *technology intelligence* (Lichtenthaler 2007). Technological intelligence is about discovering the opportunities of a new technology and how to not to miss these opportunities. Technological intelligence is often also mentioned with other terms as well, more specialized in their function, but all of those will be discussed later. More importantly, technological intelligence is executed as a method to make forecasts and to gain unique insights regarding issues within a firm's business environment (Miller 2000).

2.2.2 Business Intelligence

In the previous paragraph a more detailed focus was at the technological forecasting methods and why these forecasts can be of great importance for companies. It ended with the explanation of *technology intelligence*, on which we will now further elaborate. Basically, technological intelligence is about discovering the opportunities of a new technology and how not to miss these opportunities. It is a method to make forecasts and to gain unique insights regarding issues within a firm's business environment (Miller 2000). However, technological intelligence is among experts also known as other definitions. Although they do not differ a lot, we will emphasize all of them in this paragraph to clarify the small differences. The term *business intelligence* will be used after this analysis, because it reflects the best form of intelligence in this research.

Advantages

Technological intelligence has many advantages over an expert-based approach. First, technological intelligence handles much larger volumes of information. Humans would never have the capacity to analyse all this information or it would cost a lot of time and money to do so. Including less information would be less effective in technological intelligence. Second, a tool for technological intelligence has the capacity to analyse these large amount of information and can produce useful results. Many variables can be included and checked with statistical analysis. Last, a tool for technological intelligence has no problems dealing with updated information. Data constantly changes over time. A tool can make use of real-time data and as a result there would hardly be any time lag between the collected data and results of the technology intelligence tool (Yoon 2008).

So far there are just a limited amount of technological intelligence tools available. They focus on different sources. *Diva*, for instance, is a visualisation system for exploring document databases. As a source it looks at patent databases and scientific literature (Yoon 2008). However, it does not include other bibliometrics like for instance news articles (collected from Google News) which would be of tremendous value to say something about publications towards the consumer. Another tool is *VantagePoint* and *Aureka*. These tools can detect trends or relationships (Zhu and Porter 2002; Trippe 2003). *Techpioneer* is another tool, but has an extended feature. It can “communicate” with domain experts. As a result, it becomes possible to reflect the expertise knowledge which distinguishes itself from other tools by performing as a hybrid tool (Yoon 2008). All tools seem to have their own information sources and specialities which make them useful for one particular company, but cannot add value to another company. It must be said that not all traditional forecasting methods are fully reliable, because they rely on the intuition of domain experts and qualitative data.

The advantages of technological intelligence for technological forecasting are mentioned above. However, an untouched area is still about why traditional processes, like market research, are not sufficient (anymore) to determine the strongest strategies which are needed in a competitive environment. An answer to this question focuses on the environment where companies are currently operating in. There are many competitors with strong capabilities who see opportunities to develop new breakthrough innovations. In order to survive the competition, companies need to create competitive advantage and try to stick to that position. The tough competition is therefore the key driver and motivator for technological forecasting (Zhu and Porter 2002).

Multiple methods are explained by Michael E. Porter to achieve a competitive position (Porter 1985). Once this method is determined, the company needs to work towards this result. When a company has more detailed information about the developments and the possible diffusion of this of a new breakthrough technology they will face less risks and create more certainty to accomplish their strategy. Due to the highly dynamic environment (e.g. trends and continuous market changes) it is of paramount importance that all sources of information are utilized to the fullest. Market research will therefore not be sufficient. Information in high-technology markets is time sensitive and therefore quickly loses its value. The traditional methods rely on the expertise of market experts that work with data that contain a time lack, results in unreliable results (Cunningham, Porter et al. 2006). From an overall viewpoint it also becomes clear that market research has a validity problem due to the fact that this research is not useful in every phase and does not use rich intelligence from data resources. When we look at the technology life-cycle pattern it becomes clear that market research cannot predict much about the developments of a new technology in the innovation phase, because consumers are not yet aware of this technology and cannot judge about something they have no experiences with. Moreover, market research is quite time consuming which is costly and complex.

As a result, a methodology for forecasting which is only based on market research will in most cases not be sufficient for supporting strategic decisions. Subsequently, the possibilities to gain more insights in the future technology life-cycle of a breakthrough technology are endless nowadays. However, the term market research will not be applicable anymore, due to the fact that there is a shift towards technological intelligence. As has been explained, technological intelligence includes much more methods to make any predictions. Especially in dynamic environments, companies try to focus more in intelligence tools, rather than only on market research.

Different definitions on intelligence

So far we are discussing about technological intelligence. However, there are several different terms which represent approximately the same. Some call it technological intelligence; others call it competitive analysis, business intelligence, competitive intelligence, technology foresight, competitor intelligence, technology roadmapping, or environmental scanning (Pawar and Sharda 1997; Watts, Porter et al. 1999; Miller 2000; Zhu and Porter 2002; Hannula and Pirttimaki 2003). Still, intelligence in general has one goal. The intelligence process is based on the assumption that managers seek to become better informed about the critical issues on a formal and systematic basis. The Society of Competitive Intelligence Professionals (SCIP) formulates intelligence as:

"intelligence is the process of ethically collecting, analyzing, and disseminating accurate, relevant, specific, timely, foresighted and actionable intelligence regarding the implications of the business environment, competitors, and the organization itself" (Miller 2000).

Despite the multiple terms, business- or technological intelligence can be seen as a natural extension of corporate strategy activities. It can support competitive technological intelligence activities, strategic planning, R&D management, intellectual property management, and process improvement (Watts, Porter et al. 1999). Although they all are closely related to each other, there are some differences between each term (Miller 2000; Lichtenthaler 2003; Moss and Atre 2003; Nosella, Petroni et al. 2008). A small overview is given below.

Business intelligence

Business intelligence focuses on the monitoring of a wide range of developments across an organization's external business environment or marketplace. It is an architecture and a collection of integrated applications for operational and decision-support issues and databases that provide the business community easy access to business data.

Competitive intelligence

Competitive intelligence looks at the current potential strengths, weaknesses and activities of organizations with similar products or services within a single industry.

Competitor intelligence

This type of intelligence looks at the profile of a specific organization.

Technology intelligence

A task, which is performed independently, has the goal to exploit potential opportunities, and to secure potential threats with the help of quick collection of relevant information about technological trends in the environment of the company.

Overall, business intelligence is one of the most used terms in intelligence. It is about the analysis of information from the marketplace and the generation of recommendations for decision makers. Nowadays, business intelligence, as it will be called from now on as an overlapping term for different forecasting methodologies, is not only applicable for large companies. Smaller companies, with less than 100 employees, conduct intelligence as well. Internet is the key with business intelligence, because it is a set of interconnected networks that contains billions of computers connected to these networks. Obtaining information from the internet is therefore simple, cheap and can be done at any time of the day. Moreover, due to the large amount of internet users it becomes possible to receive information which can tell something about a large group of people.

Business intelligence tool

It has been explained what the main goals are by conducting business intelligence. It mostly supports decision making by managers on different aspects. It also offers them to keep up with changes in the marketplace. By making decisions based upon potential opportunities and threats it becomes possible to create a competitive advantage. Especially governments conduct intelligence in order to determine threats. However, firms often use intelligence more for determining the opportunities that ware undiscovered beforehand.

Basically, managers have several options in their decision making. The first option is to rely on their past experience, expertise for decision making and trusted people in their network, without making use of an intelligence tool. This option is not very rational, due to the fact that the decision is based on only a few information sources. The second option is to ignore all marketplace fluctuation completely. These managers simply execute their business as usual. Decisions will be made without the knowledge about the industry or about market changes. The managers believe their rank and title corresponds to sufficient understanding for making the decision. This option has led many firms lose market share or even bankruptcy. The last option is when managers involve intelligence work. Nowadays, this seems to be the most logical choice, especially for smaller firms (less than 500 employees), because these managers need to learn the basics of business intelligence due to the fact that such firms cannot hire intelligence professionals. The final option concerns to rely (hire) on the expertise of intelligence professionals to coordinate the information. This is for medium/large companies interesting when they face limited time to conduct a full analysis of the collected data. In order to succeed this option, it is important that the entire staff learns more about the benefits of intelligence, especially the key managers across the firm (Miller 2000).

To conclude so far, in order to make reliable forecasts that can enhance the rise of new undiscovered opportunities, companies can implement a business intelligence tool. Business intelligence greatest use is for forecasts and decision support, though it is also used for other reasons, like visualisation, geospatial analysis, reporting, and other activities (Moss and Atre 2003). The intelligence process can create competitive advantages for the firm (e.g. the competitor's next strategic move or the production capacity of the competitor). However, the scope of the intelligence process is very broad and can be applied for different issues. These different issues can for instance concern the firm's corporate strategy, operational efficiency, competitive advantage or position, or new product development planning (Miller 2000).

2.3 Creating intelligence

As has been explained earlier, this research will focus mainly on scientific- and news databases to see whether there is a correlation between the technology life-cycle and the amount of publications. Although this research will emphasize on one particular search term (the technology name itself which will be further explained in chapter 4), it is also possible to combine different search terms creating a more effective query due to the involvement of more useful search terms. In this paragraph we will try to cover factors which might ultimately lead to new search terms for queries which can be used in future research. A model of these factors will be presented. Using specific search terms in search engines will result in more effective hits relating to the query. By measuring *the raw counts* it becomes possible to detect possible growths in attention which can be related to the technology life-cycle pattern. Having this knowledge gains support for making strong strategically decisions, because it represents a measurement of popularity, activity, attention, and the chances of success. Thus, it offers the decision maker the opportunity to make a more rational decision.

2.3.1 R&D development and Google News

The development of a breakthrough technology will always develop according to several subsequent steps. Basically, if a new technology is invented, it often needs improvements to make the technology reliable enough to be implemented into a product. Once this phase is completed, the product is ready to be introduced in the market. As a result, the basic steps of development can be related to the technology life-cycle. It may be clear as well that the different steps of development focus on different actions from R&D departments. If it is requested to say something about the developments of the technology, a closer look at these actions might be fruitful because they have the capability to say something about a particular action. These actions (or milestones in the R&D process) can be used into a search term in Google News (or in other databases). However, if the query becomes too complex, the results of generated data will diminish, though they might become more reliable.

For an effective business intelligence tool it is important to find search terms which rely on the fit between many variables (e.g. socioeconomic factors, market forces, economic climate, and the new technology) (Watts and Porter 1997). However, in this research the focus lies mostly on the application of Google News. Due to the fact that Google looks for business- and news publications which are written to inform the market, we tend to focus on factors which are most efficient in this area.

Presumably, there can be thousands of search terms used in the query, because there are many factors in the technology life-cycle of a breakthrough technology which represent a fit with the technology and have the capacity to say something about the maturity of the technology. In this literature study we will try to complete a list which represents some of the most important factors which could be used for further research. However, in this thesis there is only limited amount of time, resulting in analyses which focus only on the technology name itself. Nevertheless, the model can be applied for future research.

One remark concerns the opportunities which can be generated with Google News. It mainly looks at news articles published in the past. Therefore, it will represent the awareness and popularity of the consumers, since newspapers try to write about news which is popular among consumers. Therefore, Google News will be most effective with search terms that focus on a combination of the (development of) breakthrough technology and the interest of consumers (i.e. the market).

2.3.2 Fields of factors for Business Intelligence

This paragraph will focus on possible factors that can be used in electronic databases. These factors represent something, such as progress of new developments, awareness of a new technology in the market, opportunities for breakthrough technologies, etc. By using search terms in electronic databases which relate to these factors it might become possible to make further forecasts.

Fields of factors are used in many different situations. There are factors which are used to say something about the level of R&D innovation of a company. Other factors are used to say something about the financial performance of a company or try to determine the progress of the development of a breakthrough technology. We focus on the latter in this chapter, due to the fact that these factors can be implemented in a business intelligence tool to make forecasts based on bibliometrics.

As has been explained in the previous paragraph there are basically three areas to be included in forecast methodologies. These areas are: *technology life-cycle, innovation context receptivity, and product value chain & market prospects*. It is hard to find the most useful factors which effectively represent helpful search terms in order to make reliable forecasts. Also, a complete business intelligence tool should implement both internal and external areas around a technology. Figure 10 shows how this process looks like by defining different fields which are related to the firm and can therefore be used as a source of inspiration to find proper search terms (Stopper 2002).

Basically, the internal and external fields can be used as a first step in future research, since it overviews the environment of high-tech firms. This brings us to the following step, finding useful factors divided over the technology life-cycle serving as a building block for discovering search terms for in more effective queries. Figure 11 illustrates several groups of factors concerning technology, innovations, economics, environment and society divided over the different phases of the life-cycle. This figure can be used as a start-up for future research once more detailed studies are being applied. The mentioned factors might play an important role in a specific phase of the life-cycle, and can therefore be regarded as a source for inspiration looking for applicable search terms. Search terms based on these factors might surface when future research requires combining the name of the technology with additional search terms. Figure 11 can be taken into account as a first conceptual building block, but it has to be taken into account that the model is still incomplete.

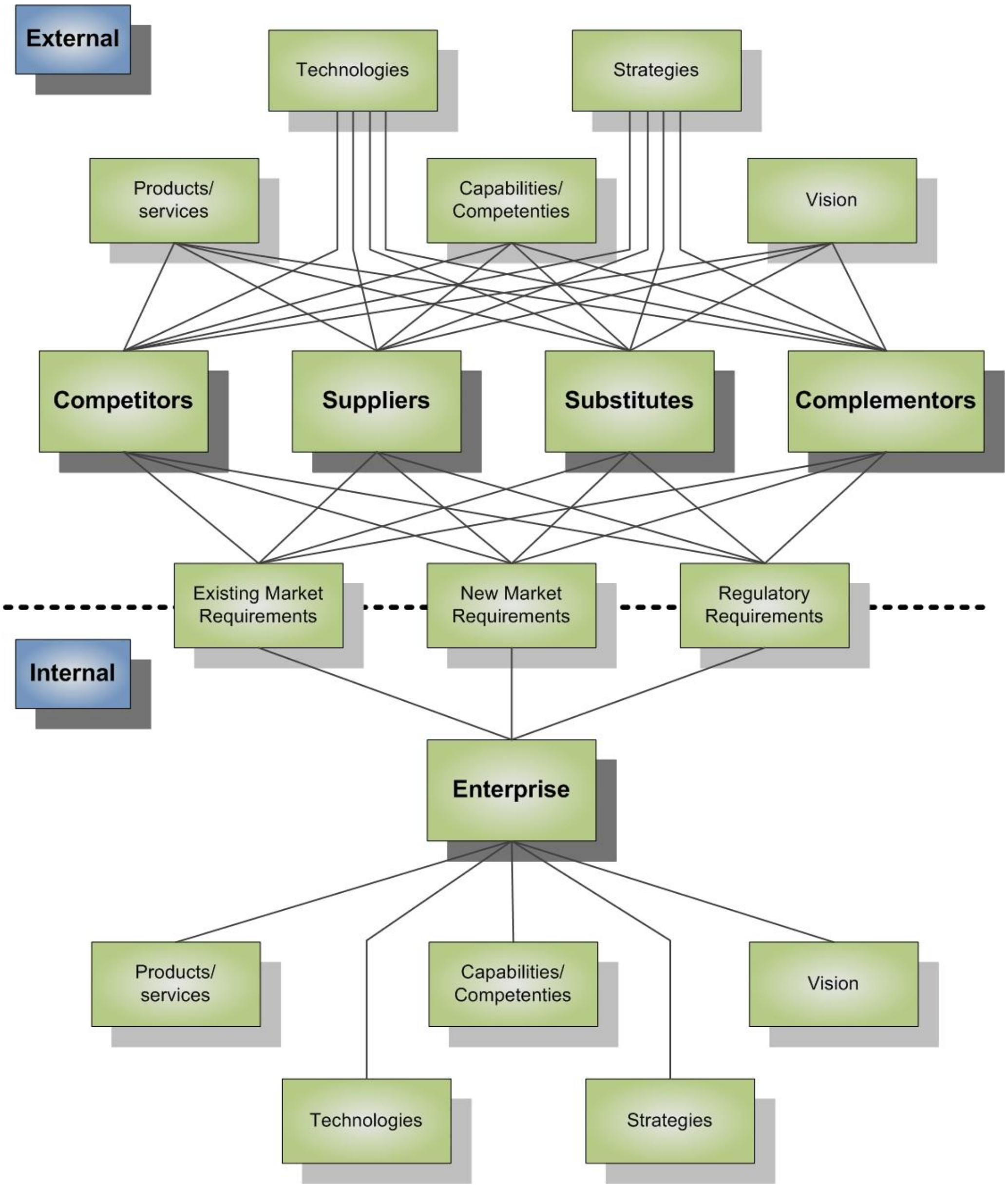

Figure 10 – Internal and external fields related to the value chain of a company and might be used to take into account for future research (Stopper 2002).

Conclusion

Both figures explain factors which are closely related to the development of breakthrough technologies and high-tech firms. They can give inspiration when future research is desirable and can give clear-cut search terms for when a more elaborated query is required. So, new search terms might be derived from these factors, using them in combination with the name of the technology itself. Then it becomes possible to measure raw counts of hits, which can ultimately be used to measure the popularity, specified on an explicit topic or breakthrough technology. Measuring the popularity by calculating the increase or decrease in publications can give tremendous information about the development, acceptation, popularity, and chances of the success of certain technologies.

Figure 11 illustrate the different factors that can be used as a source for finding targeted search terms in order to measure the amount of raw counts of hits and dive into the dynamics of the patterns. However, this model needs extra care and further exploiting before it can be used in databases when useful search terms derive.

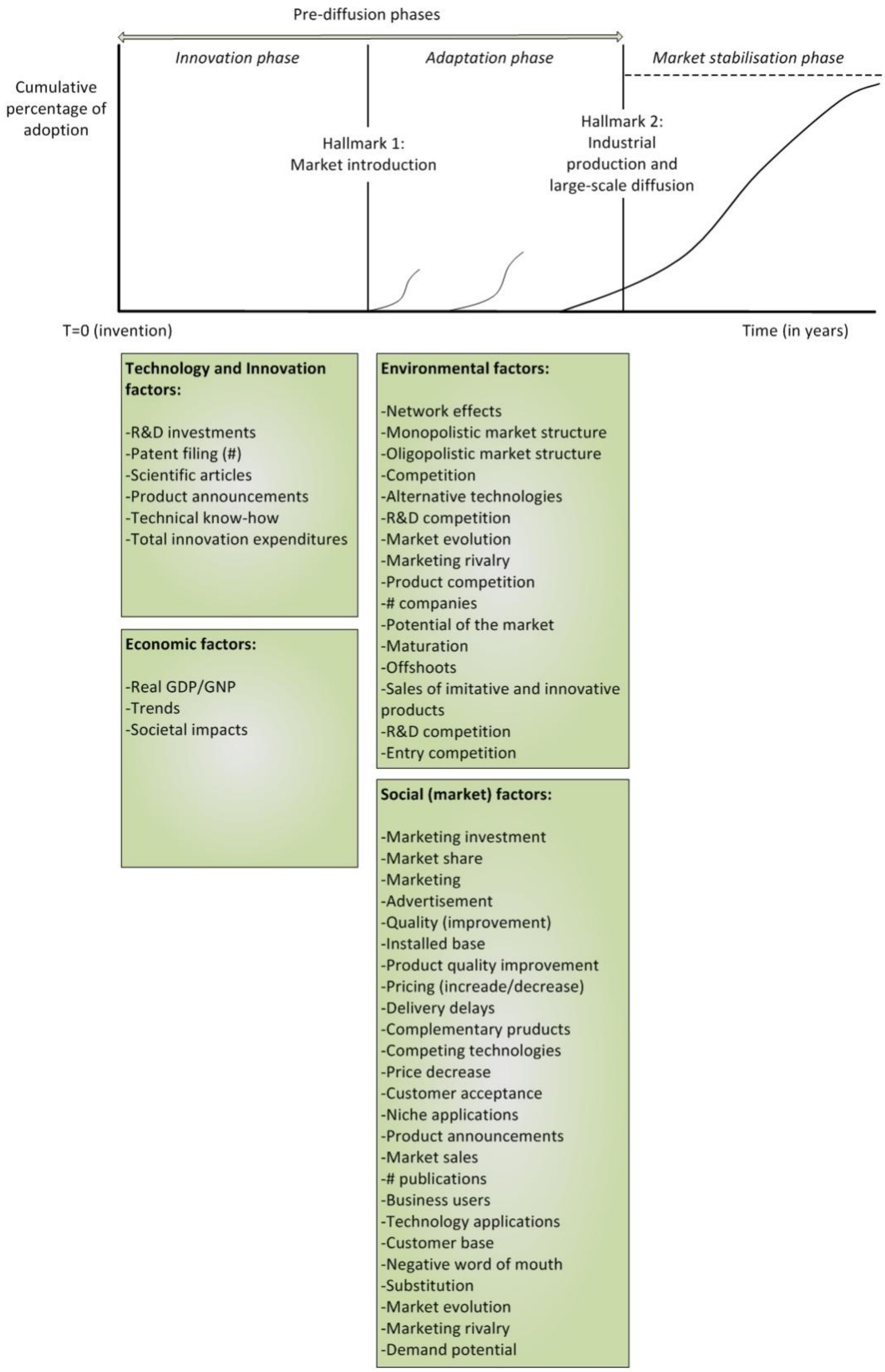

Figure 11 – Useful categorised factors which might give inspiration for future research.

2.3.3 Process of data analysis

Since the start of this thesis it is mentioned shortly why data is collected from databases, how it is collected, and how it will be treated. A thoroughly explanation of the complete iterative process is not given though, and will be explained in this paragraph. The goal is to overview how the data is treated in order to reach the final vision about the added value of electronic databases in forecasting tools. A few steps will be proposed based on literature (Porter and Cunningham 2005; Porter 2007):

- Data selection;
- Data collection;
- Data treatment;
- (Basic and advanced) analysis;
- Information representation;
- Decision making process.

Data selection

A selection has to be made which sources and/or databases will be used before any data can be collected. A decision should be made which database, publications, citations, patents or abstracts should be included. Simply choosing one will not do the trick; deliberate choices have to be made on sources that match best in the analysis.

Data collection

The data in this research will be collected from databases, aiming at the collection of hits which represent the amount of attention for a specific topic (e.g. popularity among consumers). By designing specific and effective queries a search can be executed within a particular database. After the search the raw counts will be expressed into a pattern expressing the activity of the search over time. More about this topic can be found in chapter 4.

Data treatment

The collected data will contain noise that has to be reduced. Multiple methods are available, for instance by excluding certain occurrences that are easily selected by including other information as well (e.g. the history of a technology or by including external sources which are able to filter out peaks in the patterns which are caused by external causes). By merging related (technical) topics, combining different name variations, and by excluding poor qualitative data it becomes possible to clean the data. More information can be found in chapter 4.

Analytics

Once the data is pure and clean it can be analysed in multiple ways (more information about this in chapter 4 and 5). Both qualitative and quantitative methods have potential to find results with lots of added value.

Information representation

Information on findings has to be communicated towards the decision makers. This has to be done in such a way that no data get lost during this process. A distinction should be made between "information push" and "need-to-know pull". Only relevant information should be transferred in a clear and visible manner towards its end-users in this process of transfer.

Decision making process

Finally, the decision can be taken by the decision maker. The intelligence brought by the analysis will improve the decision making process in order to gain competitive advantage, identify competitors and new markets, to gauge market potential, and to reduce risks of technology development.

3. THE ADDED VALUE OF DATABASES FOR TECHNOLOGICAL FORECASTING

In chapter two we have elaborated a bit about innovation forecasting and how technologies develop and mature over time. Also, it has been made clear that certain indicators and sources can be used to determine the future of a technology. Many different sources and databases are used to forecast the developments. This paragraph can be divided into two parts. First, we will elaborate shortly about cycles, trends and sources which can be used for intelligence. Second, we will analyse the findings in literature and check whether they match to our findings. In more detail, we will check when certain databases (scientific – and news databases) show the highest activity in a specific part of the technology life-cycle in order to forecast the breakthrough technology.

More valuable information on this topic is useful for the second analysis which will be performed in chapter 5. Basically, when it is recognized that in a specific phase of the life-cycle there is a higher level of attention, this could give the decision maker more knowledge. It tells him that certain sources in a forecasting tool *might* be more reliable or useful at a certain moment. The analysis in this chapter will greatly enhance this perspective because it will focus on the scientific- and news activity. The results can be taken towards the second analysis, because a distinction can be made how innovation or diffusion can be measured. Activity in science and news can be related to locate the most appropriate source to investigate the innovation, adaptation, and diffusion of a technology.

3.1 Cycles and trends

Chapter two already entered an important topic which should be included in this thesis. This research mainly focuses on the added value of news databases in the field of forecasting. However, in literature it is regarded that for intelligence certain forecasting sources are most useful in only a specific part of the life-cycle of a technology. We will first shortly elaborate what is explained on this topic.

Schmoch (2007) explains that a clear distinction should be made between science, technology and market activities in order to further understand technology developments. A long-term analysis of the developments of science-based technologies caused the discovery of double-boom cycles of technologies (innovation cycles). The first boom can be related to science/technology-push and the second boom can be related to market-pull. The first technology cycle can be decomposed into several steps of basic research, applied research, experimental development, and innovation. The second market cycle is described in terms of turnover and profit with an S-shaped curve, followed by stagnation and the collapse of this cycle with as a result a bell-shape cycle (Bass 1969). Basically, the cycles explained in literature primarily suggest mechanisms such as different product generations, the fluctuation between long-term and short-term activities, and the shift from product to process innovation, all in a *sequential* manner. However, although this "science-push" model, as it is referred to it general (some authors regard it as a market-pull model), is acknowledged by different literatures, Schmoch explains one shortcoming. He introduces another model, based on the "concomitance model" (Schmidt-Tiedemann 1982): the "*interaction model*" (Schmoch, Hinze et al. 1996). This innovation process can be divided into the subsequent phases of 'exploration', ‘innovation’, and ‘diffusion’. Research, technical and commercial functions develop over time *in parallel* and are linked interrelated. A further development of this process shows activities and interaction of research institutions, enterprises and the market (Schmoch, Hinze et al. 1996; Schmoch 2007).

Now, coming back to the goal of this literature part, according to the science-push model it is often expected that a first bell-shaped curve of publication activities emerge that can be linked to basic research (curiosity-research). This curve is followed by a second bell-shaped curve of patent activities which can be related to applied research and development. Market diffusion will start shortly after the patent cycle. Basically, the model is a sequential follow-up of stages (Bush 1945; Myers and Marquis 1969; Utterback 1974; Irvine and Martin 1984). Though these expectations seem very logic in general, according to several authors (Grupp, Schmoch et al. 1990; Schmoch, Grupp et al. 1991; Schmoch, Hinze et al. 1996; Lacasa, Grupp et al. 2003; Noyons, Buter et al. 2003) these activities of publications and patents continue in parallel up to market diffusion instead of sequential linear activities. Figure 12 shows in a graphical way how the chronological order of scientific, technological, and business develops over time according to many authors.

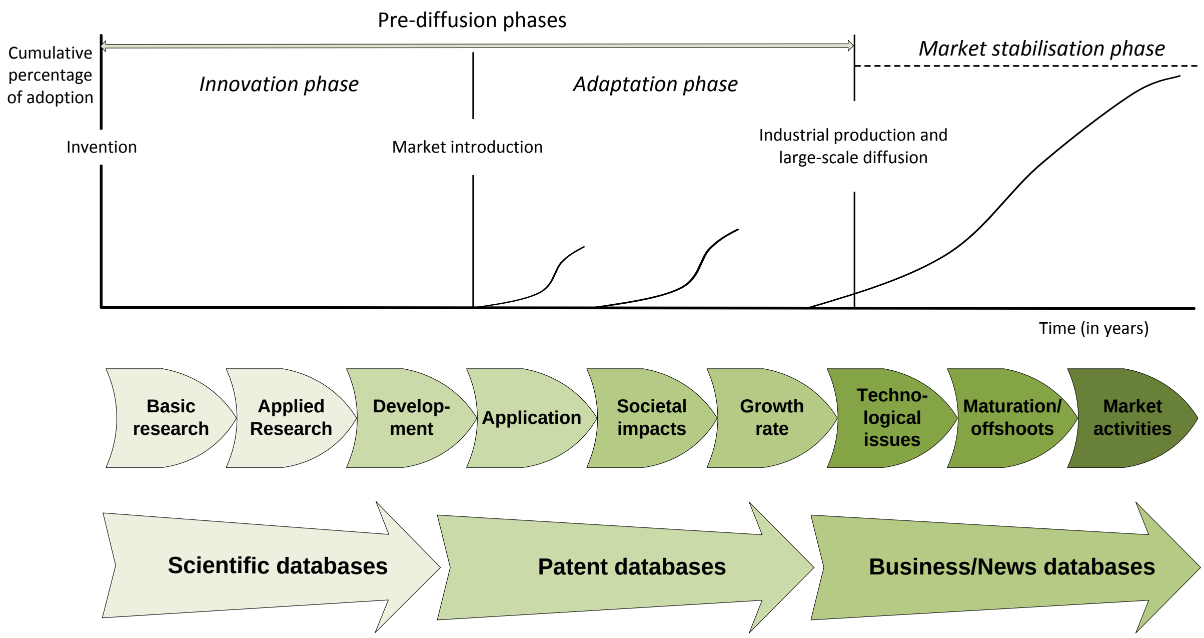

Figure 12 – The chronological order how different fields of activity are expected to develop over the technology life-cycle of breakthrough technologies according to many authors (*science-push model*).

Various electronic databases are available for specific and flexible analyses. Patent databases for instance, are regarded as good indicators due to the reflecting activities on inventions and breakthroughs, even though only 40-60 % of the patents are actually used for innovation (Grupp 1998). As a result, the number of patent applications can be regarded as an indicator for corresponding technology activities. Also in patent applications double-boom cycles have been found in some cases (Grupp and Schmoch 1992). The number of patent applications increases in the first stage of the development of the technology, finds a maximum and decreases thereafter. About 15 years after the start of the first growth, another growth occurs and even exceeds the first growth. The question rises if this phenomenon occurs due to the fact that patents expire after a certain time or that new patent applications are filed. Although this double-boom pattern is not always present (only in about 50% of science-based technologies), it explains how patent applications can be of added value for the determination of trends and how this information can be used for forecasts (Schmoch 2007). However, this says nothing about sales or turnover, but similar patterns might also occur for other (electronic) databases. Other databases could indicate other areas of progress, such as scientific databases or news databases. Though it is often expected that scientific databases show the highest activity in the beginning of the life-cycle of a technology, this might be doubtful. Especially when you realize that scientific activity not only focus on the major breakthroughs, but also on the improvement of control issues which enable a broader application spectrum. Schmoch (2007) believes that scientific activity extends over a life-cycle of a technology and starts almost simultaneously with the patent activities. His arguments are that the results of basic research are not taken into account either by technology or science, even though they often appear several decades before the first patent boom. He found an enormous time lag between the first ideas and technical realisation and a relevant scientific and technological activity reflected in publications and patents. However, he analysed only a few technological fields. Therefore it is interesting to investigate when scientific databases show their largest activity. Another reason why this topic is particular interesting is because scientific activity is regarded more reliable than technological activity (patents), because science is less sensitive to the environment and the subsequent problems (i.e. companies often switch and prioritise their actions), according to Schmoch. Also, not all companies file patents due to financial and secrecy issues. Science on the other hand shows a more stable environment (especially when it concerns basis research). Market activity on the other hand does not reach a significant peak until the beginning of the second technical boom according to Schmoch (2007). Market activity steadily increases and do not follow the cyclic course of the technological activities (patents). Again, the analysis by Schmoch was done for only a few technologies. So, double-boom cycles (science-push and market-pull) occur in about 50% of the cases and further observation on specific databases is very interesting and can add value for further analysis on forecasts.

However, Schmoch's vision can also be found in other literature sources. Rosenberg (1982) explains his vision on to what extend science is exogenous. Many scientists believe that science is an exogenous phenomenon (Kuznets 1966), a view where economic growth only occurs due to science from an external environment. However, Rosenberg believes that significant chains of causation run from economic life to science as well as from science to economic life. He proposes several reasons why he thinks that science can be of endogenous nature; one is due to the power of economists. Powerful economic impulses are shaped, directed and constrained in scientific firms, because science is a costly activity and it can be directed in such a way that large economic rewards are possible. Thus, daily economic life has become closely linked to science and therefore science has transformed in a more endogenous activity instead of an exogenous activity (Rosenberg 1982). It is technology that provides the measurement instruments to do scientific research.

From the viewpoint of Rosenberg it seems that scientific databases do not show their largest activity in the start-up of the technology life-cycle, but even more in the adaptation phase. This viewpoint can be linked to the one of Schmoch, were scientific publications grow in parallel with patents and extend over time. He reckons that scientific activity starts simultaneously with the start of patent activity or even a bit earlier. Rosenberg believes that technology provides the measurement instruments to do scientific research. As a result, this would entail that in the innovation phase (and a small part before the invention) technology is the cause for the growth of scientific activity. Thus, it seems like Rosenberg suggests that patent- and business activity indicate the technology especially in the innovation phase, although he does not mention this explicitly. However, he does mention clearly that science occurs more in the adaptation phase and the market stabilisation phase. Rosenberg also explains that in some scientific cases a double-boom occurs. He gives an example of the material Polyethylene, which shows two peaks in scientific activity. His arguments are that when this material was broadly applied in technologies, it took years before many aspects of the material could really be tested within these applications (e.g. fracturing, contamination, aging, corrosion, brittleness, etc). He explains a whole new generation of problems arose *after* it had been installed, which ultimately caused a second generation of research for a deeper understanding of the technology. In this way, a double-boom cycle could be present for scientific activity as well.

Thus, Rosenberg and Schmoch have some overlap in their viewpoints; however they also have some differences. The first difference is that Schmoch mainly focuses on the relation between science and patents, and Rosenberg mainly focuses on the relation between science and economic (business) performance. Second, they see different scientific patterns. Schmoch sees a clear growing pattern in general, while Rosenberg often sees a double-boom cycle which can be related to further scientific research after the technology is implemented and tested in the market (e.g. Polyethylene). Third, they have a different viewpoint why scientific activity occurs. Schmoch blames basic research, though this research is often not published. Rosenberg agrees with Schmoch on this viewpoint; however, he believes that economic and technical measurements instruments can be blamed as well.

Conclusion

To conclude; different databases can be used to assess the scientific, technological, and market activity. Each database shows the highest activity in a specific part of the technology life-cycle, even though some researchers (Schmoch, Grupp, Rosenberg) believe that there are some cases where certain databases are important over a much longer period over the life-cycle (Trumbach, Payne et al. 2006). Still, overall it is often regarded that scientific databases provide insights into technical breakthroughs, inventions and innovations especially in the beginning of the life-cycle. Patent databases provide insights in the developments of technologies before the large-scale diffusion occurs. Last, business and news databases offer insights in the breakthrough technology and in what way it is attracted by the interest of the marketplace (Bush, Myers and Marquis, Utterback, Watts). However, especially the latter database is hardly discovered and implemented in forecasting methodologies. So, in order to determine the degree of interest and maturity of a technology a chronological analysis of publication and patents activity seems to be of importance. Patterns can be interpreted in terms of empirical evidence of the progression of the technology life-cycle (Watts, Porter et al. 1999).

Besides this science-push theory that many scientists embrace, there are other viewpoints as well which explain that scientific-, technological-, and market activity mainly occurs in a specific part of the life-cycle. Also they believe to see certain trends or cycles. For a complete vision of the entire different viewpoints in literature, we have developed an overview of all theories. This overview can be found in Figure 13. In the following paragraphs we will analyse whether these viewpoints correlate with our findings. We will do this by looking at 14 cases from two industries by collecting data from news databases and scientific databases. We will than answer our hypotheses mentioned in chapter one. More details about the two industries, the 14 technologies, and the dates of the hallmarks will be explained in paragraph 4.2.

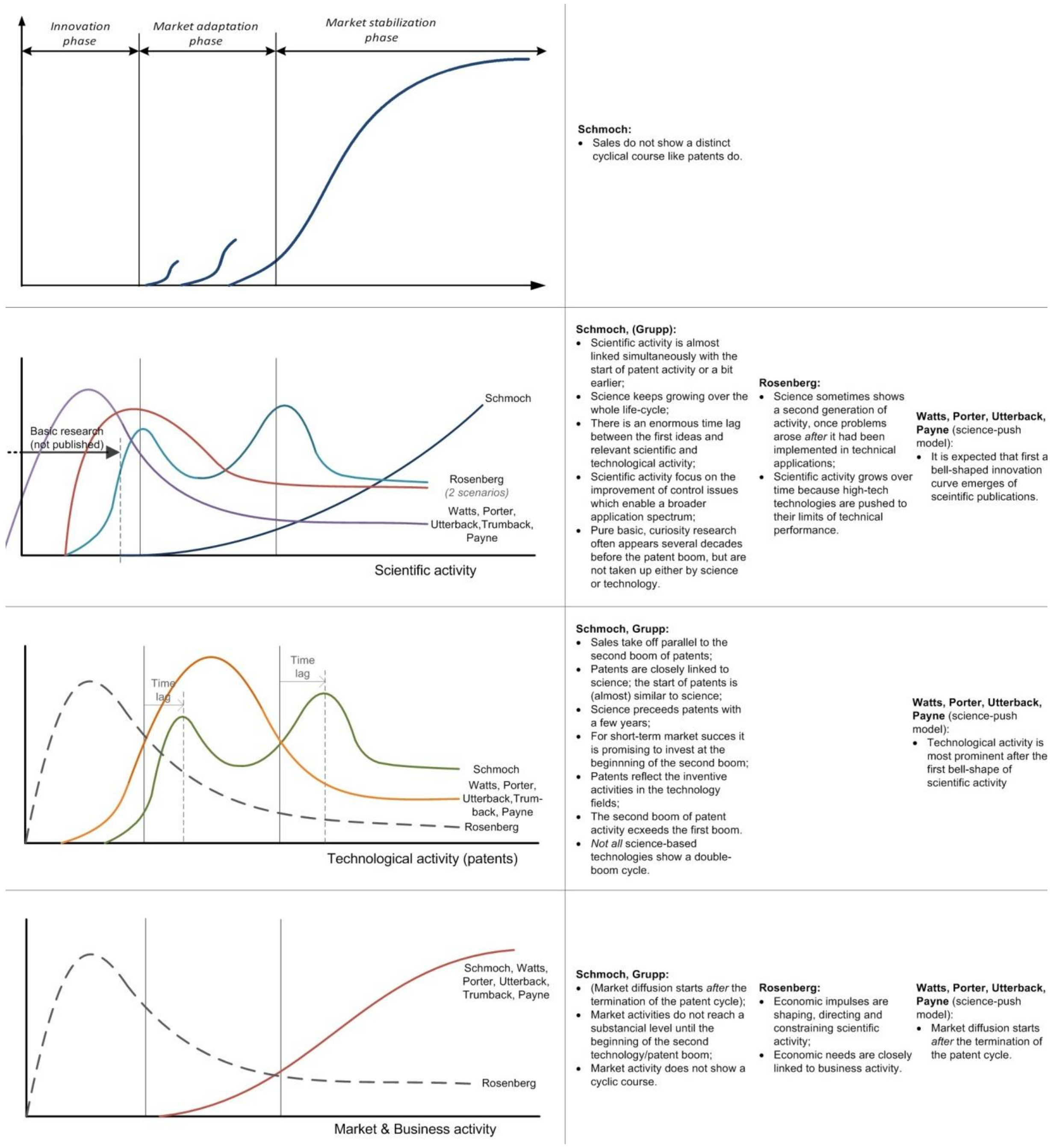

Figure 13 – Overview of different viewpoints in literature of trends and cycles.

3.2 Approach for analysis

In the next two paragraphs an analysis will be performed on news- and scientific database. During the analysis an answer will be found on the question of when activity is highest over the technology life cycle in science and news hits. Afterwards, it becomes possible to see which viewpoint described in the previous paragraph matches best with the data.

Calculating the average

The analysis will calculate the average of hits per technology phase from 14 different cases from two different industries. More information about these cases is given in paragraph 4.2.1. The average amount of hits per phase will be calculated because the length of the phases can vary quite a lot. As an example we use Paracetamol, which has an innovation phase of about 75 years and a market adaptation phase of two years. Therefore, calculating the average will reassure that we do not compare apples with oranges. The average will give us the opportunity to conclude from our results. Although the length of the innovation and adaptation phase are determined by the hallmarks, the length of the market stabilisation phase remains undefined. In order to calculate the average of the market stabilisation phase as well we need the length of this phase. The following criterion is used to calculate the average hits subsequently: *the length of the market stabilisation phase is equivalent to the average of both the innovation and market adaptation phase with a maximum of 10 years*. As a result we are able to calculate in which phase the amount of hits is highest.

Ranking

After calculating the average of hits per phase we will rank the numbers. The ranking is done by allocating a number (from one to three) that categorises each phase based on the amount of hits. A *three* is the phase with the highest activity; a *one* represents the phase with the lowest activity.

Reasons for ranking the hits are because of several reasons:

- Ranking offers a fair way to compare the phases. The amount of publications is dependent of many variables (e.g. the time period of development, complementary products and services, etc) and in order to compare the different cases in a fair way, ranking offers a solution;
- Ranking clearly visualises how the patterns of publications develop, although some information about the difference will be neglected;
- With ranking it becomes possible to test how often for several technologies most publications occurred in a certain phase. It becomes possible to look at the patterns without looking at the absolute numbers;
- Ranking offers the possibility to compare different technologies, which possibly have a different number of publications (e.g. Viagra vs. Minoxidil);
- By adding up the ranks and by calculating the mean, the possibility emerges to conclude from these findings in a clear and structured manner.

Testing

After this ranking a statistical test will be performed in order to check the significance of the results. More information about the testing will be done in the following two paragraphs. In this test we will use a significance level of $\alpha = 0{,}05$, since this is commonly used in these field of research.

3.3 Business Press and News databases

This paragraph emphasizes first on business press and news databases (Google News) and when they appear to show the highest amount of activity (i.e. amount of hits) over the technology life-cycle. For this analysis we have focused on only the 'Timeline' pattern offered by Google News due to the much better results it offers compared to the 'News Article' pattern. More information between the differences of these patterns is given in paragraph 4.2.2, and more information about Google News and its characteristics is given in paragraph 2.2.1. Table 5 illustrates an overview of the average hits found in the 14 cases from the two industries.

Absolute average amount of hits found in Google News			
Breakthrough innovation	Average hits per year in Innovation phase	Average hits per year in Market Adaptation phase	Average hits per year in Market Stabilisation phase
MATERIALS			
Astroturf	10	82	532
Kevlar	26	40	53
Dyneema	29	196	530
Nylon	117	32	2127
Teflon	35	22	44
PET	22	61	304
PHARMA & HEALTHCARE			
X-ray	1390	1310	1310
Aspirin	15	232	46
Paracetamol	2	3	5
Contraceptive pill	0	1	2
Viagra	218	7740	8923
Prozac	57	56	53
Polio vaccine	463	5800	3745
Minoxidil	1	1	71

Table 5 – The average amount of news hits calculated per year per life-cycle phase using the timeline pattern from Google News.

Ranking

Based on these results projected above we can say something about in which phase news databases show most activity. The subsequent table gives an overview of previous results, but now categorised by ranking. The ranking of the absolute hits from the previous table is shown below in Table 6.

Table of RANKING - Google News			
Breakthrough innovation	Average hits per year in Innovation phase	Average hits per year in Market Adaptation phase	Average hits per year in Market Stabilisation phase
MATERIALS			
Astroturf	1	2	3
Kevlar	1	2	3
Dyneema	1	2	3
Nylon	2	1	3
Teflon	2	1	3
PET (polyester)	1	2	3

PHARMA & HEALTHCARE			
X-ray	3	2	2
Aspirin	1	3	2
Paracetamol	1	2	3
Contraceptive pill	1	2	3
Viagra	1	2	3
Prozac	3	2	1
Polio vaccine	1	3	2
Minoxidil	2	2	3
Total	**21**	**28**	**37**
Mean rank	**1,50**	**2,00**	**2,64**

Table 6 – Each phase of the technology life-cycle ranked towards the amount of news hits.

Based on both tables projected above, we see some exceptional cases, for instance X-ray. In this case most news publications are found in the innovation phase. However, this technology has a very short developing time, because the second and third hallmark occurred in the same year. Therefore the amount of hits is equal for both the adaptation and market stabilisation phase. Still, Table 6 indicates that news databases show the highest activity in the market stabilisation phase than in the prior phases due to the fact that they represent relatively more news hits in this phase. After ranking the technologies and their associated publications we see that the innovation phase has the least amount of hits in general, and the market stabilisation has the most hits in general. The adaptation phase is in the middle of both.

From observation we also indicate that the pharmaceutical industry shows more cases where the highest activity occurs in the market adaptation phase, while in the material industry the highest activity is always found in the market stabilisation phase. We tend to believe this phenomenon occurs due to the fact that once they are launched in the market they are fully tested and approved by the FDA. Therefore, consumers do not regard this technology as a niche, rather an honest and reliable product which they are willing to pay for. The adaptation phase is relatively short to the innovation phase which results in a relatively high average of hits. For materials the applications are often much broader and therefore the adaptation phase is often longer as well. As a result, the highest activity is visible in the market stabilisation phase.

Statistical analysis

The ranking done above already gave a clear view of when activity is highest and the increase over time. However, with the non-parametric Friedman two-way analysis of variance by ranks it becomes possible to test the null hypothesis (the hypothesis that the average ranking is equal to all phases), in order to check whether the difference in ranking can be regarded significant. This statistical analysis is useful because the data from *k* matched samples (number of phases in the sample) are from an ordinal scale. The Friedman test determines whether it is likely that the different columns (in this case the ranks of the samples) come from the same population (Siegel 1988). The null hypothesis can be described as:

H_0: Each phase from the technology life-cycle has the same average ranking.

In that scenario the distribution of ranks in *each* column would be a matter of chance, and thus it would be expected the ranks 1, 2, and 3 would appear in all columns with about equal frequency. Also, if H_0 is false, then the rank totals (see Table 6) would vary from one column to another.

In this test *k* represents the ranks (*k* = 3); N represents the amount of cases (N = 14). After executing the test, the mean ranks are: innovation phase 1,50; market adaptation phase 2,00; market stabilisation phase 2,64 (also shown in the Table 6). This already indicates that hits grow over each subsequent phase. The Friedman test accepts the null hypothesis at a significance level of p = 0,008. Since we use a significance level of α = 0,05, in this research, we can reject the null hypothesis with this data (p = 0,008 < α = 0,05). Based on these findings we can conclude that all samples do not come from the same population. From the average ranks, it can be observed that news activity increases over the phases of the technology life cycle.

3.4 Scientific databases

This paragraph emphasizes on scientific databases and when they appear to show the highest amount of activity (i.e. amount of hits) over the technology life-cycle. The used scientific database is Scopus which is owned by Elsevier (main international publishers of scientific journals)[14]. Scopus is one of the largest abstract and citation databases of peer-reviewed literature and quality web sources with smart track tools[15]. Also, it clearly visualises the amount of hits per year and therefore it becomes possible to clearly do this analysis. More information on Scopus is given in paragraph 2.2.1. Table 7 overviews the average amount of scientific hits calculated per year over the technology life-cycle collected from Scopus.

Absolute average number of hits found in Scopus			
Breakthrough innovation	Average hits per year in Innovation phase	Average hits per year in Market Adaptation phase	Average hits per year in Market Stabilisation phase
MATERIALS			
Astroturf	0	0	1
Kevlar	3	0	32
Dyneema	0	1	7
Nylon	0	0	2
Teflon	0	1	5
PET (polyester)	0	2	4
PHARMA & HEALTHCARE			
X-ray	3	30	30
Aspirin	0	0	0
Paracetamol	0	0	1
Contraceptive pill	0	0	0
Viagra	5	245	260
Prozac	0	0	1
Polio vaccine	2	10	7
Minoxidil	0	45	180

Table 7 – The average amount of scientific hits per year over the technology life-cycle.

[14] *http://en.wikipedia.org/wiki/Scopus*

[15] *http://www.elsevier.com/wps/find/electronicproductdescription.cws_home/704746/description#description*

Ranking

Based on these results projected above we can say something about in which phase scientific databases show most activity. The subsequent table gives an overview of the same table, but categorised each phase in which we find most publications. The ranking of the absolute hits from the previous table is shown below in Table 8.

Table of RANKING - Scientific database Scopus			
Breakthrough innovation	Average hits per year in Innovation phase	Average hits per year in Market Adaptation phase	Average hits per year in Market Stabilisation phase
MATERIALS			
Astroturf	2	2	3
Kevlar	2	1	3
Dyneema	1	2	3
Nylon	2	2	3
Teflon	1	2	3
PET	1	2	3
PHARMA & HEALTHCARE			
X-ray	2	3	3
Aspirin	1	1	1
Paracetamol	2	2	3
Contraceptive pill	1	1	1
Viagra	1	2	3
Prozac	2	2	3
Polio vaccine	1	3	2
Minoxidil	1	2	3
Total	**20**	**27**	**37**
Mean rank	**1,43**	**1,93**	**2,64**

Table 8 – Each phase of the technology life-cycle ranked towards the amount of scientific hits.

Table 8 indicates that scientific databases show most activity in the market stabilisation phase due to the fact that the average amount of hits is found in this last phase of the technology life-cycle. This seems to be a remarkable result, because it is believed by some scientists that science is most turbulent during the first developments of the life-cycle. However, based on this table it seems that scientific databases find most results in the market stabilisation phase.

What also can be inferred from this table is that the market adaptation phase also gains more average hits per year than the innovation phase. Basically, we see the same result as in the case of news databases; however, there it was expected that this result would occur. In this case it was expected that hits would diminish after the technology life-cycle continues.

Statistical analysis

The ranking done above already gave clear view of when activity is highest. However, with the non-parametric Friedman two-way analysis of variance by ranks it becomes possible to test the null hypothesis (the hypothesis that the average ranking is equal to all phases), in order to check whether the difference in ranking can be regarded significant. This statistical analysis is useful because the data from k matched samples (number of phases in the sample) are from an ordinal scale. The

Friedman test determines whether it is likely that the different columns (in this case the ranks of the samples) come from the same population (Siegel 1988). The null hypothesis can be described as:

H_0: Each phase from the technology life-cycle has the same average ranking.

In that scenario the distribution of ranks in *each* column would be a matter of chance, and thus it would be expected the ranks 1, 2, and 3 would appear in all columns with about equal frequency. Also, if H_0 is false, then the rank totals (see Table 8) would vary from one column to another.

In this test *k* represents the ranks (*k* = 3); N represents the amount of cases (N = 14). After executing the test, the mean ranks are: innovation phase 1,43; market adaptation phase 1,93; market stabilisation phase 2,64 (also shown in the Table 8). This already indicates that hits grow over each subsequent phase. The Friedman test accepts the null hypothesis at a significance level of p = 0,00. Since we use a significance level of α = 0,05, in this research, we can reject the null hypothesis with this data (p = 0,00 < α = 0,05). Based on these findings we can conclude that all samples do not come from the same population. From the average ranks, it can be observed that scientific activity increases over the phases of the technology life cycle.

3.5 Relation between science and market activity

The mean ranks calculated for Google News and Scopus can be analysed by calculating the correlation between both. This will be done in order to say something about the differences. Below in Table 9 the mean ranks are overviewed first.

Technology life-cycle phase	Mean rank Scopus (x)	Mean rank Google News (y)
Innovation phase	1,50	1,43
Market adaptation phase	2,00	1,93
Market stabilisation phase	2,64	2,64

Table 9 – Overview of the mean ranks from Scopus and Google News.

From the data illustrated above we can calculate the correlation (*r*) between both databases using the following formula from (Siegel 1988), page 236:

$$r = \frac{\sum \mathrm{XY}}{\sqrt{\sum \mathrm{X}^2 \sum \mathrm{Y}^2}}$$

Using this formula results in *r* = 0,999794. This indicates that the correlation of the mean ranks between both databases is very strong, since it is almost equal to 1.

Besides calculating the correlation, it is also possible to calculate ratios. Appendix C gives an overview of the ratios between the scientific and news hits. It is interesting to observe from this overview that the ratio of scientific hits to news hits gets higher over time. Thus, overall scientific hits will grow faster than news hits. Although this does not necessarily mean that the interest for the technology in the scientific community grows relatively faster, it may be of interest to take this difference in publication behaviour into account if news hits and scientific hits were to be used in a business intelligence tool.

3.6 Conclusions and limitations

Both analyses on scientific and news activity shown in the previous paragraphs both give an answer to our hypotheses. We are now able to assess whether the hypotheses as they are mentioned in chapter 1 are confirmed or rejected. As has been explained in the beginning of this chapter there are basically two viewpoints in literature on this topic. The first believes that sequential steps occur from basic-, applied research, patent activities and market activities (Bush, Myers and Marquis, Utterback, Watts). The second viewpoint believes that these activities occur in a more parallel way (Schmoch, Grupp, Noyons, Lacasa, and Rosenberg); though there are discussions in what kind of pattern they emerge.

Our analysis clearly shows that both databases create a pattern of increased publications. Table 6 and Table 8 show after ranking the average amount of hits per life-cycle phase that activity in general raises parallel to the maturity of the technology. Thus, the amount of hits increases over the evolution of the technology life-cycle. Looking into detail we do see some exceptions, but in general the market stabilisation phase shows far more activity than the prior phases. Also, the adaptation phase is more active than the innovation phase when we compare the results. It is a remarkable result, because many scientists (Myers and Marquis 1969; Utterback 1974; Watts and Porter 1997; Trumbach, Payne et al. 2006) believe scientific activity is most present in the beginning of the technology life-cycle, however, this was not the case based on the results of the analysis.

As a result we believe the second viewpoint in which scientific, patent and market activity occur in parallel steps. The difference measured in activity is regarded significant after the ranking was tested with a statistical analysis. Based on our results we cannot say in further detail how the patterns develop (i.e. in double-boom cycles, linear or exponential). Also, most activity in scientific and news databases emerge in the market stabilisation phase. However, it has to be said that certain limitations were present in this analysis. First, the analysis only included 14 cases from two industries, which is only a small sample size. Second, the query consisted of only one technological search term. Discussion about whether the technological name is a correct search term, due to the fact that scientific terms of brand names are not used in this analysis. Third, there were some exceptional cases in this analysis like for instance X-ray, which shows a very short innovation and adaptation phase. Fourth, it should be included that secrecy issues might have played a role in this research due to the fact that science is related to competition and therefore secrecy on new developments might not always be published. Further research could give a more clear answer on this topic.

A final possible limitation should be explained separately from the rest. Although increased activity occurs over time because the databases find an increasing amount of hits, this does not necessarily mean that the attention is indeed increasing. It should not be forgotten that more sources are included over time by the databases, since for instance Google News is improving their database every day by including more sources. Some of the greatest newspapers started publishing at a certain time, while the innovation phase of a technology might for instance have been passed already. Thus, it might be the case that the amount of sources increases as well, resulting in more hits over time and thus more activity, without an actual increase of attention. Based on this information there are two scenarios: 1.) there is a true increase in attention or 2.) there is no actual increase in attention, but just more sources are just taken into account over time. The question emerges which scenario is occurring. An answer to this question is not easy to give, but we believe that this measuring error is present but relatively small and thus an actual increase is occurring. Reasons supporting this vision are the large time span we collected data from. Breakthrough technologies of 200 years ago and modern technologies are included as well. Unfortunately, there is no information from Google News or from Scopus findable about this topic. Only by looking at the full list of sources and by analyzing at what moment each source started publishing would eliminate this measurement error. There is a limitation present in this analysis, but we cannot say much about the level of significance. Therefore, it is assumable that there is no actual increase in activity, because the increase in activity is not an increase with a fixed amount of sources. Still, we believe it did not disturb our conclusions, because the sample size was collected from a very large time span. Also, many newspapers already started publishing before these breakthrough technologies were even invented.

To conclude about this limitation, there is a measurement error present since the amount of sources within the databases increases. Although the error is present, we believe that it does not disturb the conclusions significantly; however we need to take into account this phenomenon and its effects. However, due to the lack of information about this topic and the missing explanation from Google and Scopus we cannot say much about this limitation. A full analysis on the sources and the moment they started publishing should give more clarity.

That said, we can conclude the following on the hypotheses:

1. *Scientific databases show most activity and collect most publications in the beginning of the development of breakthrough technologies (i.e. around the innovation phase of the technology life-cycle).*
2. *News databases show most activity and collect most publications at the end of the development of breakthrough technologies (i.e. around the market stabilisation phase of the technology life-cycle).*

Based on our results, we are able to say the following. The first hypothesis is rejected, since in this analysis most scientific activity is clearly found in the last phases of the technology life-cycle.

The second hypothesis is confirmed, since in this analysis business press and news activity is clearly found in the last phases of the technology life-cycle.

4. APPROACH FOR ANALYSIS

This chapter will focus on the methodology that will be used in the second analysis of this thesis which will emphasize in on the correlation between the pattern of publications and the technology life-cycle pattern. It will begin with a selection of possible methods for analysis. Subsequently, one method will be chosen. Additionally, emphasis on the databases, the query, the data from the innovations, and the reliability of the chosen method will be included before the actual analysis will be performed in chapter 5.

4.1 Methodology of analysis

Multiple methods are applicable for determining a possible correlation between the amount of publications and the hallmarks of the technology life-cycle of a breakthrough technology. An overview of different methods is given first, subsequently; the most appropriate method will be chosen which fits this explorative research best. This is necessary due to the fact that those different methods each have their own pros and cons. Also, future research might use a different method which would be more appropriate. In general, these methods are of quantitative nature. However, there seems to be an overlap, because the qualitative information will be transformed into quantitative information.

4.1.1 Different methods for analysis

Four methods will be proposed and elaborated in this section. The methods are:

A. Slope of the publication pattern
B. Peak of the publication pattern
C. Technology life-cycle pattern of the publications
D. Full history analysis

A. Slope of the pattern

The first proposed method is about determining the slope of the pattern of publications. The slope of the curve increases once the amount of hits increases as well. When the slope is negative, the curve is going downwards. The slope is calculated by dividing the difference in the amount of publications by the difference over a time period. In mathematical terms: dy/dx. When the difference over time (dx) is kept constant (e.g. one year) for every calculation, it becomes possible to compare the slopes and determine when the slope is steepest. The steepest slope is expected to be most useful to predict the hallmark, because it represents the largest growth in publications. It is expected that the large growth of publications can be related to the awareness and interest of this new technology by consumers. When the (future) consumers are interested in this latest technology, they will consider buying this technology. Making the consumer aware of the latest innovations will, among others, be done by publications in news articles and magazines. Therefore, it is expected that a highest growth of publications (steepest slope) can be related to a growth in sales and thus for predicting the hallmarks of the life-cycle. According to Watts and Porter, a clean pattern will show a rise first, and then decline in fundamental research. Afterwards, a similar lagged pattern emerges in a more applied research database. Last, the pattern will show evidence of development, applications and possibly impact in the market (Watts and Porter 1997).

The slope of the pattern is not equivalent to the peak of the pattern, which might make it more useful than just looking at the peak of the pattern. The corresponding date of this slope tells that in this year the growth in publications is largest and this date can therefore be compared with the hallmark date of the breakthrough technology. Figure 14 shows how this looks like.

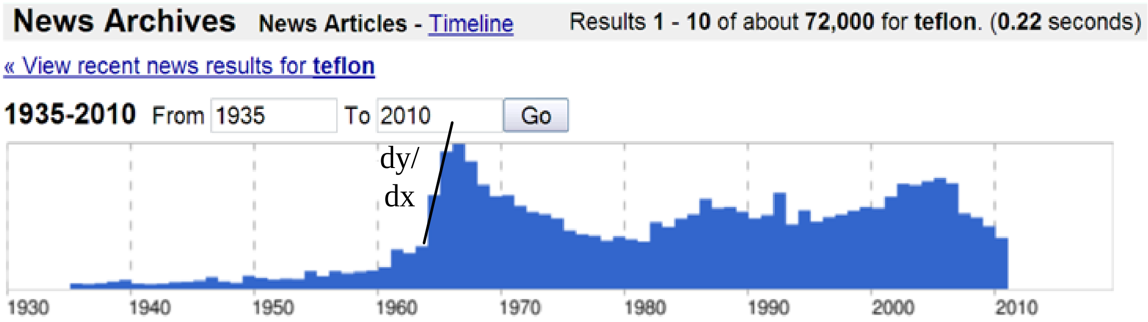

Figure 14 – Calculation of the slope of the graph.

The downside of this method is that all quantitative data needs to be collected manually, because for each year the amount of hits is shown separately in Google. These amounts of hits per year (or month) need to be collected for every query and for every technology. Also, sometimes there will be multiple peaks present with approximately the same slope. It will be hard to determine which slope corresponds with the hallmark. On the other hand, it becomes possible to compare the results of different technologies together, which can be beneficial.

B. Peak of the pattern

The second possible method is to measure the difference between the date from the peak of publications and the date of the hallmark. This difference is measured in years. With this method it becomes possible to say something about the deviation between the hallmark and the peak in publications. Figure 15 shows how this looks like. By comparing both peaks it might become possible to see a constant deviation which occurs at most technologies. It is expected that this constant deviation is happening due to a *time lag*. Often there is a time lag between the emerging technology and the moment that news articles about the technology are published. Again, the main focus lies on the peak of publications, because it is expected that more news is published around the date of the hallmark. This expectation can be supported by the fact that more people are interested in the emerging technology.

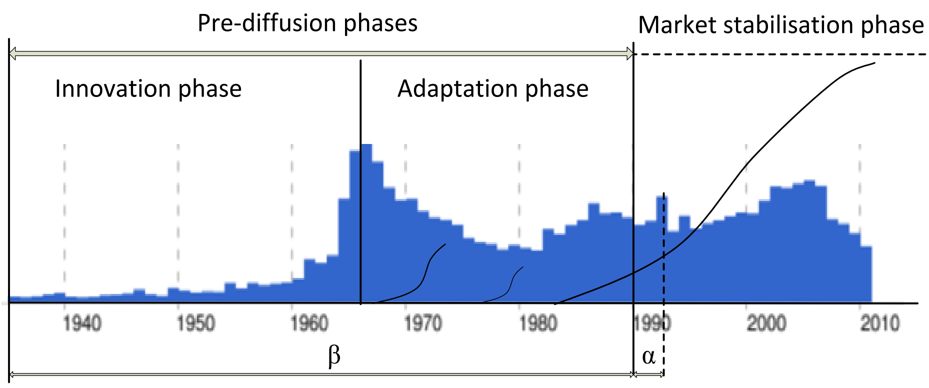

Figure 15 – An example how the peak of publications is compared with the hallmark.

The deviation between the peak of publications and the hallmark of the life-cycle can be expressed in years, but also in a percentage of the total length of the technology life-cycle (in Figure 15 shown as α). In this way it becomes possible to compare different technologies and their average deviations. The difficulty with this method is to determine the exact peak in publications. Often, there is no clear peak or there are multiple peaks in the pattern of publications. The peak that correlates to the hallmark should be determined by filtering out the other peaks which are also related to the hallmark, but with a lesser extent. However, this task of filtering and looking at multiple peaks is time consuming and might not be very reliable. There are two alternatives to solve this problem:

- Choose the peak which lies closest to the date of the hallmark within certain boundaries (e.g. 5 years). As a result, a correlation might be found between the peak and the hallmark. Having this information can be helpful for predictions, because then it is clear that a sudden high peak in publications will have as a result a high probability that the hallmark will occur or start.
- Conduct a qualitative analysis first which emphasizes on the complete history of the technology and why these peaks occur in the first place. By doing so, it becomes possible to distinguish the “real” peaks from the “fake” peaks. The “real” peak can then be used in this analysis. Unfortunately, digging into the history of each technology seems like an impossible task within the time boundaries of this research. However, it would be of great added value in this explorative research.

By choosing the first method, less time will be consumed for seeking the real peak that corresponds with the hallmark due to the fact that the focus will only be around the dates of the hallmarks and not on the rest of the technology life-cycle. This is beneficial when more search combinations will be used in Google News. All results can easily be interpreted into an overview. Table 10 shows how this would look like and the numbers are just an example. Thus, the date of each peak of a technology can be compared to the date of the hallmark.

The table shows the deviation in years between the peak in publications and the hallmark expressed in time. The plus- and minus symbol represents whether it concerns a positive or negative deviation. Since the technology life-cycles vary per case, the deviation can be calculated and expressed into a percentage from the total length of the technology life-cycle. By doing so, a possible correlation might be found when this data will be compared. Also, the average deviation for every hallmark can be calculated now. The average deviation possibly shows that the amount of publications increases only after the hallmark and thus it becomes possible to detect a possible time lag.

Technology search terms	Invention hallmark		Market introduction hallmark		Large-scale production and diffusion hallmark	
e.g.	# hits (of peak)	Deviation (in years)	# hits (of peak)	Deviation (in years)	# hits (of peak)	Deviation (in years)
X-ray	12800	+1 (1%)	18500	+2 (2%)	19400	-1 (1%)
Nylon	600	+3 (6%)	4060	0 (0%)	14200	-2 (4%)
Dyneema	1400	-4 (12%)	2000	+2 (6%)	4380	0 (0%)
…						

Table 10 – Example that shows the deviations between the hallmarks and the publications peaks.

C. Technology life-cycle based on publications

A third method to find a possible correlation is by developing a pattern of the publications. This pattern can be distinguished into phases which are separated by a "separator". The separator is not the same as the hallmark of the technology life-cycle, rather a fictive hallmark, which distinguishes the different phases in growth. The *cumulative amount of publications* can be expressed in a graph. When this is done correctly, it becomes even more easily possible to distinguish this graph in small phases based on the degree of growth. Instead of looking only at the separators, this method focuses more on the phases. By looking at the growth and whether this growth is high or low over time, a comparison can be made with the original life-cycle pattern and the associated phases. Figure 16 shows how this looks like and is based on the data collected for Astroturf.

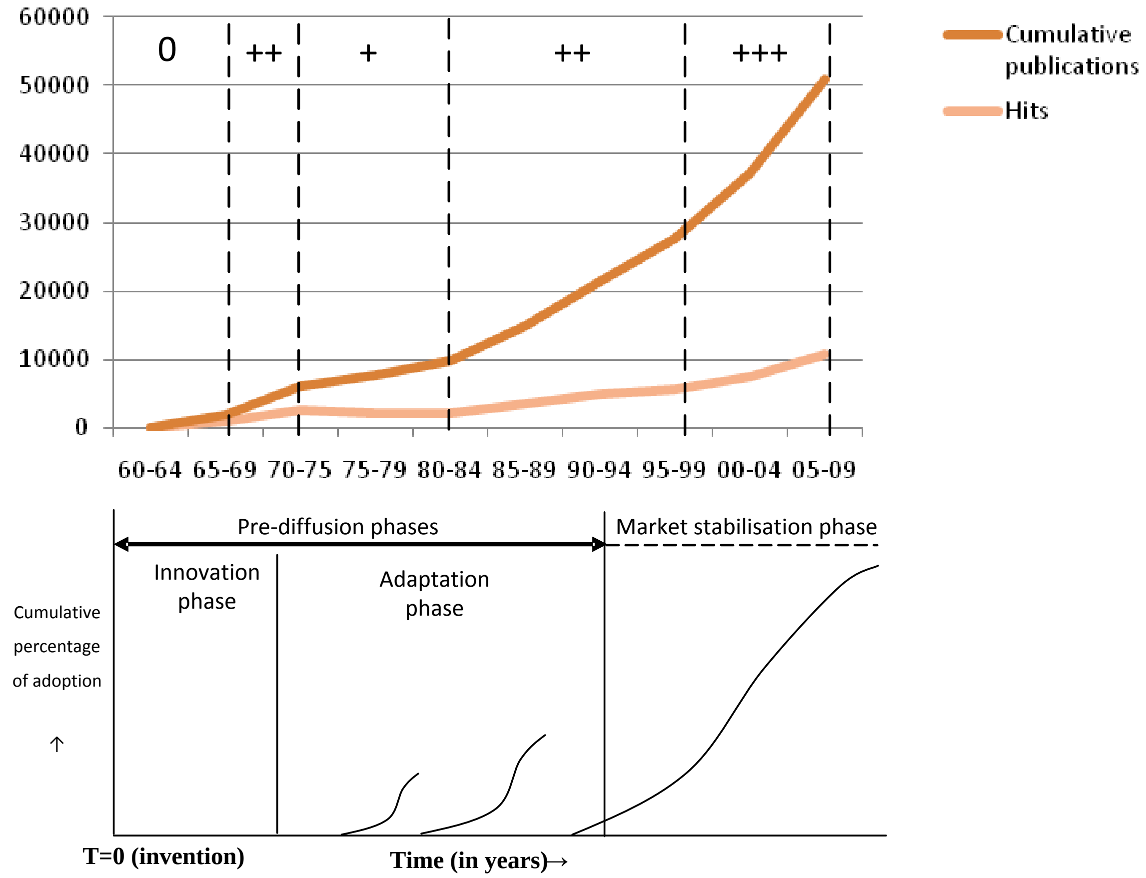

Figure 16 – A conceptual overview of how to split the publications into different phases.

Once the phases are determined it becomes possible to see whether the phases find a correlation with the phases of the technology life-cycle and whether the hallmarks of the technology life-cycle occur in a phase of low, medium or high growth. Figure 16 shows as an example that the hallmarks occur at the phases of increased publications (++).

There are many limitations though with this method. First, developing a pattern based on the cumulative amount of publications is very labour intensive. Second, it is doubtful whether the "separators" (i.e. the vertical lines that distinguish the phases) are really that clear to be found. The patterns might show a much more dynamic graph, which will result in an unusable undefined pattern. On the other hand, if clear separators emerge in the graph than we can simply compare them with the original technology life-cycle and their hallmarks, but this will not always be the case. Third, it is impossible to determine boundaries which explain whether the phase is allocated to a low, medium or high growing phase, because each case is completely different. Only on a visual basis it becomes possible to determine to what extend the growth is occurring, however it is hard to relate clear criteria to this growth factor. Last, finding a fit between both patterns seems almost impossible.

D. History analysis

A last proposed method for finding a correlation between both patterns is digging into the history of each technology and compare those with the collected data from Google News. This method is a full qualitative analysis which focuses on each particular year from both the data from Google and the history of the technology. Not only the peaks and slopes of the data are analysed, but also the history of the technology is analysed. As a result, this method looks from both perspectives and can greatly enhance the perception why certain peaks occur, without relating to the hallmark of the technology life-cycle. Figure 17 tries to support this method in a visual manner.

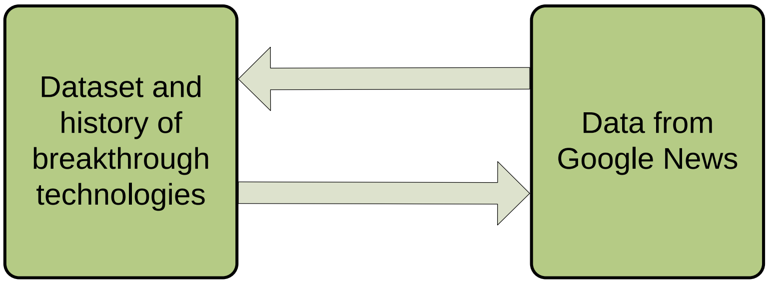

Figure 17 – A visual expression of the history method.

The main advantage of this method is the fact that it fits best in this explorative research, because it also includes the history of each breakthrough technology. The other methodologies did not include the history of the technologies. As a result, it becomes possible to say something about why certain peaks occur in the news data. As an example, you can image that once a medicine did not pass the FDA (Food and Drugs Administration), it might gain more publications in news articles, though it concerned negative publications. The data shows a clear peak in that particular year, however, it cannot be related to the hallmark of the technology life-cycle, but to the poor performance of the medicine. This example did not gain more publications due to the popularity of it, but gained more publications due to the fact that it is not safe for consumers to use. It may be clear that this method perfectly fits in this research, because these particular fluctuations in the pattern will be highlighted and noise will be filtered out. Also, future research (and the other methodologies explained before) could use these results, because the “real” peak is filtered out from the noise.

Moreover, the fact that this method creates perfect insight in the fluctuations of the pattern, it also becomes possible to say something about the usefulness of Google News. If it turns out that the hallmarks of the technology life-cycle reflect the main peaks in the publications it becomes possible to say something about the added value of Google News. One major limitation is the labour intensity which is caused by the analysis on the full history of each technology.

4.1.2 Selected method for analysis

This research will conduct the latter analysis that was proposed in the prior paragraph: *history analysis*. This method can be seen as a full qualitative research and will add much value due to the fact that is will focus on every peak, as well as the full history of the technology. In this way it becomes possible to:

- Reduce noise by explaining all fluctuations in the data;
- Distinguish the quality of the peaks;
- Explain the occurrence of each peak in history;
- Find the real added value of Google News for forecasting;
- See whether there is a correlation between both patterns;
- Fulfil the most effective research for this explorative research.

4.1.3 Criteria for analysis

This paragraph will shortly elaborate on the criteria we need to complete this analysis. Due to the fact that this analysis is of qualitative nature and not quantitative, it is hard to determine when we find a correlation between both patterns. For some cases Google finds thousands of hits per year, but other cases only show a very limited amount of hits. Therefore, we cannot simply determine criteria that indicate when we see a peak in publications. For instance, a peak can be indicated by a doubling of publications after a long steady constant period of publications. However, it is also possible that publications grow with a factor twenty in an already turbulent pattern; in that case we also indicate a peak.

Thereafter, we also look in detail to the history of each case and why certain years of the patterns show a turbulent time. By combining both analyses we are able to find possible correlation. In order to deal with this large diversity in our results, we simply analyse each case separately and do not set up any criteria which clarifies when we cross a peak in publications. Thus, analysing on a visual basis will be done in this research, without setting any criteria of when we see a correlation.

4.1.4 Reliability

The selected method for analysis creates some issues concerning the reliability. After a historical analysis is executed to reduce the noise and irrelevant fluctuations in the data we might find possible correlations between the patterns. However, due to the fact that we will do this analysis only on a visual basis it might gain issues concerning the reliability, because this analysis would only use the vision of the author. Therefore, the level of subjectivity is very high. A protocol is written to include the vision of other people as well. This protocol will explain the research procedure, the goal, the methodology in a brief way. The collected data and full analysis is included as well. With this protocol we want to support the *procedure* by increasing the validation. The interrelated reliability will not be tested with this method. Three students from the Master Management of Technology were requested to assess the analysis and the corresponding results in order to give their opinion about the procedure. After verification, the subjective procedure will gain strength due to this support. The full protocol is shown in Appendix D.

4.2 Data collection

Different data needs to be collected for this research. First, the technology life-cycle patterns of breakthrough technologies and the associated hallmarks need to be found. Second, the data on the news publications need to be found. Both fields will be explained in this paragraph.

4.2.1 Technology life-cycle patterns

The main priority of this research is to determine a possible correlation between the news and press publications and the technology life-cycle of breakthrough technologies. Data that will be collected need to be appropriate for this research, but also reliable and valid.

It starts with defining the hallmarks of the breakthrough technologies. This data has been collected, checked and verified by Dr. J.R. Ortt from the Technical University of Delft. His data is used in this thesis project. Although the database of hallmarks from breakthrough technologies is divided in multiple industries, this thesis project focuses on only two industries*: materials and pharmaceutical*. These industries are deliberately chosen based for several reasons. First, both industries show totally different technology life-cycles which make them interesting to involve in this research, because it becomes possible to find different results for the different industries. Second, due to limited time in this research it would become impossible to involve all industries and all the technologies of the database (around 110 technologies). In the material and pharmaceutical industry the applications are clear and specific and thus easier to use in this research to limit the amount of noise. For instance, we can now fully focus on technologies with only one or two applications (e.g. Viagra, Dyneema, or Paracetamol). These technologies have been developed mainly with one goal (e.g. cure people). Third, these two industries show hallmarks which can be found in Google News. As a result, these industries are most useful due to the fact that the related data can be used for finding a correlation (there are certain technologies which hardly find any result in Google News). Last, these chosen breakthrough technologies have shown the *clearest hallmarks* which makes them most reliable for using them in this research.

Below in Table 11 an overview is given of the chosen technologies and the associated hallmarks. The table shows three columns of data which represent the hallmarks of the technology life-cycle.

Breakthrough innovation	Invention	Market introduction	Industrial production and large-scale diffusion
MATERIALS			
Astroturf	1964	1965	1968
Kevlar	1965	1971	1973
Dyneema	1964	1975	1990
Nylon	1934	1937	1940
Teflon	1938	1948	1950
PET	1941	1946	1953
PHARMA & HEALTHCARE			
X-ray	1895	1896	1896
Aspirin	1853	1899	1900
Paracetamol	1878	1853	1955
Contraceptive pill	1927	1928	1944
Viagra	1994	1998	1998
Prozac	1971	1982	1982

Polio vaccine	1952	1955	1955
Minoxidil	1960	1970	1979

Table 11 – Overview of the hallmarks from two industries; materials and pharmaceutical.

4.2.2 Bibliometrics from Google News

Data from bibliometrics on the amount of news publications need to be collected as well in order to complete this research. This data will be collected from Google's latest application; Google's business press and news search engine (Google News). This application offered by Google is fairly new and is improved regularly. Google News can be regarded as an overlapping term for a database which can be subdivided into *archives* and *news*. Basically, Google News Archives looks at news older than thirty days and that is archived in the database. The other subdivision called Google News looks at most recent news less than thirty days old. This research collects most data from Google News Archive, but in this thesis we will often refer to the overall term (Google News) meaning the same thing. However, both Archive and News can be regarded useful, depending on the query. The data collection for this research used Google News Archive due to the fact that the used cases are breakthrough technologies which emerged decades ago and are much older than thirty days. Figure 18 illustrates how Google News is subdivided as explained.

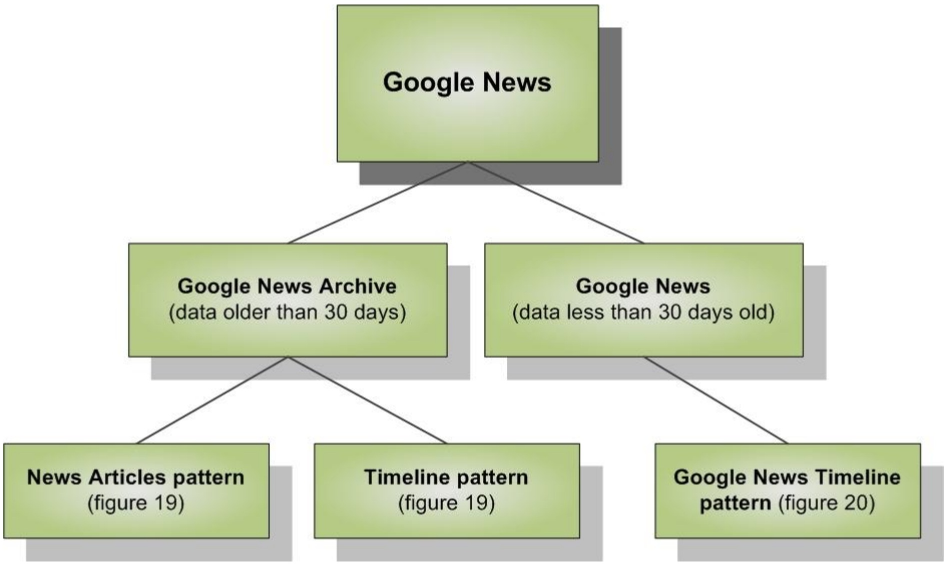

Figure 18 – Graphical overview of how Google News is subdivided.

Google News searches for business press and news articles which were published in history and are published today. Google collaborates with many different partners to digitalise historical news in order to let users find information published before the launch of the internet. Remarkably, Google offers their users to search even back to the 1800's, which immediately shows the capacity of this application.

Google News particularly seeks in news papers, magazines and historical archives from all over the world[16]. Users can search for articles, events, peoples or ideas and see how they have been described

[16] For a complete list of the hundreds of newspapers and magazines that are included in Google's News and sources that will be included in the future see this website: *http://www.google.com/support/news/bin/answer.py?answer=148418#rights*.

over time. Moreover, besides the relevant articles for the query, users can also see a historical overview of the results by browsing an automatically generated timeline.

Search results include content from a number of sources, some of them are freely accessible, others require a fee. Many historical archives are still completely offline. More information about the characteristics of this database is given in paragraph 2.2.1. Google claims that over time more news publications will be available online which will allow users everywhere to read even more of the history unfolded.[17] Also, the amount of news papers that will digitalise their unique collections will be increased with the help of Google. As a result, Google News will be improved every day and will even be more useful in the future. However, Google still explains the possibility of missing data or time period for a title due to some issues[18]:

- Google does not have the rights to display this content;
- Publication dates are missing or were not included;
- Low quality of the microfilm and issues with the scanning and conversion process;
- No search results for a specific query.

Results from the query and its limitations

The user of Google News Archive can search in selective details. The amount of hits can be given on different levels:

1. Total hits per life-cycle;
2. Hits per decade;
3. Hits per two years;
4. Hits per quarter of a year;
5. Hits per month.

As has been explained earlier, besides the quantitative data that Google News Archive offers, it also offers the same results in an automatically generated timeframe (see Figure 19). This graph immediately shows where the peak of publications is located in the pattern of publications. However, there is a limitation to this graph, because it does not represent the real quantitative data. The graph is just a visual representation of the amount of publications and is developed by computer algorithms. Therefore, this graph can only be used as an indicator in this research. The quantitative data can still be used effectively, because according to Google this information is reliable[19]. The quantitative data can be related to the amount of hits Google founds during the search. Google does not distinguish the quality of the hits, only the quantity. Figure 19 shows what is meant with the visualisation of the graph and the limitation it contains. The column of August is higher than September. However, according to the data, August contains 95 publications and September contains 105 publications. It is clear that the graph does not show this correctly.

---

[17] *http://news.google.com/archivesearch/help.html*

[18] *http://www.google.com/support/news/bin/answer.py?hl=en&answer=1005651*

[19] Contact with Google verified that the quantitative numbers presented in the upper bar of the screen (the circled part of Figure 19) are reliable. The visual pattern is based on computer algorithms and therefore less reliable.

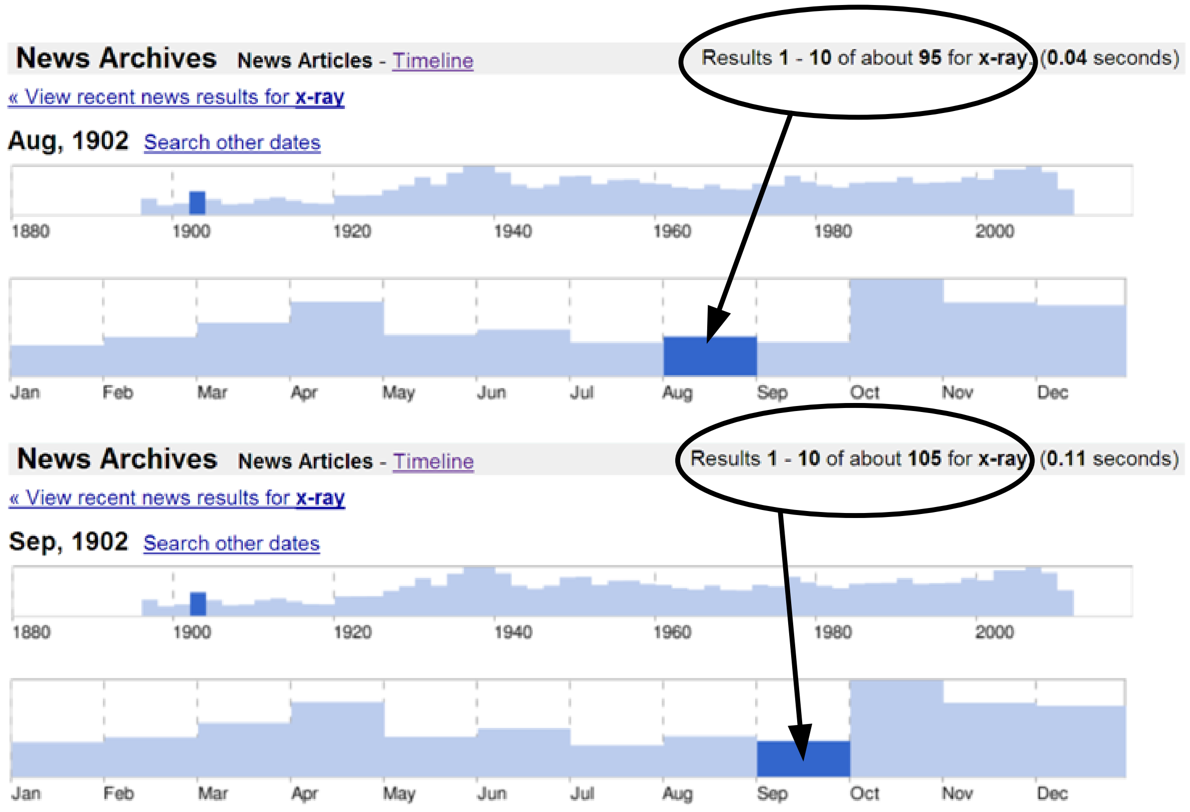

Figure 19 – Small deviation in the visual graph from Google and the quantitative data.

There is another remark which concerns these patterns of publications. Google News Archive offers two graphs for each query. The first graph shows the pattern of the news and business press publications ***(News Articles)***. The second graph shows the same results, but also includes "events" ***(Timeline)***. The timeline pattern seeks for web pages that describe events from the past, show charts and better snippets[20]. Figure 19 shows *News Articles*, but by clicking on link of *Timeline* it becomes possible to see this other pattern as well. An example of an event is, in our vision, for instance the introduction of a new technology by organizing a meeting where people can gather around and see this technology for themselves (i.e. demonstration). Especially in the days where internet and television were not present or invented yet, these events could be one of the limited possibilities to inform consumers about a new technology. However, an event can also be interpreted as compiling multiple articles which are talking about this one thing in the real world. Nevertheless, there is hardly any information available about this topic and how Google find its information. It is unfortunate for not knowing to what extend the addition of events is visible. Subsequently, it is unclear what these events involve precisely and how Google News finds them. So, the timeline pattern shows a higher level of dynamics and peaks, but Google does not mention anything what events clearly entail. Even after contacting Google the answer remained mysterious.

Still, the 'events' are not the most important difference. The timeline pattern has one major advantage over the news article pattern. It reorders the results, displaying in a graphical, chronological timeline. The timeline showed the distribution of publications over time, and by

[20] *http://googlesystem.blogspot.com/2007/12/google-news-archives-updated-timelines.html*

clicking at any point (date) on the timeline displays only the results from that point in time[21]. The news article pattern does this job in a less detailed way, resulting that the timeline patterns find more useful pattern for this analysis.

Both graphs show completely different patterns. During the first attempt to find a correlation between the life-cycle and the hallmarks, it became immediately clear that the timeline pattern creates far better results for finding a possible correlation with the technology life-cycle. This phenomenon did not only occur in one particular technology, but almost in all technologies. This can be related partly to the addition of 'events' (compiling multiple articles which are talking about this one thing in the real world), but probably more important due to the reordering of the results over a chronological timeframe. As a result it gains more reliable hits than News Articles only. In this research we will collect both patterns, however, during the analysis we will only use Timeline due to its great capacity. Further explanations supporting this vision will be given in the chapter 5.

An addition word has to be mentioned for more clarity in future research. As explained, Google News can be subdivided in archives and news. So far we have been talking about Google News Archive, since this research focuses on historical breakthrough technologies. However, if the result of this research turns out successfully and Google News will be used in forecasting tools, than Google News Archive has one major limitation. This limitation concerns the delay of thirty days, before the news will be included in Archive. The forecasting tool can neutralise this limitation by using Google News, focusing on the most recent data not older than thirty days. In that case, another timeline will be used, called Google News Timeline. Basically this application is equal to Archive, but has much more opportunities. It let the users of Google News Timeline select the sources carefully (i.e. newspapers and magazines) in which they are interested in. Selecting options such as seeking for blogs, magazines, quotes, or prizes become possible within Google News Timeline. Timeline can be regarded equal to Google News Archive; however it focuses on the latest news. Therefore, this application can be regarded more useful in most recent technologies, but cannot be used in this research due to the fact that the breakthrough technologies are selected from previous centuries. Still, a tiny explanation is given about Google News Timeline in order to clarify Google News Timeline. Figure 20 shows how this looks like. It still offers the ability to create a pattern and to measure attention among consumers; however the layout is different from Google News Archive. For clarity, all these applications can be regarded as a sub-application of Google News itself and therefore we will use this overlapping term mostly in this research. Figure 18 gives more clarity about how Google News is subdivided with their corresponding graphical patterns.

[21] *http://www.netforlawyers.com/content/google-removes-timeline-display-option-news-archive-results-clarifies-dates-assigned-results*

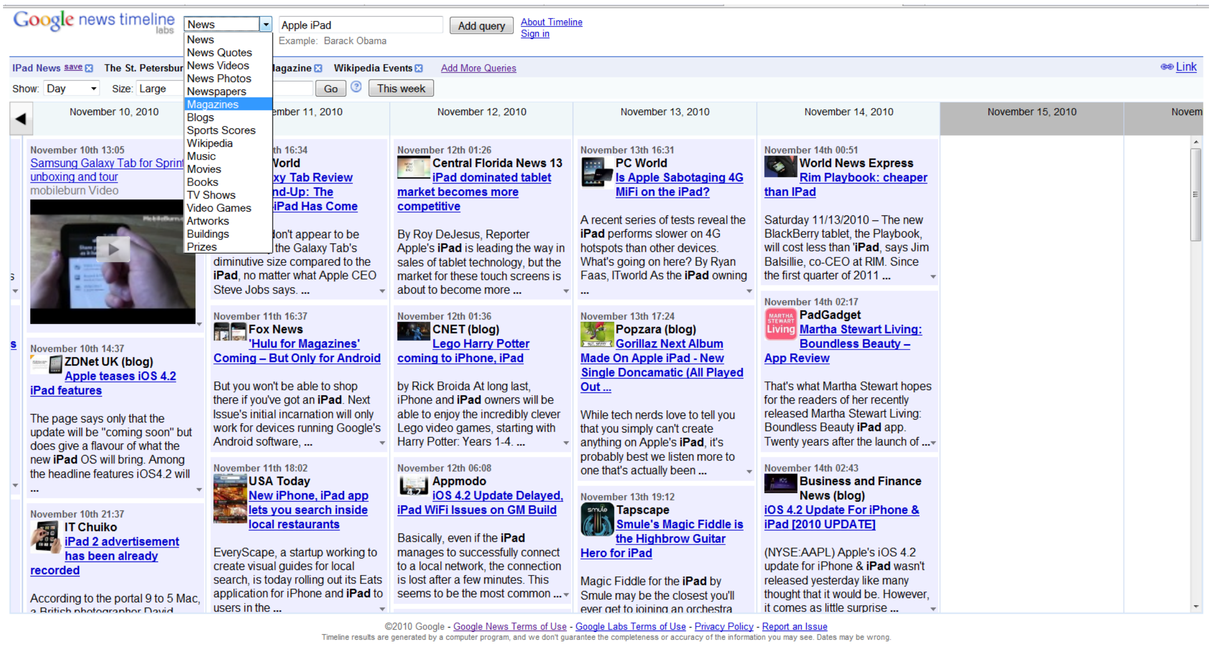

Figure 20 – Another sub-application from Google News: Google News Timeline offering equal results as Google News Archive only focusing on the most recent news facts.

4.3 Search combinations for the query

This research will not use word combination for each search, although it would be of major added value for future research. However, this paragraph will explain all the ins and outs, because it deserves some attention (for future research). We will explain the possibilities there are to make word combination in a query, why this would be beneficial and explain the further details.

4.3.1 The query in Google News

In the literature study it has been made clear that in order to define good indicators (which can be used in the query for combining search terms) it is vital that the query does not only use the term of the technology itself (e.g. X-ray). During the first developments and innovations of a technology there were certain keywords which were used. Often this was the molecular material or structure of a material. In the example of X-ray the first discovery was X-ray radiation. Only after this radiation was implemented into a device, it was called an X-ray scanner. As a result, it is necessary that during the query we also include the terms which represent the earliest translations before the final commercial name was given. This is of paramount importance to find useful results of the technology before it got its commercial name. Especially in the innovation phase, this approach will be useful.

Combining different search terms can be of major added value. The query can become more effective and more specific results might emerge due to the fact that search terms are more specified. Often, the breakthrough technology has alternative terms or names (e.g. brand names or molecular names). The publications will have a closer relation to the topic once alternative names are included in the query.

With the two industries, material and pharmaceutical, we have the advantage that the applications are limited. A medicine for instance, was mainly developed for one particular goal, which was to treat an illness or disease. The application therefore is clear and not too broad, although there are some examples of medicine that have multiple functions. Especially in electronic consumer goods, we often see that the applications are endless. In those cases it would be a lot harder to find a correlation between the publications and the hallmarks, because the breakthrough technology needs to be filtered out first before it can be used in the query.

Table 12 gives an overview of the breakthrough technologies and their sub characteristics (e.g. molecular structure, brand names) and how they can be used during the query in future research. An overview of the technologies and what they precisely entail is given in Appendix B - Qualitative analysis on news articles.

Breakthrough innovation	Terms for the query		
Astroturf	Artificial turf[22]	Chemstrand Company	Artificial grass
Kevlar	Para-aramid synthetic fiber[23]	DuPont	
Dyneema	UHMWPE[24]	UHMW	High-modulus polyethylene (HMPE)
	High-performance polyethylene (HPPE)	DSM	
Nylon	(generic family name) Synthetic polymers[25]	Polyamides	
Teflon	Polytetrafluoroethylene[26]	PTFE	
PET	Polyethylene terephthalate[27]	PET/ PETE	Trade names: Dacron, Diolen, Tergal, etc
X-ray	Röntgen radiation[28]	X-radiation	Medical X-rays
	X-ray tube	Anode	Radiography
Aspirin	Acetylsalicylic acid[29]	ASA	
Paracetamol	Acetaminophen[30] (used in U.S. Canada)	Para-acetylaminophenol	Para-acetylaminophenol
Contraceptive pill	-	-	-
Viagra	Sildenafil[31]	Pfizer (manufacturer)	Revatio (other brand name)
Prozac	Fluoxetine[32]		Prozac/Sarafem (other brand names)
Polio vaccine	-	-	-
Minoxidil	Minoxidil (off patent) nowadays)[33]	Loniten (first usement)	Rogaine/Regaine/Avacorand Mintop (other trade names)

Table 12 – Overview of the alternative names of the breakthrough technologie.

22 *http://www.madehow.com/Volume-7/Artificial-Turf.html*
23 *http://en.wikipedia.org/wiki/Kevlar*
24 *http://en.wikipedia.org/wiki/Dyneema*
25 *http://en.wikipedia.org/wiki/Nylon*
26 *http://en.wikipedia.org/wiki/Teflon*
27 *http://en.wikipedia.org/wiki/Polyethylene_terephthalate*
28 *http://en.wikipedia.org/wiki/X_ray*
29 *http://en.wikipedia.org/wiki/Asprin*
30 *http://en.wikipedia.org/wiki/Paracetamol*
31 *http://en.wikipedia.org/wiki/Sildenafil*
32 *http://en.wikipedia.org/wiki/Prozac*
33 *http://en.wikipedia.org/wiki/Minoxidil*

Table 12 already shows how many alternative search combinations can be made, but they do not even include the search terms defined in the literature study. It has already been explained that in this research we will not use word combinations for the query, because in this exploratory research it is simply too time consuming to manually collect all the generated data for these combinations. Though, for future research it would be of tremendous value. Therefore we have shortly elaborated on this topic.

4.3.2 Query design for future research

Although this research will not make use of word combinations for the query, we would like to further elaborate shortly on the possibilities which were offered for doing so for future research. Electronic databases have the capability to combine different search terms and to search more effectively. By using the correct search terms, the search creates more effective results. However, combinations might also lead to worse results and less hits when they are used improperly.

Google News can handle different combinations in order to improve the quality of the search. There are five operators (symbols) which make the search more effective[34]:

- **AND (+):** by using this operator in between two search terms, Google will search for both the first and the second search term combined (e.g. *X-ray + introduction*) and makes sure the results include common words that Google's search technology generally ignores;
- **OR:** by using this operator in between two search terms, Google will search for both the first and second search term separately (e.g. *X-ray OR Xray*);
- **MINUS (-):** by using this operator after another search term, Google will exclude all results that include this search term (e.g. *X-ray – advertisement*);
- **"…":** by using these operators for two search terms, Google will only return results that include this exact phrase (e.g. *"X-ray developments"*).
- **"site" operator:** by using this operator Google will find results from a specific site.

One remark concerns the efficiency of making large combinations. People often assume their search becomes more effectively by making large combinations of words. However, in most cases they do not. For instance, searching on X-ray, market introduction (second hallmark), and product quality, but also by filtering out possible advertisements will result in the following search combination:

X-ray + "market introduction" + "product quality" – advertisement.

As a result, the amount of hits will be very limited due to the fact that the query is too complex and too specific. A better combination would focus only on X-ray + "market introduction". In order to maintain the quality and a sufficient amount of hits to find the hallmarks in the patterns, we advise to limit the queries to a maximum of two search terms. By using two search terms as a maximum, Google News clearly shows enough useful hits for further analysis. However, if three search terms would be used in a combination the hits would diminish significantly.

[34] *http://news.google.com/archivesearch/refinesearch.html*

By setting this criterion of a maximum of two search terms per query for future research, we would be limited in making complex combinations. For instance, some breakthrough technologies which will be analysed in this research will have multiple names. An example is X-ray, which is sometimes referred as X-ray and sometimes as Xray. Also, in some cases a breakthrough technology is famous only with a commercial (brand) name. However, especially in the innovation phase these breakthrough technologies had no brand name yet and were often referred to their molecular structure or their special feature. Therefore, the name of the technology transformed over its life-cycle. It is hard to determine all the possible names of these breakthrough technologies and which ones are used most often in news articles.

To finish, query design is quite important for an effective search. It is of tremendous value if future research includes this topic in combination with additional search terms. The basics have been explained in this paragraph, relating towards Google News. Porter and Cunningham (2005) have further elaborated about this topic. It is advisable that more emphasize is given to this topic in order to create more effective queries in future research. Issues on this topic can then be highlighted and analysed.

5. RESEARCH ANALYSIS

This chapter elaborates on the collected data on the patterns of publications in order to find an answer to the research questions. The focus lies on both news publications and scientific publications. Both will be analysed in this chapter. The procedure of this analysis is subjective and in order to strengthen the validity three students from the Master study Management of Technology have executed a protocol explained in Appendix D.

5.1 Patterns of scientific publications

It has been explained in the literature search that a complete business intelligence tool includes multiple sources to make reliable forecasts about the development of breakthrough technologies. This is partly relevant due to the developments of new technologies and their different stage they are currently in. Scientific literature, for instance, would according to some scientists especially focus on science and new developments in the beginning of the life-cycle, while (Google's) News database(s) could tell much more about the popularity among future consumers (Myers and Marquis 1969; Utterback 1974). However, from our analyses performed earlier in this thesis, we have seen that scientific and market activity *increases* over time, in accordance with other researchers (Schmoch, Grupp). By including multiple sources it becomes possible to make more efficient forecasts. Although the focus in this research lies on the patterns of Google News, we wanted to include an analysis on scientific literature databases as well in order to determine whether these databases would be useful as well for a business intelligence tool and future research. If this would be the case than it could be considered to be included in future research.

The data that is collected by using scientific literature databases is shown in Appendix A. The used database is called "Scopus". Scopus offers a broad selection of scientific literature from different databases which makes is exceptional useful in this research. Moreover, it shows the amount of hits of scientific literature per year in a clear overview. Therefore, it can be used directly to determine the pattern of scientific literature published over time and there is no need to manually count the hits.

5.1.1 Analysis

The data collected from Scopus shows one remarkable result. It often has periods in time that it does not show any hits at all. In those years there were no scientific publications found. Also, initially it looks like there is no correlation at all between the patterns because there seems to be an absence in growth or large peaks around the moment of the hallmark. This makes a full historical analysis superfluous. A brief historical analysis only on the clear peaks will suffice, because there are simply only a very few peaks which can be related to in the pattern of publications.

Initially, there seems to be no correlation at all between the different patterns. However, there are some cases where there is a clear growth visible in scientific literature hits related to the hallmarks. Table 13, shows an overview of the cases and when a correlation was found and when no correlation was found. When a breakthrough innovation has a correlation with the hallmark due to a significant growth in literature compared to the years prior to this hallmark, we will mark this with a '+'. If there is no significant correlation (i.e. no growth), we will mark this with a '-'. The years in 'bold' are the hallmark years. A remark is about how a correlation is defined; this analysis is done on a qualitative way and no further boundaries are distinguished for this analysis. Thus, analysis is simply done only on a visual base as explained in 4.1.3 Criteria for analysis.

Breakthrough innovation	Hallmark 1 Invention	Hallmark 2 Market introduction	Hallmark 3 Large-scale prod./diff
Astroturf	-	-	-
Kevlar	-	-	+
Dyneema	-	-	-
Nylon	-	-	-
Teflon	-	-	-
PET	-	-	-
X-ray	-	+	+
Aspirin	-	-	-
Paracetamol	-	-	-
Contraceptive pill	-	-	-
Viagra	-	+	+
Prozac	-	-	-
Polio vaccine	-	-	+
Minoxidil	-	-	-

Table 13 – Overview of visible correlations between scientific literature and the hallmarks.

5.1.2 Conclusion

It can be concluded from Table 13 that only six correlations between the data from scientific database Scopus and the hallmarks are determined. The first hallmark (invention) does not show any correlation at all. Instead, the third hallmark (large-scale production and diffusion) shows most correlations with the breakthrough innovation, though the result is not significant since only 4 cases out of the 14 show a correlation (28 %). The second hallmark only shows 2 correlations out of the 14 cases (14%). We regard this result, 6 out of 42 cases, not significantly enough to believe that there is a correlation between both, because we tend to believe these correlations are just a coincidence due to the low chance.

Overall, scientific databases seem to be limited in the helpfulness. However, it should not be forgotten that the search terms of breakthrough technologies could be more specific (e.g. towards their molecular structure or brand name), which might result in more efficient results. Some of these breakthrough innovations are only known by their trade name instead of their molecular name. Especially in the beginning of the life-cycle, these trade names are often not very well defined yet. Consequently, it might say much more about the hallmarks. Future research could focus on more specific search terms and could also include multiple scientific databases. This analysis seems to have too little correlations to find it really useful for implementation into a business intelligence tool. However, we are able to answer our hypothesis as has been mentioned in chapter 1.

3. *There is a correlation (match) between the pattern (i.e. hallmarks) of technology life-cycles of breakthrough technologies and the pattern of publications from scientific databases.*

Bases on our findings, we tend to say that this hypothesis is rejected. We believe this because there is no correlation found for the first hallmark; only 2 correlations for the second hallmark (14%); and only 4 correlations for the third hallmark (28%) out of the 14 cases. This result is therefore significant low and therefore the hypothesis is rejected.

5.2 Patterns of news publications

The analysis on the patterns of news publications is done on a qualitative level due to the fact that the data shows different peaks and sometimes irregular patterns which can only be explained by comparing the data with the history of the technology. With this method (explained in paragraph 4.1.2 Selected method for analysis) it becomes possible to reflect to both sides of the data, thus the data collected from both Google News and the history of the data. By doing so, it becomes possible to explain different peaks and why they occur in the first place. Noise and irregularities are reduced by this method.

5.2.1 Analysis on a qualitative level

The data collected in July 2010 from Google News is given in Appendix B - Qualitative analysis on news articles. Both patterns from Google (i.e. “News Articles” and “Timeline” as has been explained in paragraph 4.2.2) are included in the tables. However, the analysis only focuses on the latter. This pattern is more complete due to the fact that it includes both news articles and is reordered over a timeframe. It gives much more effective results than the News Articles-pattern and is far more superior for this analysis.

Data which is shown in Appendix B - Qualitative analysis on news articles, is not collected for the full life-cycle, but only around the hallmarks dates. This is deliberately done in order to stick to relevant data related to the research questions. However, the data is selected with a maximum of ten years before the first hallmark and a maximum of ten years after the last hallmark. In some cases (e.g. in the case of Viagra), we have chosen for less years due to the very short technology life-cycle. In these cases we have chosen for the amount of years equal to the average length of the innovation or adaptation phase.

The appendix will further elaborate on each specific peak and each specific irregularity in the pattern, which will subsequently be compared to the history of the technology. The information of the history of the technologies is collected from a private and non-published database from Dr. Roland Ortt, who has verified each case. In this chapter we will only illustrate two exhibits of clear cases and show the results of the total analysis of each breakthrough invention in an overview and discuss the outcome.

5.2.2 Exhibits

Although the appendix elaborates on all the cases which are included in this research, we will explain two breakthrough technologies in this paragraph and why there is a correlation between the hallmarks and the pattern of publications. We will elaborate on two extreme cases. One case elaborates on peaks which clearly show a correlation with the hallmarks (Nylon) and another case which also explains a correlation in combination with other peaks and causes who interfere with this idea (Polio vaccine).

Nylon (clear case)

The complete analysis on Nylon is included in Appendix B - Qualitative analysis on news articles, but the highlights are explained here. Nylon is a synthetic plastic material composed of polyamides of high-molecular weight and usually manufactured as a fibre and can be regarded as a true breakthrough material. The applications of Nylon are quite broad, like clothing, construction materials, and sport materials. Table 14 shows the amount of publications collected from Google News over the length of the technology life-cycle. The years expressed in 'bold' are the hallmark years.

Year	News articles	Timeline
1930	31	156
1931	22	79
1932	31	49
1933	16	49
1934	20	71
1935	21	204
1936	34	77
1937	33	209
1938	59	321
1939	208	453
1940	1090	1540
1941	2100	2340
1942	2330	2500
1943	1670	1800
1944	1340	1400

Table 14 – Publication from Google News on Nylon over the technology life-cycle.

One of the reasons why the case of Nylon can be regarded as a clear example in this research is that it does not include peaks which cannot be explained by history. Therefore it represents a clear pattern that seems to correlate well to the hallmarks of the technology life-cycle.

The invention hallmark (1934) show a correlation with the amount of publications. Especially after the hallmark was crossed publication rose, because although in 1934 the invention hallmark occurred, one year later a variant of the invented nylon became Du Pont's most celebrated product. Therefore this hallmark has a sort of overlap between 1934 and 1935. However, publications clearly rose in these years. Then publications diminished and in 1937 when the second hallmark occurred the publication grew again. Again, we see a relation between both patterns. The largest growth occurs in the third hallmark in 1941 (and 1942) when Nylon gained a lot of popularity among consumers. Even after the year of the third hallmark publication rose. There are no other causes found in history which can be related to this peak in publications and therefore the hallmarks seem to be the clear cause in this case. This result is remarkable due to the fact that secrecy played a role during the development of Nylon. Still, we find a relation between the patterns, which might be even clearer when no secrecy issues would have played a role.

What makes this case useful is the fact that it shows clear peaks on the hallmark dates. Also, there is only a bit noise; indefinable peaks are hardly present in this case. Moreover, it clarifies that peaks in publications can occur also a bit later after the crossing of the hallmark of the life-cycle or when an event occurred due to a time lag (caused by the time needed to publish about a particular topic). In this case we see that the peak in publications is clearly present just after the crossing of the first hallmark. A time lag has an effect, which in some cases will result in a growth in hits in the subsequent year. This makes sense when for instance, a hallmarks or event occurs in December of a certain year. As a result, publications will rise in the year subsequent of the hallmark year.

Polio Vaccine (doubtful case)

The complete analysis of the vaccine is included in Appendix B - Qualitative analysis on news articles, but the highlights will be explained here. Polio is a virus that is found worldwide and is one of the most feared and studied diseases. It is highly infectious and it affects mainly children under three years old and it can result in paralysis. Table 15 shows the amount of publications over the length of the technology life-cycle.

Year	News Articles	Timeline
1947	13	32
1948	5	21
1949	12	25
1950	13	38
1951	15	42
1952	45	79
1953	184	270
1954	868	1040
1955	4960	5800
1956	1630	1690
1957	1300	1340
1958	535	657
1959	692	848
1960	540	675

Table 15 – Publication from Google News on Polio vaccine over the technology life-cycle.

Although there was a breakthrough development in 1948 on a possible vaccine, the publications do not show a clear growth here. In 1952, the polio vaccine was officially invented, as a result, the publications growth in that particular year and the subsequent year. Therefore, it seems that the amount of publications can be related to the hallmark of invention. However, this case is doubtful due to the fact that in 1952 and 1953 a major outbreak of polio occurred in the US. The increase in publications can therefore not be related to only the hallmark of invention, because the content of the publications is not included in this research. Still, a small growth in publications is found.

From 1952 more tests were done for three years and were regarded successful. Publications slowly rose as well in this period; however the real growth occurred in 1955. This is remarkable because this year is similar to the year of the second and third hallmark: market introduction and large-scale production and diffusion. A significant growth with a factor six is clearly visible with respect to the prior year. However, though it might seem that this peak can be related to the hallmarks, this might not be the case. Reason to believe this emerges from the fact that after 120,000 children were vaccinated of which 10 were killed and 200 were permanently paralyzed. It was regarded as one of the worst disasters of American history. Therefore, the peaks in publications can also be related to this phenomenon, instead of the crossing of the hallmarks.

What makes this example quite interesting is the fact that it shows a massive peak which cannot be related to the hallmarks only. It clarifies that a peak might be caused by a hallmark or by other factors. The content of each publication is not included in detail and therefore we cannot distinguish the reliability of such a “correlation”. In the hallmark of invention it seems like the peak in publications can be related to the hallmark; however, we also know that a major outbreak of polio occurred in 1952 and 1953. This phenomenon also occurs for the second and third hallmark when many children died after the vaccination programme. Therefore, we cannot conclude that this peak occurred with only a relation to the hallmark, but also due to other causes.

5.2.3 Causes for peaks

Peaks can occur due to multiple factors as has been explained in the previous paragraph. They might occur due to the crossing of a hallmark of the technology life-cycle; however other causes can be found as well (e.g. a major outbreak of polio in the same year of the first hallmark in the example of polio vaccine). In this paragraph we will try to explore the different causes for large peaks that cannot be explained by the crossing of the hallmark.

Though there was not always a clear cause that could explain a certain peak in publications, often there was a clear cause. These are the most common causes of peaks found in publications:

- New discoveries/improvements in the technology (Kevlar, Aspirin, Paracetamol)
- Successful tests (Aspirin)
- Patent applications (Teflon)
- Realisation of plants (Dyneema)
- Registration of a new trademark (Teflon)
- Introduction of new application of the technology (Teflon, Minoxidil)
- Emergence of a standard (PET)
- Negative physical effects on humans (bad publicity) (Paracetamol)
- Acceptance by FDA regulation and/or added to Pharmacopoeia (Minoxidil, Prozac)
- Governmental approval (Prozac)
- Removal from the market (Prozac)
- Second official introduction in the market (Prozac)
- Outbreak of a virus (Polio vaccine)
- Abandonment of a vaccine from the market (Polio vaccine)
- Severe accidents after treatment (Polio vaccine)

The causes mentioned above have resulted in a growth in publications in Google News. Some of these causes can be regarded as external due to the fact that they occurred by external factors (e.g. outbreak of a virus, emergence of a standard). However, there are also causes which can be related to the development of the technology (e.g. new discoveries, successful tests, patent applications). Basically, the causes can be categorized in these two fields, internal or external causes. However, internal factors might play a vital role for the (further) growth of a peak which ultimately supports the greatest peak that can be related to the hallmark. For instance, registration of a new trademark might result in a further growth because this standard (technology) gains more popularity among consumers before large-scale production and diffusion occurs.

5.2.4 Overview of results

As has been explained in prior paragraphs that peaks in publications can occur randomly, they also seem to occur due to the crossing of the hallmark. In this section we will give an overview of the amount of correlations we have found between the hallmarks and the peak in publications. Table 16 is developed after the qualitative analysis explained in Appendix B - Qualitative analysis on news articles.

Breakthrough innovation	Hallmark 1 Invention	Hallmark 2 Market introduction	Hallmark 3 Large-scale prod./diff
Astroturf	-	+	+
Kevlar	+	+	+[35]
Dyneema	+	+	-

[35] This correlation was found mainly in the "News Articles" pattern, not in the "Timeline" pattern.

Nylon	+	+	+
Teflon	+	-	-
PET	+	+	+
X-ray	+	+	+
Aspirin	+	+	+
Paracetamol	+	-	+
Contraceptive pill	-	-	-
Viagra	-	+	+
Prozac	-	-	-
Polio vaccine	+	+	+
Minoxidil	-	-	+

Table 16 – Overview of visible correlations between the news literature and the hallmarks.

5.2.5 Conclusions

In contrast to the analysis on scientific databases it seems like Google News is a much better database for finding a correlation with the technology life-cycle hallmarks of breakthrough technologies. Although there were some cases that showed a clear match, there was still enough evidence found to believe that the hallmarks can be related to the popularity of a new technology measured in news articles. Also, it became clear that the 'Timeline pattern' is far more efficient compared to the 'News Article pattern'. Therefore, the analysis focused mainly on the Timeline pattern.

In total there are 14 cases which each three hallmarks. Although some cases have both hallmarks (market introduction and large-scale production and diffusion) divided in the same year (e.g. X-ray in 1896), there are 42 hallmarks in total. After looking at each case and the history of developments it turns out that in *approximately half of the hallmarks there is a link between the amount of publications and the hallmarks year*. On these specific moments of correlation it became clear that publications increased when a hallmark was entering or was already crossed. Therefore, it is assumable that publications can be used as a method to predict the occurrence of a hallmark. However, the analysis also made clear that other peaks in publications emerged even though there was no hallmark present. These peaks occurred due to multiple reasons (e.g. failure of tests, new discoveries, new applications of the technology, when the FDA (did not) accepted new medicines). This makes it hard to distinguish why a peak emerges and to what they are related. Still, in some cases the peaks were undeniable closely related to the hallmark, which makes it a helpful tool to include Google News database. Also knowing that Google improves it database every day by including more sources (e.g. newspapers and magazines), the quality will improve as well in the future. Still, this analysis offers us the opportunity to answer our hypothesis:

4. *There is a correlation (match) between the pattern (i.e. hallmarks) of technology life-cycles of breakthrough technologies and the pattern of publications from news databases.*

This hypothesis is confirmed, since in approximately 50% of the cases we saw a clear correlation between both patterns. We tend to believe this result can be improved in future research, according to the methods explained in chapter 4.

6. BUSINESS INTELLIGENCE TOOL

In the previous chapter we have analysed both scientific databases and news databases and the correlation they appear to have towards the life-cycle of breakthrough technologies. The main reason for doing this research was to discover the possible added value of the bibliometrics of these databases when implemented in a business intelligence tool. This tool should be capable of making predictions, supporting managers in their daily activities and strategic decisions. Also, multiple sources should be used in a business intelligence tool in order to secure the reliability of these forecasts. When we take a look at the first discoveries on these databases we believe that it is indeed of major importance to not only rely on these databases. Scientific databases show in only a few cases a correlation, mainly in large-scale production and diffusion hallmark. However, we keep in mind that this was a coincidence. Google News is more useful since it showed a correlation in about 50% of the cases, yet the result is not fully convincing, since still 50% of the cases did not show a correlation.

This knowledge offers us the possibility to say something about the added value of these databases in business intelligence tools. First should be emphasized that, due to several reasons, these databases alone will not do the trick of making reliable forecasts. Noise is still a major disruptive factor, which can only be filtered out if the history and developments of the technology are thoroughly analysed. Second, it should not be forgotten that multiple search terms or even search combinations would probably be of much higher added value than it is regarded now. However, an algorithmic program should make it possible to include even more words in the query. A last major disturbing factor concerns the patterns of publications. They are often so dynamic that it is almost impossible to use them correctly. When each peak is constantly outdated by another peak, it is hard to determine if this peak occurs around the date of the hallmark. Companies need to reassure that they will not react too late and therefore, need to fully rely on the analysis. Nevertheless, when these databases are used in combination with other sources to make forecasts, they can become of great added value. Especially news databases have potential, since they seem to represent the popularity of consumers.

7. CONCLUSIONS AND DISCUSSIONS

7.1 Conclusions

Many conclusions were emerged from this research. We will elaborate on them step by step in a clear and structured way and give answers on the research questions. This section is divided in two parts. The first part focuses on the analysis on the phases, the second part focuses on the analysis on the hallmarks.

Part I – Phases

The goal of the first analysis (phases of the technology life-cycle) was to investigate the moment that scientific and news databases show the largest activity. The analysis was done by collecting data from both databases and calculating the average amount of hits per technology phase. The findings were remarkable, since we encountered a contradicting answer to our expectations.

a. **Scientific activity.** In the analysis, scientific activity in the technology life-cycle develops in parallel to other activities (e.g. news activity). It was expected that scientific activity would be most present in the beginning of the life-cycle (i.e. mainly before and during the innovation phase); however, we clearly saw that most cases of breakthrough technologies showed an increasing activity over time due to the growth in scientific publications. Remarkably, even in some cases the market stabilisation phase showed the largest scientific activity. Thus, if we would like to answer our research sub question, we have to say that the subsequent phase from the technology life-cycle shows more scientific activity in most cases of our analysis. So, the activity keeps (and thus publications) increase more and more over time.

b. **Market activity.** Market activity on the other hand was expected to show the highest amount of activity in the last part of the technology life-cycle of breakthrough technologies. The analysis showed that this perception was correct, since most cases showed the highest average amount of news publications in the market stabilisation phase. We clearly encountered a growth in news publications which walked parallel to other activities (patent activity and scientific activity). As a result, we can conclude that based on our findings, the market activity keeps growing in every subsequent phase of the technology life-cycle. Also, this gives us more insights in the usefulness of news databases implemented in forecasting tools.

Based on above information conclusions can be drawn about the development of scientific and news activity over time. In both cases a clear growth in publications over the phases of the technology life-cycle is noticed. The adaptation phase showed more activity than the innovation phase, but the market stabilisation phase showed more activity than the adaptation phase. Thus, activity increased over time and the phases. This result brought us to the second analysis which focused on the hallmarks of the technology life-cycle and the predictive character of the two databases. The analysis focused on a match between the databases and the technology life-cycles. Knowing that the activity in science and news increases can be included into the second analysis and thus it might say something about the usefulness of these databases and at which moment they can be used most effectively.

Part II - Hallmarks

c. **Scientific correlation between patterns.** Although in six of the 42 cases a (small) correlation (i.e. match) between both patterns was noticed, we believe this result is not significant enough to conclude that scientific databases applied in forecasting tools have a large added value. Since only the name of the breakthrough technology was used in the analysis, this could have caused a low result. However, the conclusion remains that scientific databases and their patterns show almost no correlation with the patterns (i.e. hallmarks) of the technology life-cycles.

d. **News correlation between patterns.** From our analysis we were able to investigate a possible correlation (i.e. match) between the patterns of the technology life-cycles and the patterns of news databases. In about 50% of the cases clear visual correlations were encountered. This result is significant to be included in future research, where even better results are expected to be found (i.e. using search combinations or different search terms). However, news databases can add value in forecasting tools due to the clear correlations that were found already in this explorative stage. Another remarkable result is that the correlations were found in all three phases of the technology life-cycle and therefore they can be used successfully over a large time-span.

e. **Added value of databases in general.** As a result, we claim that news databases have proved to be useful in forecasting tools. However, more research is required for improvement, but this will be explained below in the section 'future research'. Especially when news databases will be combined with other forecasting methods, it will strengthen the overall performance in forecasting.
   Scientific databases on the other hand, show low added value for future research and for application in forecasting tools. However, improved results can be expected by using more effective queries.

We are now able to fully answer the **main research question**:

- *How to predict the development of breakthrough technologies (i.e. the technology life-cycle pattern with three hallmarks and phases) with the help of electronic databases?*

Electronic databases such as Google News and Scopus offer a pattern of publications in their own field of interest. These patterns of news databases (not scientific databases) can be used to predict the pattern of development of breakthrough technologies, because some patterns show a correlation with the hallmarks. Because of these correlations, it is possible to see particular trends and peaks which can indicate a sudden change in the technology life-cycle. This phenomenon offers a company tremendous opportunities for creating a competitive advantage in a highly dynamic market environment and to successfully launch a breakthrough technology in the mainstream market.

Due to different issues (see next chapter) the reliability of the research is unclear. The results cannot be generalized, because the sample size is small and they are not statistically tested (only the first analysis about activity is statistically tested). Also, the boundaries of possible correlations were not identified by hard criteria, thus the level of subjectivity is high. Future research however, could solve these issues by exploring new boundaries and include further details on this topic. The next chapter will emphasize on these ideas.

7.2 Discussions

Several discussions could be raised on the conclusions and issues on this topic. Each of them is treated below.

Trends & cycles

Different viewpoints on trends and cycles are proposed in literature. Basically, two divisions in these viewpoints can be distinguished: a *sequential* (science-push model) and a *parallel* process of activities. These activities can be divided into scientific, technological, and market activity. Based on the analysis performed in chapter 3, the results tend to bend towards the *parallel* viewpoint.

Some authors (Schmoch, Grupp et al. 1991; Lacasa, Grupp et al. 2003; Schmoch 2007), are firm believers that the developments of these activities (i.e. scientific, technological, and market) occur in parallel. However, other authors (Myers and Marquis 1969; Utterback 1974; Watts and Porter 1997), believe this process develops sequential. They argue that science shows most activity in the beginning of the life-cycle, that technological activity is highest around the adaptation phase, and that market and business activity is highest in the market stabilisation phase. However, the analysis performed in this study matches to the vision of Schmoch (and Grupp), since publications in science and business/news both independently increase over time. However, discussions may rise on two issues. *First,* Schmoch and Grupp see a double-boom cycle in technological activity and they see a steady growth in scientific activity; however, they acknowledge that this does not occur in all science-based technologies. They carried out their study focused on 'robotics' and 'immobilised enzymes' and technological subfields included. The question rises how frequent this phenomenon actually occurs (also measured in more industries) and whether this cycle can be regarded as an exception or as the standard? *Second*, the analysis we have performed has limitations as well (further explained in the next chapter), because the sample size is small and only two industries have been taken into account. Thus, both scenarios have their limitations and their arguments. However, could both scenarios really occur or is the appearance of one of the two scenario simply coincidence? We do not know this, but it might be assumable that in certain industries one scenario will occur more often than another.

To further elaborate on this topic, we go back to the basic thoughts by simply distinguishing three main players in the development of breakthrough technologies: *scientists, companies and consumers*. From this basic perspective it is more than assumable that scientists are mostly interested in their own operating field: science. Consumers are mostly interested in obtaining and using the latest technologies. Companies just want to make profit and try to turn scientific knowledge into market success. From this simple, but plausible idea, it seems logical that science shows most activity in the beginning of the life-cycle. Market interest becomes larger when applications of this technology are implemented in products. Thus, from this very basic perspective it still seems plausible to embrace the science-push model. Now, we include another issue. This study focused deliberately on breakthrough technologies, since they are included in science for many years. Why is it then, that we see such contradicting results regarding our intuition? Keeping in mind that these breakthrough technologies have a long history of scientific research, why is it then that scientific activity grows over time and is almost not present in databases at the beginning of the life-cycle? It might be the case that basic research is not always published (maybe due to secrecy reasons?). Applied research on the other hand might be better financed in general due to the high expectations of the applications. This brings us to another factor.

Now it is time to include another viewpoint from literature. Rosenberg (1982) explains strong arguments why science is not only exogenous. He explains that science does not only occur from an external environment, but can be regarded as an endogenous process as well. According to his vision, significant chains of causation run from economic life to science as well as from science to economic life. Based on this viewpoint, it becomes plausible that technological and market activity should be present in the beginning of the technology life-cycle as well. It also becomes plausible that scientific activity will grow over time, especially once applied research is approved by managers of high-tech companies. It may be clear that there is still much uncertainty around this topic, because different viewpoints seem plausible. However, this analysis clearly showed that the activity in all fields increased parallel over time. Still, it would be incredibly interesting to further investigate the cause of these phenomena's.

Google News

This section can be divided into two parts, touching the topic of the sources within the databases and the timeline patterns created based on the hits of the databases.

Used sources within the databases

More discussions can rise concerning the sources of the databases. Especially Google News is constantly increasing the amount of sources within their database. The sources however, have different dates starting with publishing. Is the collected data therefore reliable when we see an increase in publications, or is this increase caused just by the increase of more sources? What is the effect of this phenomenon and how fixed are the sources over time? So many questions emerge touching this topic. This brings us to the two performed analyses which have their limitations, but both bring discussions as well. The main question that rises is about to what extend the sources vary. The subsequent question is: "what is the real effect?" Even more questions about the reliability of the analyses and the results soon emerge afterwards. The limitations and shortcomings will further be explored in the next chapter, but based on those presented arguments many discussions can still rise because there is not one clear answer to be given. The effect is present in Google News and Scopus, but both databases do not mention anything about this cause and the effects. We can argue whether this unclear phenomenon disturbs our analysis. We tend to believe it does not significantly. Due to the wide range of cases covering a large time span we believe the effect is small. Also, many sources (newspapers) have started publishing long before the invention of many breakthroughs of the technologies.

Timeline patterns

Another issue is raised for discussion. Google News Archive is used to identify the pattern of news publications for the historical breakthrough technologies used in this research. However, doing research on modern technologies would probably require more recent news, since Google News Archive includes news which is at least thirty days old (explained in paragraph 4.2.2). So, this delay needs to be included once the focus lies on recent breakthrough technologies. The discussion emerges on the two possible patterns from Google News Archive, which are divided in patterns of 'News Articles' and 'Timeline'. During this research it became clear that the timeline pattern is far more useful due to the better match with the technology life-cycles of the cases. However, a real distinction between both patterns is still vague. The timeline pattern seeks for web pages that describe *events* from the past and reorders the hits, displaying the pattern in a graphical and chronological timeline. However, it has to be said that this notion was not obtained from Google itself and thus assessment of the reliability of this statement is poor. Still, the timeline pattern matches better than the news articles pattern. This raises the question what the exact differences are between both patterns, since the news article patterns sometimes shows a match with the life-cycles of the cases as well, though less significant. Google does not mention much about the differences. Even after a request for more clarity from Google, the answer remained unclear. Are there really more sources included in the timeline pattern or are the hits just reordered over a chronological timeframe? Unfortunately, this topic remains unclear to some extent, but we do know that the timeline pattern is more useful. It is just still unclear why this is the case due to the lack of information available on this topic.

Patterns of publications

Random peaks in patterns

The patterns created by databases can, in some cases, form a useful indicator in forecasting tools. However, the peaks, in particular business and news patterns show many random ones. Some of them are simply explainable by certain occurrences; like for instance successful tests, emergence of standard, severe accidents. Although many of those causes can simply be filtered out, there are still some random peaks present in the patterns which cannot be explained. The question rises how these random peaks can be filtered out if they do not represent anything useful.

An answer to this question cannot be given directly. As said before, it is best if this knowledge about patterns is used in combination with other forecasting tools so that their reliability is supported by one another. However, is there no other way to eliminate those random peaks? First of all, there is possibly a solution, when multiple search terms are used in the query. Doing so correctly might result in a more detailed pattern which correlates even better. Still, random peaks will always be present emerging without any clear cause. Also, using different search names for the technology (i.e. brand names, trade names, and molecular names) might influence the results positively. Discussions might surface about combining multiple search terms and the possibility that the amount of hits reduce significantly, even though the level of efficiency might enhance. This might eventually lead to a pattern of publications that is simply too underdeveloped to be used in this research.

The capacity of publication patterns in forecasting tools

From the beginning of this research we have talked about forecasting methods implemented in business intelligence tools. We elaborated on the correlation between the technology life-cycles and publication patterns and how applied in forecasting tools. Bringing this topic to the surface does bring along certain questions.

Basically, we have studied the correlation (i.e. match) between both patterns. Also, future research will most likely enhance this vision, but the issue emerges of the *real* added value of having this knowledge. This research looked at historical cases and their correlations, but how can this knowledge be used in current cases? An example; if you as a manager want to know the popularity of your breakthrough technology in the market, you can measure the market activity with bibliometrics collected from business and news databases. However, if you see a distinct growth in hits, you still cannot say anything about the real popularity among future consumers. This issue is comparable to the stock exchange. People use different tools to predict the stocks, but these people never know when the *real* peak is crossed. The same holds for bibliometrics. If the hits are increasing, you will never know when the real peak is crossed or when the real peak will emerge. Thus, we are able to *track* historical patterns and their correlations, but until today we cannot clearly use this information for forecasting. We can only use it as a forecasting indicator, in which the added value will enhance once it will be combined with other forecasting tools.

8. LIMITATIONS AND FUTURE RESEARCH

This chapter will elaborate on the limitations of this research and give recommendations for future research.

8.1 Research limitations

8.1.1 Reliability and validity

Before delving into the issue of limitation, a distinction should be made between two forms of limitations. We introduce *reliability* of a measurement instrument (concerning its overall precision and accuracy). Reliability is often measured in repeatability. The other form concerns the *validity* of the instrument. This is important because it provides insights into the degree of which a measurement instrument achieves its aim. It focuses on the question: does it actually measure what you want to measure? Both will be explained regarding this thesis, before further elaborating on specific limitations of this study (Velde, Jansen et al. 2004).

The empirical data for this research are collected from two different sources. The data for the technology life-cycles is gained from the dataset from Dr. R. Ortt, who in his career as a scientist collected, improved, and verified his data over and over again. We consider his data as reliable. The other data are gained from databases. The issue of repeatability could be raised for this source, since Google News constantly updates its database with more newspapers, business magazines and other news sources. Therefore, the results from the data collection performed in July 2010 would not be repeatable when the same search would be conducted at another moment in time. Thus, we would not end up with the same dataset when executing the original data collection method. Data from Scopus should in theory give the same results when repeating the search, since Scopus uses the same search sources over and over again.

We are also able to discuss the validity of this research. We have examined the sources from our databases (Google News and Scopus) and based on that, we believe that Scopus indeed gives us insight in scientific activity and Google News on the popularity of a technology among (future) consumers. We have found evidence in literature forming a good indicator for measuring the activity on these different fields. Though, we could raise the question whether the sample size is large enough and whether we have drawn the right conclusions from our analysis. A larger sample size would strengthen the validity. Also, we could raise the issue whether it is fair to compare the technology life-cycles with the patterns of publications. Is it really true that scientific and market activity can be measured in scientific and news databases? We claim this is valid, because databases can be seen as an indicator for activity when authors find it valuable to inform consumers and other scientists on new discoveries, developments and other happenings. As we have also explained in our methodology, we claim that this is a valid way to investigate the added value of databases in forecasting tools, because it includes large amounts of data on a determined topic.

8.1.2 Limitations of this research

This study has several limitations in general. Firstly, the analyses are both limited to only 14 cases of breakthrough technologies selected from two industries (materials and pharmaceutical). The relatively low amount of cases can never reassure a certain level of significance. In order to upgrade the reliability of the analyses it would have been better if a larger sample size would have been used, based on more industries (e.g. telecommunication and electronics). All these technologies have been launched successfully in the market stream. It may be clear that unsuccessful technologies could not be included due to the fact that their technology life-cycle would not have been complete.

In order to develop forecasts for new technologies, the real reliability of these databases cannot be discussed. Only conclusions about the found correlations in the successful technologies can be drawn (14% in scientific databases and approximately 50% in news databases). However, now we face another limitation of this analysis. Although the first analysis about scientific- and news activity was statistically tested on significance, we have to face the high subjectivity of the second analysis. This subjective analysis consisted of a full historical study, followed by correlation analysis between the patterns measured with the naked eye of the researcher. Thus, there was an absence of a statistical testing and a limited level of objectiveness. From this point of view we cannot generalize based on our results of the second analysis, we can only use it as an indicator to perform future research. In the next paragraph more details will be mentioned for future research in order to reduce these limitations.

Secondly, one major limitation in this research is the lack of word combinations in the query. This seems to be important, since technologies often have multiple names (brand-, trade-, molecular names) which could create more useful results. Thirdly, this qualitative analysis has an inevitably error since not *all* histories have been collected into the smallest detail in the database from Dr. Roland Ortt. Although all data is checked and verified, there is a limited risk of containing small gaps. Some historical information is simply not (yet) findable due to the incompleteness of filing or due to secrecy issues. As a consequence, there might be a gap in the analysis on the historical developments of each technology. Fourthly, we cannot say anything about the reliability of Google News or Scopus and their corresponding hits. Those databases are external sources of data collection. Therefore, it is possible that the amount of hits found in a certain year of the pattern of publication contain a bias, since it is using standard algorithms that are not checked or verified. Google already claimed that the graphical representation of publications (publication pattern as shown in Figure 2), has a clear deviation compared to the actual amount of hits. Therefore, this graphical pattern can only be used as an indicator. This same deviation is applicable in the other algorithms without the awareness of the user of this database, although Google claims that the quantitative numbers mentioned in their search tool are fully reliable (explained in paragraph 4.2.2).

A final limitation further elaborates on the databases. Although increased activity is measured in science and news during the analyses since the databases find an increasing amount of hits, this does not necessarily mean that the real attention is indeed increasing. It should not be forgotten that more sources are included over time by the databases, since for instance Google News is improving their database every day by including more sources. It might be the case that the amount of sources increases as well, resulting in more hits over time and thus more activity, without an *actual* increase of attention. An explanation to what extend this phenomenon occurs is not easy to give, but we tend to believe that this measuring error is present but relatively small and thus an actual increase in activity is occurring. Reasons supporting this vision are the large time span we collected data from. Also, many sources used in Google operate already for a very long time. Unfortunately, there is no information from Google News or from Scopus findable about this topic. Only by looking at the full list of sources and by analysing at what moment each source started publishing would eliminate this measurement error. Then, it would become possible to calculate the average amount of hits, per phase *and* in relation to the amount of sources, resulting in a fair comparison between the phases. On the other hand, this phenomenon of increasing sources does not seem to be a major problem for finding correlations between the hallmarks and the publications pattern, because the analysis focused on the correlation between the patterns instead of the difference between the phases. The correlations can become more reliable due to the increase in sources. Thus, this would be applicable for the second analysis performed, but not for the first. So, there is a limitation present in this analysis, but we cannot say much about the level of significance. Therefore, it is assumable that no actual increase in activity is present, because the increase in activity is not an increase with a fixed amount of sources. However, for finding a correlation this does not seem to be a major problem.

8.2 Directions for future research

Future research should focus on the following issues in order to address the limitations mentioned above. This thesis gained a lot of new knowledge for implementation in future research and therefore can be regarded as a building block for the future. In this paragraph we will overview the directions for future research.

Google News

Since Google News can add much value to forecasting tools, the main direction for future research involves a full quantitative study on Google News. This is required since the executed analysis in this thesis has limitations and was of qualitative nature. The subjective character should be transformed into an objective character. We suggest choosing a quantitative method as explained in paragraph 4.1.1, which includes the calculation of the *slope* and the *peaks* of the patterns in order to determine the relation between both patterns. Further research should define a peak in publication in a more mathematical manner, resulting in an easier way to assess when a *real* peak occurs. Next, a statistical correlation between the hallmarks and the peaks in publication can be determined. This further research is quite important in order transfer from a subjective point of view to an objective point of view.

Next, including within this quantitative study we recommend an algorithmic model being capable of handling large amounts of data for simplifying the process of data collection. The large amounts of data from Google News can then be collected automatically by which they can be analysed more thoroughly (like distinguishing data per month instead of per year). It offers the ability to include more cases in the analysis.

Subsequently, we recommend enlarging the sample size on different areas. This can be done by including a wider range of industries next to the material- and pharmaceutical industry (e.g. telecommunications and electronics). Doing so will improve the insights on the added value and the usefulness of Google News in general and it offers to measure the differences between the industries. Also, the sample size will enlarge by collecting more successful cases of breakthrough technologies from different industries. The reliability will enhance and noise will be reduced.

To enhance the effectiveness of the query it is advisable to create search combinations. Not only the 'name of the technology' should be used, but also alternative names such as trade-, brand-, molecular names. These alternative names should be combined with the factors as presented in the model of Figure 11 in chapter 2. This model can be regarded as a source of inspiration for expanding search terms. The concept of combining search terms reduces noise significantly resulting in more effective analyses for future research on this topic. A last remark concerns to involve the criteria for making search combinations as presented in paragraph 4.3. Future research on query design is important, because it enhances the efficiency of the query.

This proposed methodology for future research presented above makes it possible to find answers to remaining questions. One particular question concerns the *time lag* which arises due to the fact that publications often emerge after a certain event occurred. For instance, once a new technology is launched into the market, it takes several days or months before publications emerge on this topic. If this market introduction occurs in December, it is assumable that publications emerge in January. Then, they will be listed as an article published in a new year. Thus, the effects of this time lag can be analysed by a quantitative analysis. Future research could also indicate more about the real correlation with an unambiguous significance.

Last, it is important to investigate the effect of increasing sources within Google News. The first analysis (in chapter 3) found an increase in scientific- and news hits over the technology life-cycle, however, it was not visible whether this increase was caused by a higher level of *real* attention or due to an increase in sources over time. In the analysis the amount of attention increased, but not with a fixed amount of sources leading to a measurement error. Future research could indicate the effect of this phenomenon by looking at each source and at what moment they were included in the database. Then it becomes possible to calculate the average amount of hits per phases *and* in relation to the amount of sources.

Patent databases

Future research should include patent databases as well for analyses. Besides the further exploration on Google News, it is advisable to include a full quantitative study on patent databases as well. These databases represent the technological activity and proved to have distinguishable trends. Moreover, combining patent databases with news databases will improve the capacity of forecasting tools even further.

Social Media

A final word is about social media. Nowadays there are many social communities interlinking many people from all over the world. Examples of social media websites are for instance Facebook, LinkedIn, and MySpace. Some of these community sites offer their users to create blogs, profiles groups, and all sorts of other applications. Due to the exceptionally strong network between friends, family and colleagues news rushes through this network. Not only does it reach many people at once in a short period of time, people can also discuss and join groups when they like specific topics. Facebook for instance has over 500 million users in July 2010[36]. Future research could include social network websites in order to investigate the added value of measuring attention among consumers. With a focus on groups and blogs on certain topics might enhance the perception of attention among consumers. Social media might enhance the capacity of forecasting tools to a great extent.

[36] *http://nl.wikipedia.org/wiki/Facebook*

REFERENCES

Abernathy, W. J. and J. M. Utterback (1978). "Pattern of Industrial Innovation." Technology Review: 41-47.

Agarwal, R. and B. L. Bayus (2002). "The Market Evolution and Sales Takeoff of Product Innovations." Management Science **48**(8): 1024-1041.

Arkes, H. R. (2001). Overconfidence in Judgemental Forecasting. Principles of Forecasting: A Handbook for Researchers and Practitioners. J. S. Armstrong. Dordrecht, Kluwer Academic Publishers**:** 365-386.

Armstrong, J. S., Ed. (2001a). Principles of Forecasting: A Handbook for Researchers and Practitioners. Dordrecht, Kluwer Academic Publishers.

Armstrong, J. S. (2001b). Selecting forecasting methods. Principles of Forecasting: A handbook for Researchers and Practitioners. J. S. Armstrong. Dordrecht, Kluwer Academic Publishers**:** 365-386.

Bass, F. M. (1969). "A New Product Growth for Model Consumer Durables." Management Science **15**(5): 215-227.

Bass, F. M. (2004). "Comments on "A New Product Growth for Model Consumer Durables": The Bass Model." Management Science **50**(12): 1833-1840.

Booz, Allen, et al. (1982). New Product Management for the 1980s. New York, Booz-Allen and Hamilton, Inc.

Brenner, L. A., D. J. Koehler, et al. (1996). "Overconfidenc in Probability and Frequent Judgements: A Critical Examination." Organizational behavior and human decisions processes **65**(3): 212-219.

Brentani, U. D. (2001). "Innovative versus Incremental New Business Services: Different Keys for Achieving Success." The Journal of Product Innovation Management **18**: 169-187.

Bush, V. (1945). Science, The Endless Frontier. Washington, DC, Public Affairs Press.

Chandrasekaran, D. and G. J. Tellis (2007). "Diffusion of New Products: A Critical Review of Models, Drivers, and Findings." Review of Marketing: 39-80.

Chandy, R. K. and G. J. Tellis (2000). "The Incumbent's Curse? Incumbency, Size and Radical Product Innovation." Journal of Marketing **64**(3): 1-17.

Christensen, C. M. (1997). The Innovator's Dilemma. When New Technologies Cause Great Firms to Fail. Boston, Harvard Business School Press.

Christensen, C. M. and J. L. Bower (1996). "Customer Power, Strategic Investment, and the Failure of Leading Firms." Strategic Management Journal **17**(3): 197-218.

Church, J. and N. Gandal (1993). "Complementary network externalities and technological adoption." International Journal of Industrial Organization **11**: 239-260.

Cooper, A. C. and D. Schendel (1976). "Strategic Responses to Technological Threats."

Crawford, C. M. (1977). "Marketing Research and the New Product Failure Rate." The Journal of Marketing **41**(2): 51-61.

Crawford, C. M. (1979). "New Product Failure Rates - Facts and Fallacies." Research Management **September**: 9-13.

Crawford, C. M. (1987). "New Product Failure Rates - A Reprise." Research Management **July-August**: 20-24.

Cunningham, S. W., A. L. Porter, et al. (2006). "Special issue on tech mining." Technological Forecasting and Social Change **73**: 915-922.

Dunphy, S. M., P. R. Herbig, et al. (1996). "The Innovation Funnel." Technological Forecasting and Social Change **53**(3): 279-292.

Garcia, R. and R. Calantone (2002). "A critical look at technological innovation typology and innovativeness terminology: a literature review." Journal of Product Innovation Management **19**(2): 110-132.
Gaynor, G. H. (1993). Exploiting cycle time in technology management. New York, McGraw Hill.
Golder, P. N., R. Shacham, et al. (2009). "Innovations' Origins: When, By Whom, and How Are Radical Innovations Developed?" Marketing Science **28**(1): 166-179.
Golder, P. N. and G. J. Tellis (1997). "Will it ever fly? Modeling the takeoff of really new consumer durables." Marketing Science **16**(3): 256-270.
Green, S. G., M. B. Gavin, et al. (1995). "Assessing a Multinational Measure of Radical Technological Innovation." IEEE Transaction on Engineering Management **42**(3): 203-214.
Grupp, H. (1998). Foundations of the Economics of Innovation - Theory, measurement and Practic. Cheltenham/Northampton, Edward Elgar.
Grupp, H. and U. Schmoch (1992). At the crossroads in laser medicine and polymide chemistry. Patent assessment of the expansion of knowledge. Dynamics of science-based innovation. H. Grupp. Springer, Berlin/Heidelberg, New York**:** 269-302.
Grupp, H., U. Schmoch, et al. (1990). Developing industrial robot technology in Sweden, West Germany, Japan and the USA. Measuring the Dynamics of Technological CHange. J. Sigurdson. London/New York, Pinter Publishers**:** 106129.
Hannula, M. and V. Pirttimaki (2003). "Business Intelligence Empirical Study on the top 50 Finnish Companies." Journal of American Academy of Business, Cambridge **2**(2): 593-599.
Haubt, R., M. Kloyer, et al. (2007). "Patent indicators for the technology life cycle development." Research Policy **36**: 387-398.
Hauser, J., G. J. Tellis, et al. (2006). "Research on Innovation: A Review and Agenda for Marketing Science." Marketing Science **25**(6): 687-717.
Hippel, E. v. (1986). "Lead Users: A Source of Novel Product Concepts." Management Science **32**(7): 791-805.
Irvine, J. and B. R. Martin (1984). Foresights in Science - Picking the Winners, F. Pinter.
Karakaya, F. and B. Kobu (1994). "New product developments process: An investigation of success and failure in high-technology and non-high-technology firms." Journal of Business Venturing **9**(1): 49-66.
Kleinknecht, A., K. v. Montfort, et al. (2002). "The non-trival choice between innovation indicators." Economic Innovations New Technologies **11(2)**: 109-121.
Kleinschmidt, E. J. and R. G. Cooper (1991). "The Impact of Product Innovativeness on Performance." Journal of Product Innovation Management **8**: 240-251.
Kuznets, S. (1966). Modern Economic Growth - Chapter one, Yale University Press, New Haven.
Lacasa, I. D., H. Grupp, et al. (2003). "Tracing technological change over long periods in Germany in chemicals using patent statistics." Scientometrics **57**(2): 175-195.
Langley, D. J., N. Pals, et al. (2009). "Imitation analysis - Early prediction of the market demand for major innovations." European Journal of Innovation Management **12**(1): 5-24.
Lee, Y. and G. C. O'Connor (2003). "New Product Launch Strategy for Network Effects Products." Journal of the Academy of Marketing Science **31**(3): 241-255.
Leifer, R., G. C. O'Connor, et al. (2001). "Implementing Radical Innovation in Mature Firms: The Role of Hubs." Academy of Marketing Executive **13**(3): 102-113.
Levintal, D. A. (1998). "The Slow Pace of Rapid Technological Change: Gradualism and Punctuation in Technological Change." HousIndustrial and Corporate Change **7**(2): 217-247.
Lichtenthaler, E. (2003). "Third generation management of technology intelligence processes." R&D Management **33**(4): 361-375.
Lichtenthaler, E. (2007). "Managing technology intelligence processes in situations of radical technological change." Technological Forecasting and Social Change **74**: 1109-1136.
Lieberman, M. B. and D. B. Montgomery (1988). "First-mover advantages." Strategic Management Journal **9**(S1): 41-58.

Lilien, G. L. and E. Yoon (1990). "The Timing of Competitive Market Entry: An Exploratory Study of New Industrial Products." Management Science **36**(5): 568-585.
Linden, A. and J. Fenn (2002) "Emerging Technologies Hype Cycle: Trigger to peak."
Lynn, G. S., J. G. Morone, et al. (1996). "Marketing and Discontinuous Innovation: The Probe and Learn Process." California Management Review **38**(3): 8-37.
Maier, F. H. (1998). "New Product diffusion models in innovation management - a system dynamics perspective." System Dynamics Review **14**(4): 285-308.
Mansfield, E., J. Rapoport, et al. (1981). "Research and innvation in the mordern corporation: Conclusions." Corporate Strategy and Product Innovation: 416-427.
McIntyre, S. K. (1998). "Perspective: Market Adaptation as a Process in the Product Life Cycle of Radical Innovations and High Technology Products." Journal of Product Innovation Management **5**: 140-149.
Meade, N. and T. Islam (2006). "Modelling and forecasting the diffusion of innovation - 25-year review." International Journal of Forecasting **22**(3): 519-545.
Meldrum, M. J. (1995). "Marketing high-technology products: the emerging themes." European Journal of Marketing **29**(10): 45-58.
Meldrum, M. J. and A. F. Millman (1991). "Ten risks in marketing high-technology products." Industrial Marketing Management **20**(1): 43-50.
Miller, J. P. (2000). Millenium Intelligence - Understanding and Conducting Competitive Intelligence in the Digital Age, Thomas H. Hogan, Sr.
Min, S., M. U. Kalwani, et al. (2006). "Market Pioneer and Early Follower Survival Risks: A contingency Analysis of Really New Versus Incrementally New Product-Markets." Journal of Marketing **70**: 15-33.
Moerloose, C. d., Ed. (2000). New Product Success Factors, New Products Typology and Moderating Impact on Success Factos According to the Type of Newness. Proceedings of the 7th International Products Developments Management Conference. Leuven, Belgium.
Mohr, J. J., S. Sengupta, et al. (2005). "Introduction to High-Technology. ." Marketing of High-Technology Products and Innovations: 1-31.
Mohr, J. J., S. F. Slater, et al. (2006). Foundations for successful high-technology marketing. Managing Technology and Innovation. R. M. Verburg, J. R. Ortt and W. M. Dicke, Routledge Taylor and Francis Group: 84-105.
Moore, G. A., Ed. (1991). Crossing the Chasm: Marketing and selling high-tech products to mainstream customers, New York: Harper, 1991.
Moss, L. T. and S. Atre (2003). Business Intelligence Roadmap - The complete project lifecycle for decision-support applications.
Myers, S. and D. G. Marquis (1969). Succesful Industrial Innovations. Washington, DC, National Science Foundation.
Nasakawa, T. and T. Nagano (2001). "Text analysis and knowledge mining system." IBM System Journal **40**(4): 967-984.
Nosella, A., G. Petroni, et al. (2008). "Technological change and technology monitoring process: Evidence from four Italian case studies." J. Eng. Technol. Manage **25**: 321-337.
Noyons, E. C. M., R. K. Buter, et al. (2003). "Mapping Excellence in Science and Technology Across Europe." Life Sciences - Report to the European Commission CWTS, Leiden.
O'Connor, G. C. (1998). "Market Learning and Radical Innovation: A Cross Case Comparision of Eight Radical Innovation Projects." Journal of Product Innovation Management **15**(2): 151-166.
Ortt, J. R. (2010). Understanding the Pre-diffusion Phases. Gaining Momentum: Managing the Diffusion of Innovations - Series on Technology Management (In Press). J. Tidd, Imperial College Press. **15**.
Ortt, J. R., D. J. Langley, et al. (2007). "Exploring the market for breakthrough technologies." Technological Forecasting and Social Change **7**(9): 1788-1804.

Ortt, J. R. and J. P. L. Schoormans (2004). "The pattern of development and diffusion of breakthrough technologies." European Journal of Innovation Management **7**(4): 292-302.
Ortt, J. R. and P. L. Schoormans (2004). "The pattern of development and diffusion of breakthrough communication technologies." European Journal of Innovation Management **7**(4): 292-302.
Ortt, J. R., C. M. Shah, et al. (2007). Strategies to commercialise breakthrough technologies. International Association for Management of Technology, IAMOT 2007 Proceedings**:** 1-20.
Ortt, J. R., C. M. Shah, et al. (2008). "Commercialisation Breakthrough Technologies: Scenario's and Strategies." Management of Technology Innovation and Value Creation: 205-220.
Ortt, J. R., S. Tabatabaie, et al. (2009). From invention to large-scale diffusion in five high-tech industries. International Association for Management of Technology, IAMOT 2009 Proceedings**:** 1-15.
Pawar, B. S. and R. Sharda (1997). "Obtaining Business Intelligence in the Internet." Long Range Planning **30**: 110-121.
Porter, A. L. (2007). "How "Tech Mining" can enhance R&D Management." Research Technology Management: 15-20.
Porter, A. L. and S. W. Cunningham (2005). Tech Mining - Exploiting New Technologies For Competitive Advantage, John Wiley & Sons, Inc. Hoboken, New Jersey.
Porter, A. L. and M. J. Detampel (1995). "Technology Opportunities Analysis." Technological Forecasting and Social Change **49**: 237-255.
Porter, A. L., A. T. Roper, et al. (1991). Forecasting and Management of Technology. New York, Wiley.
Porter, M. E. (1985). Competitive Advantage - Creating and sustaining superior performance, The Free Press.
Rosenberg, N. (1982). How exogenous is science? Inside the Black Box - Technology and Economics. Cambridge, Press Syndicate of the University of Cambridge**:** 141-159.
Schmidt-Tiedemann, K. J. (1982). "A new model of the innovation process." Research Management **25**(2): 18-21.
Schmoch, U. (2007). "Double-boom cycles and the comeback of science-push and market-pull." Research Policy **36**: 1000-1015.
Schmoch, U., H. Grupp, et al. (1991). "Monitoring science-based biotechnology by means of patent indicators." Research Evaluation **1**(1): 61-68.
Schmoch, U., S. Hinze, et al. (1996). The role of scientific community in the generation of technology. Organisation of Science and Technology at the Watershed.**:** 3-138.
Siegel, S., Ed. (1988). Nonparametric Statistics For The Behavioral Sciences, McGraw-Hill, Inc.
Slater, S. F. and J. Mohr (2006). "Successful Development and Commercialization of Technological Innovation: Insights Based on Strategy Type." The journal of product innovation management **23**: 26-33.
Song, X. M. and M. M. Montoya-Weiss (1998). "Critical development activities for really new versus incremental products." Journal of Product Innovation Management **15**(2): 124-135.
Stopper, J. M. (2002) "Roadmapping Continuous Innovation."
Stremersch, S. and G. J. Tellis (2007). "Indirect Network Effects in New Product Growth." Journal of Marketing **71**: 52-74.
Taschner, A. (1999). Forecasting new telecommunication services at a "pre-development"product stage. The future of the telecommunication industry - Forecasting and demand analysis. G. G. Loomis and L. D. Taylor. Dordrecht, Kluwer Academic Publishers**:** 137-165.
Tauber, E. M. (1974). "How Market Research Discourages Major Innovation." Business Horizons **17**(3): 22-26.
Tellis, G. J. and P. N. Golder (1996). "First to Market, First to Fail? Real Causes of Enduring Market Leadership." MIT Sloan Management Review **37**(2): 65-76.
Trippe, A. J. (2003). "Patinformatics: Tasks to tools." World Patent Information **25**: 211-221.
Trumbach, C. C., D. Payne, et al. (2006). "Technology mining for small firms: Knowledge prospecting for competitive advantage." Technological Forecasting and Social Change **73**: 937-949.

Tushman, M. L. and P. Anderson (1986). "Technological Discontinuities and Organizational Environments." Administrative Science Quarterly **31**(3): 439-465.

Utterback, J. M. (1974). "Innovation in industry and the diffusion of technology." Science **188**: 620-626.

Utterback, J. M. and W. J. Abernathy (1975). "A Dynamic Model of Process and Product Innovation." Omega, The Internation Journal of Management Science **3**.

Vanston, J. H. (2003). "Better Forecasts, Better Plans, Better Results - Enhance the validity and credibility of your forecasts by structuring them in accordance with the five different ways people view the future." Research Technology Management **January-February**: 47-58.

Vanston, L. K. and R. L. Hodges (2004). "Technology forecasting for telecommunications." Telektronikk **4**.

Velde, M. v. d., P. Jansen, et al. (2004). Guide to Management Research Methods, Blackwell Publishing.

Veryzer, R. W. (1998a). "Discontinuous Innovation and the New Product Development Process." Journal of Products Innovation Management **15**(4): 304-321.

Veryzer, R. W. (1998b). "Key Factors Affexting Customer Evaluation of Discontinous New Products." Journal of Products Innovation Management **15**(2): 136-150.

Watts, R. J. and A. L. Porter (1997). "Innovation forecasting." Elsevier Science Inc **56**: 25-47.

Watts, R. J., A. L. Porter, et al. (1999). "Functional Analysis: Deriving System Knowledge from Bibliographic Information Resources." Information - Knowledge - Systems Management **1**: 45-61.

Windrum, P. and C. Birchenhall (1998). "Is product life cycle theory a special case? Dominant designs and the emergence of market niches through coevolutionay-learning." Structural Change and Economic Dynamics **9**(1): 109-134.

Yoon, B. (2008). "On the development of a technology intelligence tool for identifying technology opportunity." Expert System with Applications **35**: 124-135.

Zeithaml, V. A. (1988). "Consumer Perceptions of Price, Quality, and Value: A Means-End Model and Synthesis of Evidence." The Journal of Marketing **52**(3): 2-22.

Zhu, D. and A. L. Porter (2002). "Automated extraction and visualization of information for technological intelligence and forecasting." Technological Forecasting and Social Change **69**: 495-506.

APPENDIX A - QUALITATIVE ANALYSIS ON SCIENTIFIC LITERATURE

The tables below represent the amount of hits found in the scientific database Scopus. Some years are missing, but when this occurs it means that there was no scientific literature available in that year. Analysis and conclusions of this data can be found in chapter 5.

Astroturf	# Hits
1964	0
1965	0
1968	0
1970	2
1973	1
1974	3

Teflon	# Hits
1938	0
1948	1
1950	2
1951	3
1952	1
1953	5
1954	8
1955	12
1956	13
1957	15
1958	13

Kevlar	# Hits
1965	0
1966	4
1970	11
1971	0
1973	7
1974	26
1975	33
1976	31
1977	73

Nylon	# Hits
1934	0
1937	0
1939	1
1940	1
1941	2
1942	2
1943	5
1944	2

PET	# Hits
1935	2
1938	1
1939	1
1941	0
1942	1
1946	1
1948	2
1949	1
1950	1
1951	5
1952	3
1953	2
1954	2
1955	8
1956	1
1957	7
1958	6
1959	5

Dyneema	# Hits
1964	0
1970	1
1975	0
1985	1
1986	2
1987	4
1988	3
1989	5
1990	5
1991	6
1992	7
1993	5
1994	2
1995	7
1996	11
1997	13
1998	11
1999	5
2000	8

X-ray	# Hits
1895	0
1896	30
1897	28
1898	7
1899	9
1900	2
1901	14
1902	9
1903	6
1904	6
1905	15

Contraceptive pill	# Hits
1927	0
1928	0
1944	0
1957	1
1961	1
1963	3
1964	4
1965	9

Aspirin	# Hits
1853	0
1899	0
1900	0
1958	1
1962	1
1970	1
1973	1
1978	2
1979	2
1980	1

Paracetamol	# Hits
1878	0
1953	0
1955	0
1961	1
1963	1
1964	4
1965	19
1966	39
1967	48
1968	66
1969	30
1970	63

Viagra	# Hits
1994	0
1996	3
1997	15
1998	245
1999	299
2000	235
2001	249
2002	316

Minoxidil	# Hits
1960	0
1970	0
1972	3
1973	20
1974	66
1975	61
1976	76
1977	85
1978	97
1979	120
1980	150
1981	186
1982	174
1983	174
1984	166
1985	183
1986	152
1987	165
1988	161
1989	166
1990	149

Prozac	# Hits
1971	0
1982	0
1986	2
1987	1
1988	15
1989	31
1990	85
1991	104
1992	108

Polio vaccine	# Hits
1952	1
1953	1
1954	4
1955	10
1956	4
1957	3
1958	6
1959	8
1960	6

APPENDIX B - QUALITATIVE ANALYSIS ON NEWS ARTICLES

The tables below represent the amount of news and business press articles (i.e. total amount of hits) which are generated by Google News. Data is not collected for the full life-cycle, but around the hallmarks dates. This is deliberately done in order to stick to relevant data. However, the data is selected with a maximum of ten years before the first hallmark and a maximum of ten years after the last hallmark. In some cases (e.g. in the case of Viagra), we have chosen for less years due to the very short technology life-cycle. In these cases we have chosen for the amount of years equal to the length of the innovation or adaptation phase.

The focus in this analysis is not only on the hallmarks, but analysis on arbitrary peaks is also included. These can be explained by the history of the breakthrough innovations. The years which are expressed in '**bold**' can be considered as the hallmark years. Last, as has been explained earlier, the focus only lies on 'Timeline' because this pattern is far superior to 'News Articles' due to the fact that it includes both news articles and reorders the hits in a chronological order. Also, in this appendix all technologies are shortly explain by their function and application. All information is collected from the private and non-published database by Roland Ortt, who verified each technology and its history.

Astroturf

Year	News articles	Timeline
1959	0	3
1960	0	3
1961	0	3
1962	1	7
1963	4	5
1964	0	10
1965	0	27
1966	54	120
1967	65	99
1968	237	285
1969	445	523
1970	619	788
1971	503	620
1972	373	441
1973	357	408

Astroturf is also knows as artificial grass and was invented as a substitution for natural grass. The name Astroturf was named after the company that produced the first ever sporting ground that was made of artificial grass. An official definition is: A plastic fabric which is applied in replacement of natural grass, and is herby being used as a surface on which sports are played or used as a children's playground.[37]

The data clearly shows that in and after the year of invention (1964), the amount of publication increases compared to the previous year. Market introduction (1965) shows an even further increase of publications, especially a significant increase of publications one year later (1966). The third hallmark large-scale production and diffusion (1968) is also clearly visible in the data, because after a small decrease in publications in 1967 a sudden significant increase emerges and continues to rise till 1970. This increase of popularity can be related to the fact that more and more Astroturf was laid on soccer fields.

[37] Private database by Dr. R. Ortt

Kevlar

Year	News articles	Timeline
1961	0	9
1962	0	11
1963	0	2
1964	0	12
1965	0	69
1966	0	20
1967	0	27
1968	0	15
1969	0	17
1970	0	10
1971	0	49
1972	0	30
1973	1	20
1974	27	51
1975	28	78
1976	31	64
1977	38	54

Kevlar can be defined as a super strong fiber with long chains that are aligned in one direction which makes it strong. It is used in many applications like bulletproof clothing, radical tires, and flame-resistant fabrics. It was invented in 1965, which the data also represents with a significant increase in publications. After this year the publications became stable and did increase. In 1970, Du Pont showed interests in commercialising the fibers. One year later Kevlar was introduced in the market (1971). In that year the amount of publications also increased, though not as significant as with the moment of invention.

In 1972 a large plant was built, which resulted in large-scale production and diffusion of Kevlar in 1973. Marketing campaigns were started to sell larger amounts of fibers. After large-scale production and diffusion the amount of publications increased as well. The applications of Kevlar increased, especially when Firestone and Goodyear started producing Kevlar reinforced radical tires. The amount of publications became stable after this year, though with a small peak in 1975 which can be related to new research which explained possible carcinogenic threats.

Dyneema

Year	News articles	Timeline
1954	0	3
1955	0	2
1956	0	2
1957	0	4
1958	0	2
1959	0	5
1960	0	15
1961	0	9
1962	0	11
1962	0	2
1964	0	12
1965	0	69
1966	0	20
1967	0	27
1968	0	15
1969	0	17
1970	0	10

Dyneema is a strong fiber made of polyethene. The chains are crystallised and aligned parallel to each other, giving the strength of the fiber. It is used for many different applications, like (fishing and sailing) lines and bulletproof clothing.

Already in 1960 the data shows a small increase of the amount of publications. However, this peak cannot be related to the history of the development of Dyneema. On the other hand, during and after the invention (1964) there was a clear increase of publications in news articles. The data of the publications and the data of the history of Dyneema are related here. In the beginning of the 1970s strong poly-ethyleen fibers are formed with a new ‘surface growth’ method. The amount of publications stay relatively stable during this period, until the second hallmark occurs; market-introduction (1975). In this year the amount of publications increases again to a new peak.

Although the ‘surface growth’ method was used for the development of Dyneema, in 1976 another methods was developed by DSM; ‘gel-spinning’ process. It takes three years to further develop this technique before it was patented in 1979. After several years of further developments, a joint

1971	0	49	venture of DSM and Toyobo, and the start of a pilot plant in Japan, the start of Dyneema fiber production in Europe, the third hallmark of large-scale production and diffusion occurred in 1990. The amount of publications slowly increased as well during this period. Therefore the hallmark is not particular clearly visible compatible to the data, although one year later the amount of publications did show a higher growth than previous years and also reached a peak, because after this year the hits diminished again. After the third hallmark more and more plants were realised over the world and the publications slowly increased as well, though no further relation seem to be present compatible with the history of Dyneema.
1972	0	30	
1973	0	20	
1974	28	52	
1975	31	81	
1976	32	66	
1977	38	56	
1978	56	86	
1979	57	87	
1980	78	136	
1981	77	111	
1982	124	174	
1983	184	266	
1984	184	251	
1985	203	254	
1986	254	329	
1987	211	293	
1988	308	375	
1989	302	373	
1990	363	422	
1991	392	513	
1992	350	449	
1993	373	461	
1994	382	459	
1995	358	423	
1996	407	472	
1997	444	525	
1998	565	644	
1999	794	927	
2000	845	924	

Nylon

Year	News articles	Timeline
1930	31	156
1931	22	79
1932	31	49
1933	16	49
1934	20	71
1935	21	204
1936	34	77
1937	33	209
1938	59	321
1939	208	453
1940	1090	1540
1941	2100	2340
1942	2330	2500
1943	1670	1800
1944	1340	1400

Nylon is a synthetic plastic material composed of polyamides of high-molecular weight and usually manufactured as a fibre and can be regarded as a true breakthrough material. The applications of Nylon are quite broad, like clothing, construction materials, and sport materials.

The data starts with a peak in 1930, but cannot be related to the history of Nylon. Nylon was invented in 1934. According to the data it is clear that the amount of publications increased as well, especially in 1935. After 1935 Nylon lost its popularity in news articles, but became very popular again in 1937, the year of the second hallmark, market-introduction. This popularity maintained for two years, before the amount of publications really increased significantly. This significant growth started one year before the third hallmark; large-scale production and diffusion. Therefore, in 1940, 1941 (third hallmark), and 1942 the hits reached a clear peak, which can be related to the hallmark and the popularity it gained in that period.

Teflon

Year	News articles	Timeline
1930	14	24
1931	11	13
1932	8	10
1933	11	16
1934	14	23
1935	15	21
1936	8	20
1937	12	22
1938	12	138
1939	16	22
1940	16	28
1941	6	31
1942	10	11
1943	6	8
1944	7	13
1945	14	39
1946	11	40
1947	7	15
1948	12	21
1949	14	23

Teflon is a material and is composed from polytetrafluoroethylene (PTFE). It is well-known as a plastic with the lowest coefficient of friction. It is mainly used as a non-stick coating for pans and other cookware due to its heat-resistance, non-corrosive, inert and slippery characteristics.

In the beginning of the 1930s PTFE was patented, before it was officially invented in 1938. The data of publications clearly shows an significant increase in the year of this hallmark. There is no denying that both are interrelated to each other. After the first hallmark the publications diminish, partly due to difficulties scientist encounter with the further developments.
In 1944, DuPont registers the trademark Teflon, which possibly relates to the fact that in 1945 and 1946 the amount of hits also increase. Especially in 1946, when the first products were already commercialized on a small scale under the commercial trade name Teflon, publications increased. Then Teflon was used in different applications like o-rings, valves, and other industrial applications in 1948. This year is also known for its second hallmark, market-introduction. The amount of publications on the other hand shows no significant increase in that year. This partly also applies for third hallmark, large-scale production and diffusion. Only a small increase in publications is visible in the same year. Thus, only the first hallmarks seem to have the clearest relation with the pattern of publications.

1950	14	34
1951	19	25
1952	25	32
1953	24	32
1954	46	91
1955	37	51
1956	40	75
1957	33	58
1958	50	92

Still, the data shows in 1954 a new peak in publications, while the hallmarks already occurred. According to the history of Teflon, in 1954 an engineer discovered a process to adhere Teflon to aluminium. This process resulted in much more application of Teflon, like non-stick cookware and protection tools. One year later, the process was patented, and in 1956 the non-stick cookware was commercialized into the market.

PET

Year	News articles	Timeline
1935	0	31
1936	2	15
1937	0	17
1938	0	33
1939	0	23
1940	6	29
1941	1	46
1942	0	19
1943	0	15
1944	0	12
1945	0	20
1946	0	37
1947	1	25
1948	10	47
1949	2	24
1950	4	78
1951	37	82
1952	85	132
1953	128	194
1954	154	207
1955	181	251
1956	304	382
1957	198	245
1958	303	390
1959	379	456

PET is a material which sometimes is also known as PETE or PETP. The material is composed from polyethylene terephthalate. It is a thermoplastic used in synthetic fibers, beverages, food and other liquid containers, thermoforming applications. In the textile industry it is referred as “Polyester” and in the packaging industry it is referred as PET. The material contains many different trade names, which creates difficulties to search effectively to publications. Searches on PET and polyethylene terephthalate resulted in very poor hits with much noise. Therefore, we chose for the search term “Polyester”, which offered much better results.

Before 1940 there were some developments on polyesters, but they were unsuitable for fibre spinning. However, in 1941 they found a solution for this and PET was invented and patented. According to the data, in that same year there is a clear peak present. In 1946 the material was introduced in the USA under the brand name “Dacron”. In 1948 there was a new introduction; Terylene a brass-bobbin yarn for lace-making. The amount of publications also increased in this year, although this was not the official year of market-introduction (which was 1946). After some years of further developments, a standard emerged in 1950, which resulted in a continuous growth in publications. In 1953, the third hallmark occurred. The publications show a clear growth as well in that year, although more publications emerged.

The difficulty with this innovation is that it entails many different trade names and different variations of the material. As a result, it is hard to relate the history of the material to the data.

X-ray

Year	News articles	Timeline
1885	2	32
1886	4	35
1887	3	33
1888	3	23
1889	3	20
1890	5	41
1891	3	25
1892	3	25
1893	1	35
1894	5	32
1895	4	1390
1896	609	1310
1897	558	758
1898	398	547
1899	352	484
1900	266	394
1901	479	685
1902	892	1130
1903	873	1140
1904	794	1030
1905	703	928

X-ray is an electromagnetic radiation of an extremely short wavelength produced by the deceleration of charged particles of the transitions of electrons in atoms. The applications of X-ray are mainly for imaging human tissue and surgical applications, but also used for material testing and food inspection. However, other applications became more popular later in the life-cycle of X-ray. In the beginning X-ray was only popular in the medicine world.

In November 1895 X-ray was invented and within weeks after Röntgen revealed the first X-ray photographs in January 1896, news of the discovery spread throughout the world. 1896 was also the year of the second and third hallmark, market-introduction and large-scale production and diffusion. Then afterwards, the rays began to exploit more and more for medical purposes. The data collected by the Timeline of Google clearly shows an enormous peak in these two years. Publications slowly diminish after these two years.

In 1901 Röntgen was awarded of the first Nobel Prize for Physics. This might be related to the increase of publication after that year; however this cannot be said with certainty. The peak in publications in 1902 till 1904 also cannot be related to the history of X-ray, due to the fact that data is missing in these years.

Aspirin

Year	News articles	Timeline
1843	0	1
1844	0	2
1845	0	2
1846	1	5
1847	2	3
1848	2	2
1849	2	7
1850	0	7
1851	1	1
1852	1	13
1853	3	29
1854	1	1
1855	1	2
1856	0	6

Aspirin also called acetylsalicylic acid (ASA) derivative of salicylic acid is a drug based on the compound called salicin. Aspirin (technology) is defined to be the combination of synthesizing salicyclic acid and buffering it. The drug is used for multiple ailments, like headache, muscle and joint aches, reduction of fever and swelling and mild infections. Also, it is used in the treatment of such conditions as unstable angina or following a minor stroke or heart attack because of its ability to inhibit the production of blood platelet aggregates, which may cut off the blood supply to regions of the heart.
Aspirin has a long history of developments. Most remarkable is the long pause of almost 50 years between the first and second hallmark. However, in that pause there are some peaks which might be explained by the history of Aspirin.

Aspirin begins to be more popular in 1852; however, the history of Aspirin does not show any relation to this peak. One year later in 1853 on the other hand, there is reason to believe

1857	1	2
1858	2	6
1859	1	19
1860	2	10
1861	4	5
1862	2	4
1863	2	26
1864	0	4
1865	0	7
1866	1	2
1867	3	8
1868	2	19
1869	1	5
1870	2	4
1871	1	4
1872	2	2
1873	4	7
1874	4	13
1875	2	7
1876	4	26
1877	1	2
1878	5	12
1879	11	18
1880	2	2
1881	2	2
1882	3	11
1883	4	8
1884	3	13
1885	6	10
1886	5	17
1887	6	30
1888	7	11
1889	2	26
1890	2	6
1891	7	13
1892	5	8
1893	6	33
1894	8	25
1895	6	16
1896	4	25
1897	2	129
1898	0	35
1899	5	232

that the peak in publications can be related to the first hallmark; invention.

Than for a few years there were little publications about Aspirin. Publications increased again in 1859, when a Professor (Kolbe) was able to proof new improvements on syntheses. However, it will take 15 years before he realised an improved version which would be used on a larger-scale.

More and more improvements occur. One major improvement was done by Jahoann Kraut in 1869, who synthesized the compound in a purer form. Unfortunately, one year prior to this year, publication increased already and can therefore not directly be related to this major improvement. Other peaks can also be found in the years to come but cannot be related to the history of Aspirin.

One very clear peak can be found in 1897. Chemist, Felix Hoffman, chemically synthesizes a stable form of ASA powder that relieves his father's rheumatism. The compound later becomes the active ingredient in aspirin named "a" form acetyl, "spir" from the spirea plant (which yields salicin) and "in", a common suffix for medications. However, the compound had already been invented decades before in 1853 by French chemist Charles Gerhardt and was being manufactured by the "Chemische Fabrik von Heyden Company" in 1897. This resulted in a major peak in publications. One year later, Aspirin was ready to be commercialised, which was done in 1899. It was the first drug that was commercialised in tablet form. Only one year later in 1900 Aspirin started its large-scale production and diffusion. It became the best-selling drugs in the world. Publications were tremendously high in 1899, but decreased during the third hallmark.

Thus, all hallmarks can be found in the history of Aspirin. However, there are some small peaks which cannot be explained by the history.

1900	4	88
1901	3	59
1902	4	47
1903	8	42
1904	3	29
1905	6	19
1906	2	17
1907	51	21
1908	47	70
1909	75	67
1910	133	86

Paracetamol

Year	News articles	Timeline
1868	0	1
1869	0	1
1870	0	1
1871	0	0
1872	0	1
1873	0	7
1874	0	0
1875	0	0
1876	0	1
1877	0	4
1878	0	7
1879	0	0
1880	0	0
1881	0	0
1882	0	0
1883	0	1
1884	0	1
1885	0	3
1886	0	5
1887	0	6
1888	0	1
1889	0	1
1890	0	1
1891	0	0
1892	0	0
1893	0	23
1894	0	1

Year	News articles	Timeline
1920	0	0
1921	0	0
1922	0	3
1923	0	0
1924	0	2
1925	0	0
1926	0	0
1927	0	0
1928	0	2
1929	0	2
1930	0	3
1931	0	1
1932	0	0
1933	0	1
1934	0	0
1935	0	1
1936	0	1
1937	0	0
1938	0	1
1939	0	3
1940	0	0
1941	0	0
1942	0	0
1943	0	2
1944	0	0
1945	0	1
1946	0	2

1895	0	2	1947	0	1
1896	0	0	1948	0	23
1897	0	2	1949	0	5
1898	0	0	1950	0	4
1899	0	15	1951	0	5
1900	0	1	1952	0	1
1901	0	0	**1953**	0	5
1902	0	0	1954	0	0
1903	0	2	**1955**	0	12
1904	0	0	1956	0	14
1905	0	1	1957	0	1
1906	0	0	1958	0	4
1907	0	0	1959	0	0
1908	0	1	1960	0	1
1909	0	0	1961	0	1
1910	0	0	1962	0	1
1911	0	4	1963	0	11
1912	0	5	1964	0	1
1913	0	0	1965	0	1
1914	0	0	1966	0	6
1915	0	0	1967	3	5
1916	0	6	1968	0	5
1917	0	0	1969	0	3
1918	0	2	1970	1	4
1919	0	0			

Paracetamol (ecetaminophen) is one of the most popular and widely used drugs for the treatment of pain and fever and it is available in both liquid and tablet form. The drug has many different brand names all over the world; however, they all come from its chemical name (Para-acetyl aminophenol). Paracetamol is often perceived as the alternative to Aspirin and not considered to be in the same class of painkillers. Paracetamol kills the pain effectively, but does not like morphine or codeine kill the pain at brain level, which makes them highly addictive. Compared to Aspirin, Paracetamol does not belong to the non-steroidal anti-inflammatory drugs category. In this way Paracetamol distinguishes itself from other painkillers.

Paracetamol has a long technology life-cycle. Especially the innovation phase creates a long process of trial and error. As a result it takes 75 years before Paracetamol is launched in the market after it was invented. The cause of this phenomenon will be explained below around the year 1893. The first hallmark (invention) is in 1878. The timeline-data shows a small peak in publications, but it is doubtful whether this amount is significant due to the fact that that was a similar peak five years prior to invention. However, that peak cannot be related to the history of Paracetamol. After the invention there were hardly any publications until 1893. In that year, Germany's leading physiologist (Von Mering) investigated Paracetamol and concluded that although it was an effective analgesic and antipyretic, in some cases the use was associated with the development of methaemoglobinaemia which resulted in the cause of blue lips and darkening of the skin. Although this phenomenom was rarely fatal, it was very alarming for patients and their carers. As a result, further investigation of Paracetamol was abandoned for over half a century. Today, experts expect that Von Mering did wrong observations.

Only a few years later in 1899 there emerged another small peak in publications, when a new discovery was done about acetanilide and that it metabolizes in the body to become acetaminophen. However, publications did not increase tremendously. Only in 1948, five years before the occurrence of the second hallmark, publications rise again. History explains that in 1946 and 1947 more developments and discoveries emerged on Paracetamol. In 1948, two researchers finally determined that the use of Paracetamol did not have the toxic effects of acetanilide. From 1951 to 1953, publications explained that Paracetamol was a safer alternative to Aspirin. In 1953, the second hallmark finally emerged and Paracetamol was officially introduced in the market. However, initially there does not seem to be a correlation between publications in that year and the hallmark. Only in 1955, when Paracetamol was produced and diffused on large-scale, publications rose a bit. It was clearly marketed to physicians and pharmacists that Paracetamol did not contain the undesirable effects of Aspirin (e.g. stomach irritations). In 1956 and 1956 publications on Paracetamol clearly growth.

A final peak in the data is shown in 1963 when Paracetamol was added to the British Pharmacopoeia and its popularity as an over the counter analgesic increased rapidly.

Contraceptive pill

Year	News articles	Timeline
1917	0	0
1918	0	0
1919	0	0
1920	0	0
1921	0	2
1922	0	1
1923	0	1
1924	0	0
1925	0	0
1926	0	0
1927	0	0
1928	0	2
1929	0	1
1930	0	0
1931	0	0
1932	0	2
1933	0	1
1934	0	1
1935	0	0
1936	0	1
1937	0	0
1938	0	0
1939	0	0
1940	0	1

The contraceptive pill was developed for contraceptive and regulation of menstruation and inter-menstrual bleeding. The pill contains components like estrogenic or progestational steroids or a combination of both. Most oral contraceptives contain a combination of both steroids, and prevent the release of eggs from ovaries.

Although the amount of publications is very limited during the search on contraceptive pill, the first "peak" can be found in 1921. At that time, a researcher Haberlandt finished some important experiments successfully.

The first hallmark occurs in 1927, followed one year later by the second hallmark; market introduction. Market introduction was only marketed as a medicine, not as a contraceptive. However, there is no increase in publications and therefore the data shows no clear significant correlation with the hallmarks. Though, in 1928 there is the same amount of publication as in 1921. However, we do not see this as a significant amount.

The third hallmark, large-scale production and diffusion, occurs in 1944. It is remarkable that in that same year publication do not show anygrowth. However, one year later publications reach a new "peak", but not significant.

In 1951 there is a real peak emerging. According to the history of Contraceptive pills, new discoveries were found. A variant of progesterone was discovered in the lab. Since 1951 different funds were collected for more research with the goal

1941	0	0
1942	0	2
1943	0	0
1944	0	0
1945	0	3
1946	0	0
1947	0	1
1948	0	1
1949	0	2
1950	0	4
1951	0	9
1952	0	2
1953	0	3
1954	0	1
1955	0	4
1956	3	6
1957	0	1
1958	0	6
1959	2	6
1960	2	52
1961	8	28

to develop a more efficient form of hormonal contraceptive. In 1955 two researchers from Harvard had found a solution. They came with a progestogen pill that would keep a woman from ovulating; therefore she could not get pregnant. The pill was called birth control and was approved by the Food and Drug Administration in 1960. It is remarkable that publications also reach a sky-high peak, which did not emerge before during the life-cycle of the Contraceptive pill. As a result, the hallmarks do not show a clear correlation with the data, however all the developments and progress in 1960 did show a correlation with the data.

As a result, in this case the hallmarks show a very poor correlation with the data. Other important developments were clearly visible in the data.

Viagra

Year	News articles	Timeline
1990	0	120
1991	0	74
1992	0	90
1993	14	60
1994	37	71
1995	53	106
1996	165	326
1997	219	370
1998	7740	13400
1999	6090	7470
2000	4710	5900
2001	4630	5590
2002	6550	7730

Viagra is a medicine which is fairly new in the market. It is a medicine which is rapidly absorbed in the human body and capable of being taken as a tablet. Viagra's effects is to inhibit an enzyme known as phosphodiesterase type-5 (PDE5), which naturally occurs in erectile tissue. PDE5 can break down cyclic guanosine monophosphate (GMP), the substance that is produced during sexual arousal and causes vascular and muscular changes that lead to an erection. Therefore, men who produce too little cyclic GMP and men who products too much PDE5 have problems obtaining and maintaining an erection.
It is remarkable that Viagra is introduced only four years after the invention, because normally medications have a long adaptation phase. In the 1980's NO (a gas found in automobile emissions) seem to have the effect of relaxing the muscle walls of blood vessels. In 1992 NO's effects seemed so formidable that *Science magazine* named it the molecule of the year. In 1991 and 1992 different trails were performed with sildenafil (trade named Viagra) with the limited result of lowering blood pressure. However, these studies learned that sildenafil causes penile erection. Treatment for impotence in that time was very limited. In those years, there were some

publications on Viagra, however, the trade name Viagra was not given yet.

In 1993 it became clear in tests that sildenafil was very effective for the stimulation of erections. One year later the drug was officially invented, allowing men to reverse erectile dysfunctions. Publications did growth as well, though not very significantly. In 1996 Viagra was patented. Viagra was approved by the Food and Drug Administration in 1998. Immediately, the second and third hallmark occurred; market introduction and large-scale production and diffusion. Sales immediately took off and became a great success between 1999 and 2001. Publications also took off since 1998. An immense peak is clearly visible in that same year. Later on publications diminished, but Viagra stayed quit popular because publications on Viagra kept coming on a constant level.

Prozac

Year	News acticles	Timeline
1957	39	51
1958	27	31
1959	22	24
1960	35	42
1961	48	54
1962	37	38
1963	29	31
1964	29	30
1965	36	38
1966	36	37
1967	41	42
1968	36	36
1969	45	45
1970	41	46
1971	38	42
1972	32	39
1973	40	44
1974	53	55
1975	55	64
1976	45	45
1977	80	81
1978	67	71
1979	68	75
1980	59	71

Selective Serotonin Reuptake Inhibitors (SSRI) are a class of antidepressants drugs and mainly used to treat depression. Nowadays it is also used for other disorders such as panic disorder, bulimia nervosa, insomnia and obsessive-compulsive disorder. SSRI's block the presence of serotine (a neurotransmitter), which causes a chemical connection in the brain that moderates a person's mood. SSRI's are well known by many different brand names; one of the most famous is Prozac. Overall, there are four types of drugs to treat depression. However, SSRI is known as a safer drug in case of an overdose and has far lesser side-effects than the other three types of drugs. For this data collection, the search term SSRI is used due to the fact that this term includes an overall term which is applicable for the many different brand names (i.e. Prozac is not used).

In 1953 it was noticed that it might be possible that serotine plays an essential part in people's brain to stay sane. In 1957 it was first discovered that behavioral changes can be influenced with the help of drugs. Much research was done for a decade, but little evidence was found that serotoninin could influence 'chemical imbalance'. Only in the late 1960s there was news about serotine that seemed to control mood and drive. In 1971, the compound Prozac was developed it was noticed that the serotonin and might influence depressions. The first hallmark was emerged. However, the amount of publications does not show a clear growth.

Then, in 1974 lab tests were in progress on fluxetine (trade name Prozac) and Prozac was patented. In 1978 serotine in

1981	36	41
1982	51	56
1983	41	48
1984	58	60
1985	58	64
1986	29	33
1987	4	54
1988	10	71
1989	8	16
1990	5	38
1991	8	42
1992	3	19

connection with fluxetine, and although the results were promising, not much was published due to competitive reasons. However, a small increase in publications is found in the data in 1977 till 1980. It takes many years before the clinic trails on flucetine were completed and submitted by the Food and Drug Administration. Then it took four years for the government to approve Prozac. Market introduction and large-scale production and diffusion started in 1982 with Zimilidine (one trade name of SSRI) and removed from the market one year later due to a potential danger for the user.

Some years after this event, more developments occurred and a new series of patents were filed. In 1986 Prozac started to take off in sales in Europe, after the trails were published one year before. In 1987 Prozac was also approved by the FDA in the US and in 1988 Prozac was officially launched in the market. In that same year it was made clear that Prozac was entirely safe. This is also noticeable in the amount of publications. In the 1990’s, the antidepressant group of drugs known as the SSRI’s (Prozac, Zoloft, Paxil) became the best-known drugs compared to their competitive drugs. Prozac took off the fastest. However, the data does not show this very clearly.

Polio vaccine

Year	News Articles	Timeline
1947	13	32
1948	5	21
1949	12	25
1950	13	38
1951	15	42
1952	45	79
1953	184	270
1954	868	1040
1955	4960	5800
1956	1630	1690
1957	1300	1340
1958	535	657
1959	692	848
1960	540	675

Poliomyelitis (acknowledged as Polio) is a viral infection of intestinal tract which belongs to the class of picornaviruses and the virus is found worldwide. There are three types of poliovirus. Type 1 is responsible for most large outbreaks, type 3 occasionally involved. Type 2 is associated with sporadic clinical cases. It is a virus which is one of the most feared and studied diseases in the beginning of the 20th century. The virus is highly infectious viral disease and it effects mainly children under three years old and can result in paralysis. There are two highly effective vaccines available: Oral Polio Vaccine (OPV) and Inactivated Polio Vaccine (IPV).

Vaccine developments for polio had begun in the early 1990’s. Many attempts failed, partly because researchers did not know there was more than one virus. In 1948 a breakthrough occurred when a group of researchers successfully cultivated the poliovirus in human tissue in the laboratory. Subsequently in that year, two research streams were focusing on the virus and a possible vaccine. Publications, however, do not show significant growth in that year or the years to come. In 1950, the first polio vaccine was claimed to be developed, however, it was still in the research stage and would not be ready to use.

A major outbreak of polio occurred in the US Outbreak 1952 and 1953. In 1952, the first successful vaccine was developed using a mixture of the three types of virus. The first hallmark was established. For three years more tests were done (also on people) and were regarded as successful. Publications slowly rose as well until 1955, when the second and third hallmark emerged. A polio vaccine was launched in the market and large-scale production and diffusion started as well. This is clearly visible in the data due to the fact that a significant peak was reached. This can also be related to a severe failure when 120,000 children were vaccinated. Ten of them were killed and 200 were permanently paralyzed and therefore it was one of the worst disasters in American history. The entire vaccine program had to be abandoned. In 1957 an improved vaccine was developed. Publications clearly show the popularity on this topic because publications were never been this high so far. One year later, when new tests turned out successful and soon worldwide distribution started again. Although publication on polio vaccine remained popular, there does not seem to be a further relation with the history of polio vaccine.

Minoxidil

Year	News articles	Timeline
1950	0	0
1951	0	0
1952	0	0
1953	0	0
1954	0	0
1955	0	1
1956	0	0
1957	0	0
1958	0	0
1959	0	1
1960	0	0
1961	0	0
1962	0	0
1963	0	0
1964	0	1
1965	0	0
1966	0	0
1967	0	4
1968	0	2
1969	0	0
1970	0	0

Minoxidil is a drug known for its ability to slow or stop hair loss and promote hair re-growth. At first it was discovered as a drug for lowering blood pressure. Minoxidil has different trade names in the market like Rogaine and Regaine, however Minoxidil is the overall ingredient in this medicine. Therefore this search term is used during this search. The history of Minoxidil starts in 1960, which is also the year of invention. However, there are hardly any publications available until large-scale production and diffusion takes over. From that moment publications increase significant.

After the invention in 1960 t here are many tests are done and many failed. Side-effects had to be diminished before it cure could be marketed. In 1970 the product is introduced in the market after it was approved by FDA for an emergency-use protocol of two weeks (therefore it can be considered as a niche application). Also this hallmark does not show any correlation with the amount of publications; possibly due to the short time it was marketed. In 1979 Minoxidil was first used exclusively as an oral drug (trade name Loniten) to treat high blood pressure (another characteristic of Minoxidil). Only in 1988 Minoxidil was introduced as a hair stimulant. This very first hair restoration medication was approved by the Food and Drugs Administation in order to treat hair loss. It became a hit in the market, after thorough testing and successful clinical trials. Publications also increase, however not that much as in other cases and also not in the same hallmark years.

1971	0	1	
1972	0	0	
1973	0	1	
1974	0	0	
1975	0	0	
1976	0	0	
1977	0	4	
1978	0	0	
1979	0	6	
1980	0	13	
1981	0	10	
1982	0	5	
1983	13	16	
1984	21	33	
1985	85	122	
1986	105	165	
1987	116	156	
1988	114	182	
1989	108	140	

APPENDIX C - RATIOS SCIENTIFIC AND NEWS DATABASES

Chapter 3 showed analyses and the moment when activity is highest in scientific and news databases. However, the ratios between both databases are interesting to observe as well. The overview below shows the ratios between both databases and ranked each result with identical criteria as explained in paragraph 3.3 and 3.4. Based on these findings it becomes visible to observe that the mean ranks increase over time. With the overview below in this appendix it is shown that the ratio of scientific hits to news hits gets higher over time. Thus, overall scientific hits will grow faster than news hits. Although this does not necessarily mean that the interest for the technology in the scientific community grows relatively faster, it may be of interest to take this difference in publication behaviour into account if news hits and scientific hits were to be used in a business intelligence tool.

Absolute number of hits						
Google News - news database (A)				**Scopus - scientific database (B)**		
Average hits per year Innovation phase	Average hits per year Market Adaptation phase	Average hits per year Market Stabilisation phase	**Breakthrough innovations**	Average hits per year Innovation phase	Average hits per year Market Adaptation phase	Average hits per year Market Stabilisation phase
			MATERIALS			
10	82	532	**Astroturf**	0,00	0,00	1,44
26	40	53	**Kevlar**	3	0,00	32
29	196	530	**Dyneema**	0,17	0,79	7,20
117	32	2127	**Nylon**	0,00	0,33	1,67
35	22	44	**Teflon**	0,00	1	5,17
22	61	304	**PET (polyester)**	1	1,86	4,43
			PHARMA & HEALTHCARE			
1390	1310	1310	**X-ray**	3,00	30	30
15	232	46	**Aspirin**	0,00	0,00	0,00
2,07	2,50	4,60	**Paracetamol**	0,00	0,00	0,60
0	0,69	2,44	**Contraceptive pill**	0,00	0,00	0,00
218	7740	8923	**Viagra**	4,50	245	260
57	56	53	**Prozac**	0,00	0,00	1,50
463	5800	3745	**Polio vaccine**	2	10	7
0,70	0,67	71	**Minoxidil**	0,00	45	180

Rank (scientific hits/news hits)				Ratio scientific hits/news hits		
			MATERIALS			
1	1	3	**Astroturf**	0,0000	0,0000	0,0027
2	1	3	**Kevlar**	0,1154	0,0000	0,6038
2	1	3	**Dyneema**	0,0057	0,0040	0,0136
1	3	2	**Nylon**	0,0000	0,0104	0,0008
1	2	3	**Teflon**	0,0000	0,0455	0,1174
3	2	1	**PET (polyester)**	0,0455	0,0305	0,0146
			PHARMA & HEALTHCARE			
1	3	3	**X-ray**	0,0022	0,0229	0,0229
-	-	-	**Aspirin**	0,0000	0,0000	0,0000
1	1	3	**Paracetamol**	0,0000	0,0000	0,1304
-	-	-	**Contraceptive pill**	0,0000	0,0000	0,0000
1	3	2	**Viagra**	0,0206	0,0317	0,0291
1	1	3	**Prozac**	0,0000	0,0000	0,0283
3	1	2	**Polio vaccine**	0,0043	0,0017	0,0019
1	3	2	**Minoxidil**	0,0000	67,1642	2,5352
18	**22**	**30**	**Total**			
1,50	**1,83**	**2,50**	**Mean ranks**			

APPENDIX D - PROTOCOL

This protocol is developed in order to verify *the procedure* of the qualitative analysis which is conducted for finding a possible match between the technology life-cycles and the patterns of publications (both scientific and news). Three fellow students from the Master Management of Technology have followed and executed this protocol. The protocol is written to support the analysis which contains a high level of subjectivity. The construct validity of the procedure that is used is can be validated in a way to enforce it. However, it does not test the interrelated reliability of the analysis. The protocol just explains in detail how the procedure looks like and the vision of three students is requested to validate this subjective procedure.

Step 1. Background

Being able to forecast certain developments on breakthrough innovations, on new technologies & applications, on market trends, and on scientific discoveries would give a company great future possibilities because it offers them creating stronger tactic strategies and the opportunity to gain a competitive advantage. However, to create this status would need many different tools to forecast in a reliable way. One (new) way to do so might be with the help of scientific and news databases. These databases search on publications related to the topic the searcher is interested in. It offers the opportunity to count the amount of hits over a particular time span. It is expected that over the technology life-cycle of a breakthrough technology there are certain phases and hallmarks that represent a moment of that time span. It is expected that the amount of hits from scientific and news databases show a certain peak or growth during this hallmark (i.e. the moment of invention, the moment of market introduction, and the moment of large-scale production and diffusion in the mainstream market). This thesis focuses on a match or “correlation” between both patterns. If a correlation would be found, it might become possible to use this application in a forecasting tool (business intelligence).

Step 2. Request

What we ask from you in this protocol is your vision on the data we have collected from three sources: 1.) the dates of the technology life-cycles dates; 2.) the patterns of scientific databases; 3.) the patterns of news databases. The data is attached to this protocol and we have given our results from our analysis. You might expect that the patterns are expressed in a graph; however, we have chosen to express the pattern in a table of data in order to clearly see the amount of hits per year. We would like to look at the fluctuation in the patterns and compare them with the years of the technology life-cycle. The years expressed in “**bold**” can be regarded as the hallmarks: 1.) invention; 2.) market introduction; 3.) large-scale production and diffusion.
We have executed a full historical analysis first and filtered out the irrelevant causes of fluctuations in the patterns. Afterwards, we have searched for correlations between the patterns based on a visual basis. Re results have been marked with a + or -, depending whether a match was found. We would like to hear your opinion about the validity of the procedure we have used to find these correlations. We hope to hear your opinion whether you agree or disagree with this procedure. We request you to only include the pattern “Timeline” in the data from the news database to validate the procedure.

Step 3. The data

The data is attached to this protocol, but the overview of the result is projected below for both the scientific and news database.

Scientific database (Scopus)

Breakthrough innovation	Hallmark 1	Hallmark 2	Hallmark 3
Astroturf	-	-	-
Kevlar	-	-	+
Dyneema	-	-	-
Nylon	-	-	-
Teflon	-	-	-
PET	-	-	-
X-ray	-	+	+
Aspirin	-	-	-
Paracetamol	-	-	-
Contraceptive pill	-	-	-
Viagra	-	+	+
Prozac	-	-	-
Polio vaccine	-	-	+
Minoxidil	-	-	-

News database (Google News)

Breakthrough innovation	Hallmark 1	Hallmark 2	Hallmark 3
Astroturf	-	+	+
Kevlar	+	+	+
Dyneema	+	+	-
Nylon	+	+	+
Teflon	+	-	-
PET	+	+	+
X-ray	+	+	+
Aspirin	+	+	+
Paracetamol	+	-	+
Contraceptive pill	-	-	-
Viagra	-	+	+
Prozac	-	-	-
Polio vaccine	+	+	+
Minoxidil	-	-	+

Step 4. General rules of thumb

In order to detect a correlation between the patterns we have used the following rules of thumb. We do not distinguish clear criteria of when we see a correlation, because this is an impossible task due to the large variety of patterns. We believe to see a growth in publications for instance when after a steady period of hits the amount of hits slowly increase (for instance double). However, we also see a correlation when hits grow extremely exponential. Thus, there are different viewpoints of how to detect a peak. However, we cannot set clear criteria for this; we simply do this by on a visual basis.

Step 5. The assessment

Please give your vision on whether you agree or do not agree with our vision on a correlation between the patterns. Do you agree with the procedure used in this subjective environment? You can do this with an explanation why you agree or not. If you believe this procedure is sufficient for this analysis please tick your agreement. Your viewpoint will be taken into account and therefore we ask you to check this honestly. Thank you for your cooperation.

Step 6. The assessment of the procedure

Student 1	AGREED
Student 2	AGREED
Student 3	AGREED