Expertise
and insight
for the future

Tommy Nurmi

# How a business can use automated systems to process written human language in order to enhance customer services online.

Metropolia University of Applied Sciences

Bachelors of Business Administration

European Management LXB14S1

Bachelor's Thesis

23.4.2020

Metropolia
University of Applied Sciences

| | |
|---|---|
| Author | Tommy Nurmi |
| Title | How a business can use automated systems to process written human language in order to enhance customer services online. |
| Number of Pages | 45 Pages |
| Date | 23.04.2020 |
| Degree | Bachelor of Business Administration |
| Degree Programme | European Management |
| Instructor/Tutor | William Simcoe |

This thesis focuses on the use of Natural Language Processing (NLP) and how it can be used to leverage competitive advantage. The main concern of NLP is how business are able to effectively process data through clear cut frameworks, use this pipe-line process in order to solve business problems involving customer services, and assist in the enhancement of customer oriented business strategies.

This topic is relevant because NLP and big data is recognized by business giants such as Amazon, Microsoft, and Google, as a critical tool to enable clear insights into Natural Language Understanding (NLU) which assists in the decision making process. By understanding customer expectations and sentiments through data, businesses are able to differentiate themselves from the competition, building a unique selling point (USP) through the cumulative efforts of utilizing NLP models to garner insights and sentiments from unstructured data.

The effective use of NLP tools and techniques by businesses requires the formation of clear objectives and metrics where quality data will be processed, and a suitable model will be deployed and evaluated. This requires an effective framework, which will determine the success of a quality NLP project in the scope of business. Qualitative research methods are utilized in this evaluation using secondary resources.

NLP is a function that helps businesses gain competitive advantage. This thesis examines the framework, limitations, and feasibility of NLP. This is significant because businesses

can use the frameworks and evaluative arguments provided to support decision making, as well as clarify the different ways NLP can be leveraged. A form of text clustering in the field of NLP referred to as sentiment analysis and topic modelling will be explained and examined through case studies. It was determined through the various case studies of Utopia Analytics, Dex, and Finnish government that the integration of systems which utilize NLP enables business to differentiate from the competition and create value for society. How effectively NLP can be utilized to enhance business functions will depend on quality data preparation, deployment of correct NLP models, and evaluation of the whole process under an approachable business framework.

| Keywords | Automated systems, Unstructured data, Natural Language Processing, Customer Service, Competitive Advantage |
| --- | --- |

## Contents

# 1 Introduction

As it becomes increasingly cost effective and reliable, businesses will consistently identify new and effective methods of automated data processing. Processing and understanding the complexity and nuances of human language through unstructured data will generate actionable information to help in the decision making process. Data is easily available in a wide variety of formats. The growth of digital channels which customers use to engage with businesses provides the opportunity for greater volumes of data to leverage.

Companies that are leveraging this data are able to create new services and products to exceed customer's expectations. However, 'the larger part of enterprise data, nearly 80 percent, is unstructured and has been much less accessible.' (Accenture 2019) To mitigate this challenge, businesses have developed the capability and capacity to process and evaluate data by deploying NLP models. These models include "applications ranging from trend identification and summarization, content and ad generation. All of these dynamics center around one thing: the customer." (Leggett 2020) Businesses continue to leverage these applications to improve business operations and reach objectives more effectively than companies who are not using NLP.

The value that is being created through NLP is scalable, adaptable, and compatible with most existing business architecture and infrastructure. Using NLP will develop greater understanding of customer's beliefs, needs, wants, and expectations. Businesses which use NLP tools including sentiment analysis and topic modeling are reaping the benefits of sustainable competitive advantage, and developing unique selling points.

## 2 Methodology

The author will analyze case studies to show how businesses use sentiment analysis and topic modeling to improve KPIs and enhanced business capabilities such as time management and active engagement of customers.

There are many methods and frameworks in business intelligence and data analytics to generate actionable information. The CRISP DM framework allows for an iterative process. Business go back and forth between phases as deeper understanding of the data and customer is developed. This will assist in better implementation of models in the context and scope of the objective. Furthermore complex data preparation eco-systems such as Hadoop are compatible with CRISP DM. Hadoop is implemented during the data preparation stage. When the data is prepared, all the other stages in the CRISP DM framework will remain unchanged. Explanation of Hadoop and evaluation of the limitations and biases when gathering data, and deployment of NLP models will be critiqued. Solutions to the disadvantages and drawbacks of using this framework and methodology will assist the reader in providing a balanced perspective on the feasibility of NLP, in the context of business and the customer. NLP is the most effective when there is quality data and a clearly defined business objective. To understand what quality data is, and how can it be gathered from information and messages, will be one section of the literature review. Businesses that evaluate the source of the data in context to their business objective will determine the impact of project. Success will be determined by indicators found within company financial statements such as increased cash flow or reduced operating expenditure. The challenges and context of customer service in the digital environment is reviewed for a greater understanding of how to define business objective, and choose the suitable framework. Flexible framework is needed when working with NLP. This will assist businesses to evaluate data in terms of quality, and deploy NLP models effectively.

# 3 Literature Review

NLP allows one to assimilate large quantities of data into meaningful insights and actionable information. The main challenge with NLP is generating information from large volumes of scattered data may result in confirmation bias. Businesses can act on information gained through NLP with a higher degree of confidence and use it to make decisions. However, this information could potentially be inaccurate because it is an accumulated result of various messages, and emotional feedback. It is a flow of messages from various channels of communication, and engagement. This can be subject to anchored beliefs and prejudices when engaging with a business's product and services. A counter argument to this is that NLP tools such as sentiment analysis can improve the effectiveness of decision making and value creation. However, businesses must have clear deliverables and objectives in which NLP can be integrated into existing business infrastructure. Moreover, the meaning of the information gathered from NLP will only be useful when the business can use it to enhance the decision making process, or use the information to create new and better products and services. To build the foundation for this practical understanding, various classical and modern perspectives of what actionable information is, what modern customer service demands, how to define clear business objectives, all in the scope of NLP systems, will be discussed in the following literature review.

## 3.1 Knowledge and Information Acquisition

Information is the flow of messages, while knowledge is created and organized by that flow of information, anchored on the commitments and beliefs of its holder. Information is the commodity capable of yielding knowledge. Knowledge is identified with information produced (or sustained) belief. The information a person gathers or receives is relative to the source of where the information is taken from. (Dretske 1981) This understanding emphasizes an essential aspect of knowledge that relates to how humans initiate action and make decisions. (Mackay 1985) Information is a necessary medium or material for initiating action and can be viewed from syntactic and semantic perspectives. Syntactic information is that which is measured with regard

to structure and rules. in NLP this refers to the grammatical relationship between words, which can be measured qualitatively. A semantic perspective on information refers to the meaning and value, which is measured quantitatively. (C, Shannon 1949) How actionable knowledge can be utilized to make decisions and enhance business functions will depend on the capacity to address semantic information in a quantifiable context, while taking into account the latent syntactic qualitative element.

NLP generated knowledge can yield successful action when applied to a business context. In business context this can be determined and evaluated through success indicators and quantitative metrics evolved through information analytics. The form of the message always changes according to the consumer platform and ecosystem. However the message that is to be understood is largely gained through qualitative inference. What makes this message valuable is how much actionable value can be derived from it. A statistical probability model, NLP tool, or advanced machine learning algorithm can be useful to further the understanding made through quantitative inferences. Searle's linguistic theory on communication states that there is a clear relationship between language and human action in terms of the "intention" and "commitment" of speakers. (Searle 1969) However, Finkelstein's argument is that one should not focus solely on quantitative and measurable indicators because it often provides an incomplete view of the larger context. These should be taken into consideration when observing and evaluating data in the form of customer feedback when using NLP tools such as semantic analysis.

"If we as scholars want to know why organizations do the things they do, or perform the way they do, we will not focus strictly on objective contextual factors. The mainstay constructs of strategy management researchers- factors such as the environment, competitors, allies, and the company's resources – will provide us woefully incomplete explanations of company behaviors. Instead, we need to also consider, in an integral way, the biases and dispositions of the people at the top of the firm. In doing so, we will find that human factors – deriving from personality, experiences, values social connections, fatigue, envy, and so on – play a substantial role in affecting organizational outcomes." (Finkelstein et, al. 2009)

Knowledge that can be expressed in words and numbers only represents one portion of the entire body of possible knowledge. (Mackay 1985) "Explicit" or codified knowledge refers to knowledge that is transmittable in formal, systematic language. This is generally found in structured, empirical data. "Tacit" knowledge has a personal quality, which makes it hard to formalize and communicate. Tacit knowledge is unstructured data. Tacit knowledge is rooted in action, commitment, and involvement in a specific context. (Grant 2007 173-180) For businesses to structure and process the human language through reasoning and inference using both tacit and explicit knowledge in the form of unstructured data 'is an important aspect of intelligence.' (Yang, Deng 2018)

## 3.2 Purpose of NLP

Converting disordered, scattered information into actionable intelligence is the purpose of NLP. NLP helps to analyze and convey meaning in the human language. (Gupta 2019) Business can utilize NLP applications in several ways. "The simplest is query understanding and content understanding – to improve both user experience and insight discovery. Understanding the user's natural language query inputs (text or speech) provide better, more targeted responses by understanding the user's questions and intent." (Accenture 2019) NLP assists in identifying alternative solutions to solving problem through gathering customer sentiments. This will convey deeper meaning and insight and help to understand customer behaviors, wants, goals, and beliefs.

The start of an NLP business project starts with text mining. "Text mining is the process of discovering information in large text collections, and automatically identifying interesting patterns and relationships in textual data. (Feldman, et, al. 2007) Finding meaning and developing business insights requires the text to be changed into a structured form. This structure can be syntactic in nature, capturing the grammatical relationships in the text, or semantic, capturing the meaning conveyed by the text. (Verspoor, Karin, et al. 2012) This helps businesses develop actionable information from customer feedback. "By deconstructing topics from masses of text, (this) allows companies to see what common issues, complaints, or positive or

negative sentiment customers have about products. These insights can lead to high-value outcomes, such as improving or creating new products that deliver a better user experience, responding timely to safety issues, and identifying which product lines are most popular with consumers." (Teradata 2018)

By defining clear businesses objectives, the information acquired by NLP is used by businesses to help in understanding the customer for better decision making in the short and long run. After the inputs and outputs are considered and conditioning patterns are taken into account, without a KPI, or a clearly defined business problem to be solved, it is difficult to judge the value and content of the information derived from NLP. There is limited consensus amongst researchers and practitioners on what project success factors are, what variables should be measured, and how they should be evaluated (Unterkalmsteineret, et al. 2011: 7-8).

## 3.3 Measuring Project Success

Organizations continually seek to leverage new and innovative ways to increase the likelihood of project success. The meaning of project success varies widely across literature, yet most definitions declare a project successful if it delivers within budget, on time, with acceptable quality, and meets overall stakeholder objectives. Mature enterprises define project-specific success criteria based on organizational needs and adaptable frameworks which aim to increase rate of project success. Selecting conductive methodologies to drive project success is often a trial and error process. However 'it is often possible to draw parallels from similar models.' (Boyd-Graber 2017: 112-114) "One should take into account that there is no standard set of criteria nor ruleset available for project managers to readily apply to facilitate project success." (Forney, Sandra, et al. 2019) Knowing what will work for project success is at best a trial-and-error process.

Project success factors can be measured through different quantitative indicators. These include customer retention rate, active engagement rate, and customer churn rate. These will provide indications on areas or functions that need improvement, and areas which can be developed into a USP for a business to gain competitive advantage.

When project categorization systems and criteria are not logically matched with project objectives, characteristics and environment, many projects are reported to fail to deliver on time, budget or do not give value to the client. (Ahimbisibwe, et, al. 2015) It is in the best interest for a business to define quantitative, measurable indicators of success before collecting data. Conversely, a business can collect data first. With the new understanding gathered from this data, leverage it to form new objectives based on what the market and consumer's needs. This data collected has the potential to develop into actionable information that supports decision-making, create value, and eventually sustainable competitive advantage. (Provost, Fawcett 2013) After defining KPIs, businesses should develop strategies to address the various challenges of customer service.

### 3.4 Challenges of Customer Service

The challenge of customer service online is to augment business functions with automated systems without losing the human connection. "Companies have lost sight of the importance of human interaction and often make it too difficult for consumers to get the right level of help and service that they need" (Morgan 2016) Trust is among the key mindsets and attitudes of successful human-machine collaboration. Initially, cultural resistance may be strong because the relationship between the inner workings of an artificially intelligent machine and the results it produces can be rather obscure. In a sense, it is no longer the algorithm but mainly the data used to train it that leads to a certain result. Humans will need some time to adjust to this shift. Getting started early not only helps produce results quickly but also helps speed up an organization's journey toward embracing the full potential of AI." (McKinsey 2017: 9-11)

Balancing trust and criticism of technology is a difficult task. A machine's success in achieving objectives is determined by the perseverance of those aiming to utilize it. Increasing the responsibility of automation for business functions using NLP will improve perceived value on paper. However, to reduce or maintain the number of external variables to a degree will allow the business entity to map out solutions. This will assist to gauge volatility and risk within a closed system. Applying this to a larger scale is what is analyzed in the analysis section.

Marketing has been defined as the task of finding and filling needs. "People satisfy their needs and wants with products. A product is any offering that can satisfy a need or want, such as one of the 10 basic offerings of goods, services, experiences, events, persons, places, properties, organizations, information, and ideas." (Kotler 2000: 6-9) When the business is able to address the customer's needs, and when there exists a clear requirement for the need to be filled, customer services are inherently enhanced. Customer service excellence has proven to be the only defensible long-term business strategy. (Gartner 2011: 5) Companies focus on researching current customers to identify their problems, gather new ideas, and to test proposed product improvements and marketing mix changes; this is where data mining and NLP can be utilized to augment business functions, generating significantly new products, services, business formats; establish new price points; develop new channels; raise service to an unbelievable level. (Kotler 2000: 6-9) Creating the framework for the synthesis of finding and filling customer's needs will help to define what the companies unique selling point is, as well as derive competitive advantage from this.

To learn and change there needs to be actionable knowledge in order to do so. This can be perceived as one of the main challenges in customer services. Organizations aiming to create competitive advantage by pushing the boundaries of knowledge should carefully consider the nuanced tradeoffs between specialized and diversified researchers when strategizing about hiring the optimal types of expertise. Recognizing and integrating new knowledge is positively associated with the propensity to diversify successfully and develop effective strategies for successful business practices. (Nagel, Teodoridis 2019). A critical element to achieving this is the framing of what success can be defined as, and how one can measure it. Once measurable, it can be evaluated, critiqued, and improved/adjusted according to the businesses wants, goals, and beliefs.

The challenge with the development and integration of data analytics is that, one is limited by the steep learning curve involved with data preparation. Businesses are reducing barriers to entry into NLP. These barriers to entry are being removed by vendors who are providing easy and low cost cloud storage, data processing, and analytics in all-in-one, low cost bundles." Vendors are providing easy and low cost

cloud storage, data processing, and analytics in all-in-one packages. This is defined as SaSS (Software as a Service). "They seek to create interactions that feel tailored to the customer. Most companies still haven't mastered social media, yet alone mobile messaging." (Morgan 2016) Amazon is one case where NLP analytics have been utilized in order to exceed customer satisfaction. It contains a vast and efficient marketplace of different functions where business problems are identified and solved through comprehension and deep understanding of customer wants, goals, and beliefs. Amazon has made progress building a compelling customer journey. Within the Amazon ecosystem contains a large and efficient marketplace which is innovating on all sides of the consumer equation. They involve virtual assistants with purchases customers make, and provide personalized content and experiences for the customer. By providing SaSS, "Amazon Comprehend and Amazon Transcribe services allow VidMob to build high-quality machine learning text analysis, enabling us to help brand clients understand content performance in ways never before possible. We are able to transcribe text from video content, and quickly analyze it using Comprehend." (Amazon 2020) This will help to provide the 'one on one' personalization that customers demand. Also it can provide understanding where and how to homogenise services, which would be important in the in any business function when services need to standardized.

A major challenge of customer service is exceeding customer satisfaction in a sustainable and continuous fashion. Sentiment analysis and topic modeling will be the synergetic tools that will be examined in this thesis that can provide this comprehensive clarity. Identifying data patterns will help to adapt changing wants, goals, and beliefs, as well as changing trends and external risks. Topic modeling, which brings order to unstructured data, and semantic analysis which provides semantic coherence and clarity, are approachable tools in this endeavor. It can help to provide a wide variety of actionable information, and filter everything that is not essential to creating value. "Analyzing product and customer behavior provides valuable insights into what consumers want, how they interact with products, and where they encounter usability issues. " (Accenture 2019) These insights can lead to new feature designs new products through understanding latent wants in the data.

As a tool and as a service, these tools can help to create a personalized '360 degrees' immersive experience (Gartner 2011:5). Increasing the elements related to flexibility of product, emotional content through story-telling, product personalization, and ease of use. (Accenture 2019) Data mining techniques and NLP overcomes many limitations of traditional statistic, enabling manipulation of dispersed data sets to acquire information." (Menon, Pradeep 2017). "It is possible to track relevant issues by identifying trending topics in real time. The challenges in these types of applications are the size and the quality of the data, making model evaluation a very important step before drawing conclusions." (Särkiö 2019: 60) Enabling greater clarity of unstructured data which is used to analyze customer behavior and maximize service value. (Provost and Fawcett 2013) In this digital landscape there are plenty strategies that companies may focus on in terms of marketing, corporate social responsibility, and LEAN management. However these are not leverage the freely available data. The issue is that most think that NLP is a very sophisticated tools which the average business or start-up will have difficulties to learn and take advantage of.

Continuous and sustainable innovation is difficult, but the most successful companies and individuals are those that continue to find methods to enhance business functions through new methods of data insights, and adapt accordingly to exceed all expectations. Empirical testing of these methods requires a clear strategy and experience that most businesses lack. This can be seen as a disadvantage and risk and a difficult task to quantify sentiment and the reality of what customers are doing (physical actions) versus what they are saying online (verbal feedback). Measurement of the theoretical constructs has repeatedly proven to be a difficult task (Boyd, Berg, et al 2013).

## 3.5 Effective Customer Service strategy

Thriving in today's market requires businesses to pivot strategy and operations according to real time intelligence gathered online. Companies that know their customer is a successful one. A sustainably competitive company will be able to anticipate customer needs and desires and do so quickly and decisively. (Leggett 2020)

This can be achieved through NLP, but it can also be hindered by NLP. "These hyper-connected manufacturing collaboration systems face the challenges of extracting, processing, and analysing data from multiple distributed web sources." (Lin, et, al. 2016: 18-23) The degree of complexity and layers of different systems need to be streamline to make progress toward more seamless and simpler customer experiences.' (Morgan)

Companies are gaining new product segments and customer insight from their data warehouses by applying new and more effective data mining techniques. Companies are unable to sustain competitive advantages outside of patents, copyrights, superior locations, proprietary information, etc. "Competitors are quick to copy any advantage through benchmarking, reverse engineering, and leapfrogging. Companies believe that their only sustainable advantage lies in an ability to learn faster and change faster." (Kotler 2000: 6-9)

Firms need to ensure they provide the necessary solutions to enhance customer service so that customers can rely on the business more. NLP augmented customer service creates opportunities to offload the human-to-human touch points when they're either inefficient or unnecessary. Allocating repetitive tasks including processing of unstructured data to automated systems will create space for more effective decision making solutions and growth. "It's become conventional wisdom that interpersonal skills like empathy will be one of the roles left to humans" (Dzieza 2020) The benefits of using the frameworks described in this thesis, and tools such as topic modelling and sentiment analysis, is that there is a relatively low learning curve involved for businesses and star-ups to leverage said tools effectively.

In order for companies to survive, building a customer aligned approach through the use of NLP strategy will help to increase autonomy and accountability. The global economy must meet the demands and expectations in every region. At the same time Customers expect issues to be resolved fast. The technology should allow the brand to focus more on finding and understanding the customer, regardless of channel—and allowing the agent to easily pop in and offer service." (Morgan 2016) Modern customer service augmented with NLP removes unnecessary friction and burden, creating a more optimized system. (Accenture 2019) "Employees want to be productive, but the

organization too often gets in their way. Our research indicates that the average company loses more than 20% of its productive capacity — more than a day each week — to what we call "organizational drag," the structures and processes that consume valuable time and prevent people from getting things done. These structures and processes that hinder organization daily operations can be identified through unstructured data." (Mankins 2017)

Successful businesses that wish to enhance productivity will seek to eliminate 'organizational drag'. Flattening organization's structure and aligning their operating model with NLP can help to understand true sources of value in their business. "This will create ways of working that allow employees to focus their time on delivering for customers and shareholders." (Mankins 2017) A streamline system is where businesses are seeking to identify bottlenecks in the daily operations and ways to optimize every 'cog in the machine'. Through the manipulation of data, business can understand how it can be utilized effectively to enhance business functions. This can be a important factor in modern customer service; exceeding customer satisfaction and beating the competition. To develop a visionary customer service and a more successful business it is important to have synthesized a coherent framework to realize these objectives.

# 4 Methods

## 4.1 Framework for Utilizing NLP

Lack of cohesive frameworks that connect business objectives with data analytics hinders the development of competitive advantage in customer services. The ability to extract information from text can help to improve effectiveness of daily and long run operations. Text annotation provides the basis for extracting recurring features or themes. It is regarded as an application area with high computational requirements (Nesi, P., G Pantaleo, and G. Sanesi 2015)Therefore businesses that lack the necessary frameworks will see this as a substantial threat for the competition that is able to make use of feature extractions such as topic modeling. Businesses need to leverage frameworks that will process data and analyse it through a unified workflow solution.

This will determine how successfully businesses can add value through NLP. The main challenge with utilizing NLP is, despite widespread application of NLP, it still lacks support and guidance from a cohesive framework. (Zhou, L. and D. Zhang 2002) A methodical process from start to finish needs to be in place for extracting insights from data.

This process should have clear and distinct stages with clear deliverables in terms of empirically quantifiable success. Otherwise putting structure and filtering data will only confuse and misguide the business looking to use unstructured data as a guide for understanding and decision making. As the scope of a business's objectives becomes clearer, they can start to collect data using a framework such as Hadoop to process and integrate it into CRISP DM. The Cross Industry Standard Process for Data Mining (CRISP-DM) is one example of a working model to integrate with business functions that assist in problem solving and enhancing business functions. (IBM) Hadoop has the ability to analyse unstructured data present in different machines and databases at different locations. It is a Java based programming framework that quickly and cost efficiently processes data using an environment known as parallel programming.

The Hadoop Ecosystem System is being adopted to reduce the complexity of moving data to and from the big data cloud platform in a very cost effective way. (Lin, et, al. 2016: 18-23) It is most useful and effect in the process of transforming the data from its different modes into uniform data. Businesses manipulate, store, and process data with Hadoop, as it is used as the main analytical framework. This is due to the strong community recognition, and limitless knowledge hub; where all algorithms and scripts are available for free. Since Hadoop programs are run across individual nodes that make up a cluster, these clusters provide a high level of fault tolerance reducing risk since the framework can transfer data from failed nodes to other nodes. (Miguel, J., S. Caballé, and F. Xhafa. 2017) It is favoured among data scientists and artists to prepare data which is increasingly used by all business intelligence functions due to its flexibility, scalability, fault tolerance and cost effectiveness. Along with clear KPIs (key performance indicators) using the framework process showed in figure 1. (NIST)

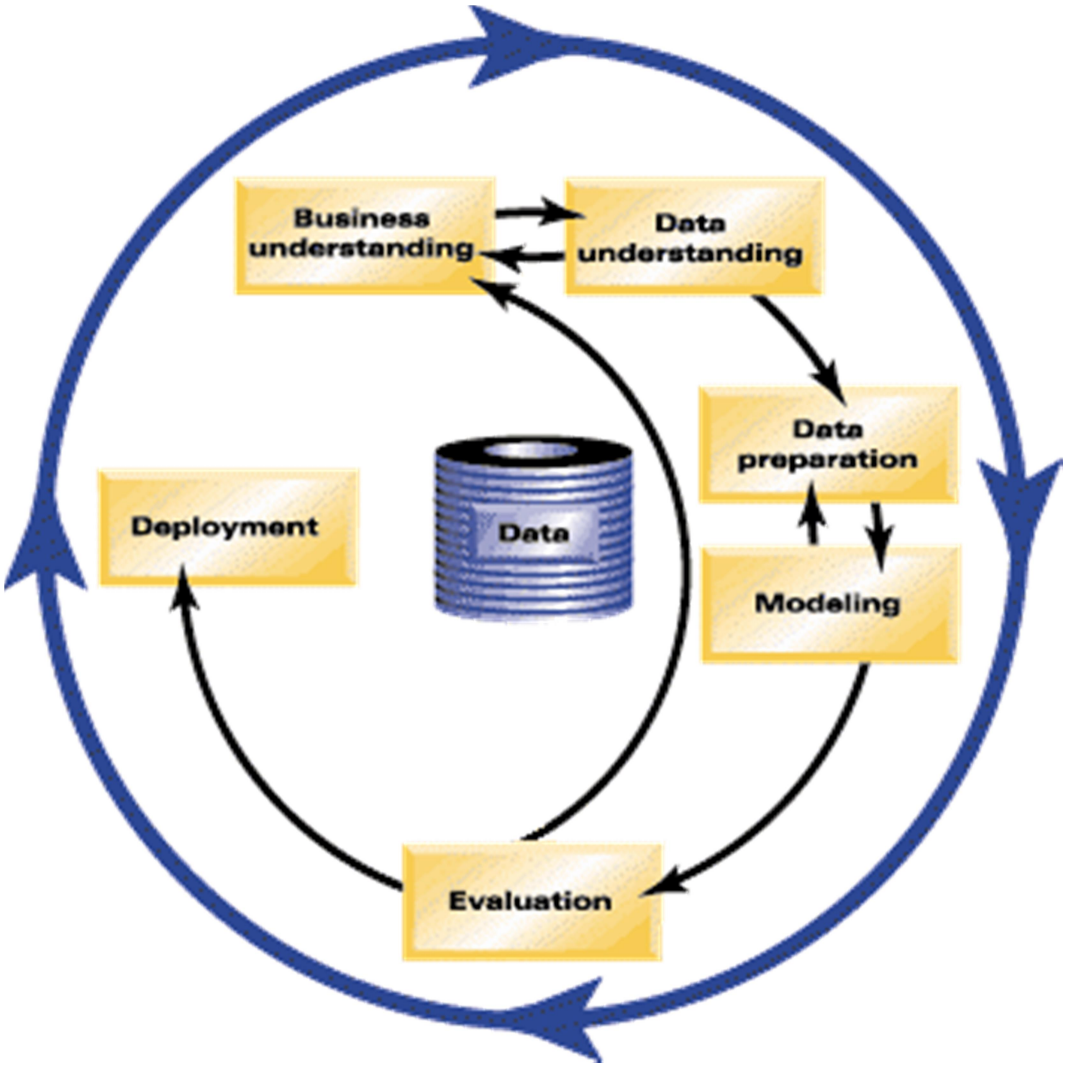


Figure 1: CRISP Process (IBM 2012)

To develop a unique selling point or sustainable competitive advantage to enhance customer services, businesses must begin with how data is used. How it is used and how it is implemented into the iterative process illustrated in figure 1. Businesses will go through different stages according to their data collection methodologies and how this relates to the understanding of one's own business objectives. Data ingestion is the first step and it is the process of data collection and import into the system either for storage or database building, moreover for further data analysis. (Krstić, et. Al 2019) This enables the conditions to gain business and data understanding for further action. This kind of data can be extracted from social media channels, government databases, internet discussion forums, etc. "The high quality of the source data is a

precondition for high quality results, as is the case in any data analysis." (Särkiö 2019: 60)

The outcome of the project is only going to be as effective as long as there is accurate data in relation to the scope of the business problem. This can be determined as a risk, because of the difficulties associated with defining what high quality data is, how it can be found, and in what context can it be leveraged. Asking these questions in the beginning of CRISP DM will bring structure and test any underlying presumptions or bias. This might aid in clarifying what data represents due to outliers, anomalies, and variables. It helps to formulate the initial impression on which businesses can begin their initial push for collecting actionable information to enable competitive advantage.

By identifying the business problem and setting clear objectives and quantifiable goals, the business can find the solution which best fits the given situation. Framing the data understanding into a business context will allow business to determine what kind of KPIs or metrics can be used. Without standards to benchmark against or success metrics, the business will not know how to gauge and perceive their success within this framework. This is also a risk or drawback to using NLP. Businesses need to clarify what the problem is and collect data according to this. To translate a business problem into an AI and data science solution, you need to understand the problem." (IBM) Understanding the problem can develop initial hypothesis generated from the data and the models. Furthermore, help to examine any weaknesses or faults within the data processing techniques. These additional findings will strengthen the businesses analytical experience in the future and help to make more informed decisions with confidence.

Without understanding of how the data can be used for business functions, and how one can leverage any available data, there will be a lack of clarity when moving along this framework. The two arrows pointing parallel in figure 1 accentuates this concept. Both the data and business understanding should either be equal or meet each other for this process to work. After defining business problem and identifying the topics that can be solved with machine processing, business can move to the next stage of data preparation.

Data preparation's main function is to extract, store, process, and move data to be used for further functions. This is where Hadoop is utilized. After the data is stored then it can be moved to data management systems using automated commands that execute this function. This is also known as parallel computing. Hadoop splits programs and data across many nodes in a cluster. Raw data, unstructured, structured, or semi-structured, is stored on the HDFS (Hadoop file system). After this processing stage, businesses can start to use data mining and business intelligence tools to gain access to insight and relationships within the data using free to use or low cost third party APIs. (Application programming interface) Some examples of effective APIs include Google Cloud Platform, Amazon Comprehend, Teradata RACE (Rapid Analytic Consulting Engagement). There is a diverse range of business providing APIs and consultation, which will be discusses in the analysis and results section. Further technical explanation of Hadoop frameworks will be omitted from this thesis in order to focus the scope of it within the context of business and developing competitive advantages. Businesses will benefit from developing systems that utilize CRISP DM as a business data framework in parallel with analytical framework Hadoop.

Next begins the next phase of CRISP DM which is data preparation. Businesses can use a systematic approach of performing statistical and visual analysis to see what the data's characteristics are beyond formal modelling and hypothesis testing. (Menon, Pradeep 2017)

Data preparation can be referred to as 'Exploratory Data Analysis (EDA). "It is different to statistical graphics although the two terms are used almost interchangeably. Statistical graphics is a collection of techniques. All graphically based and all focusing on one data characterization aspect. EDA encompasses a larger venue; EDA is an approach to data analysis that postpones the usual assumptions about what kind of model the data follow with the more direct approach of allowing the data itself to reveal its underlying structure and model. EDA is not a mere collection of techniques; EDA is a philosophy as to how we dissect a data set; what we look for; how we look; and how we interpret." (NIST 2012)

Business activities in this stage are often iterative as the need for new data arises to support the modelling and evaluation stages. The outcome of data preparation is a

data set that is to be used in the first iteration of modelling. (Spalek 2018: 177) A risk in this critical stage in the CRISP DM process is that businesses will rely on the third party partners and not take responsibility for their own data preparation. This will inhibit complete data and business understanding, which can reduce the effectiveness of this framework.

Through investigation of the nuances within the data, the first conception of what kind of NLP applications and models can be deployed. With deployment of these developed models, experimentation of different models to choose the model that best fits the scope is paramount in the development of understanding the data. This is an iterative process. This means that it can be helpful and usually necessary to take a step back and start over from the beginning of CRISP DM with greater understanding.

Enterprises can leverage NLP applications and modelling in several ways. Businesses can deploy different models in order to achieve greater business and data understanding. This can add value to the business and the customer. The simplest is query understanding and content understanding with the objective to improve both user experience and generate deeper understanding of customer's wants, goals, and beliefs. This understanding is developed through customer's natural language query inputs in the form of text and speech. "This provides better, more targeted responses by understanding the user's questions and intent. It assists in identifying out-of-scope requests and present intelligent alternatives." (Accenture 2019) These models should be continuously monitored to observe the mechanism and behaviour in relation to the prepared data.

An advantage of using this framework is that a test environment can be piloted before deploying into the real world. This can be piloted on a focus group that will mimic the forecasted environment. By working in this way, the next phase after deployment where evaluation is made can work in parallel with deployment of the model. Furthermore, when the insights and understandings are acted upon, evaluation can be made in real time by collected unstructured data again. This is extremely beneficial to finding alternative solutions and going back between stages to re-calibrate, adjust, and adapt accordingly. During the evaluative process, it is crucial for business to re-examine whether the given model, current understanding, and deployment of models

can actually help the business become more successful and create value. Otherwise this framework will become abstract and not practical, which is one factor that can be subject to criticism. Moreover, wants, goals, and beliefs of the customer, and the business, should be re-calibrated accordingly to the new information generated. Otherwise, business can be in danger of acting upon data which is out of date.

The process of using new insights to make improvements within the organization is the objective of CRISP DM. During deployment phase, this can mean "using the insights gained from data understanding to elicit change in your organization" (IBM). Impact of this decision making process is determined by classical metrics and KPIs such as return on investment, customer retention rate, and customer active engagement. Some of these indicators can be gathered through unstructured data, and some have to be observed through financial accounting systems. This can help to evaluate whether the process has helped the business become more successful. Other perspectives of success can be taken into account. Business can observe whether the product or service they are offering is matching customers' expectations through sentiment analysis through customer complaints.

"We use the concept of measurement perspective to define and categorize how the improvement is being assessed. Concretely, a measurement perspective describes the view on the improvement, i.e. which entities are measured in order to make the change visible in either a quantitative or qualitative manner. We derived from which measurement perspective [sic] an initiative is evaluated by interpreting the metrics which were described in the study and from the attributes they are supposed to measure."(Unterkalmsteiner, et al. 2011: 7-8)

There are various factors described of what makes this a dynamic and flexible framework. However, every framework and methodology will have its limitations and drawbacks.

### 4.1.1 Evaluation of bias and feasibility of CRISP DM Framework

Methodical assessment and evaluation through each individual stage in CRISP DM framework can help to determine inaccuracies. A business must consider the source of the gathered data, and this will help to determine quality of data, and data bias. Data bias can be viewed as ones propensity to be inclined to believe in the collected data. This can be because it is in favour of their own interests or agenda (the data's can show something they wish to be true), or even because of the difficulties it took to procure it. A way to minimize this is to take a larger sample size for the data from a better source and do a longitudinal study. This can help to discern patterns and justify truths to a higher degree of confidence.

Project bias also needs to be taken into consideration. This can occur when business owners with limited knowledge of how data analytics create incentive to push a data based project in hopes of gaining any kind of profit from it. This starts the project off on the fallacy of high success rates for data analytic projects. The bias can be mediated by stressing clear KPI's and business objectives. Also whether or not NLP based strategy can fit with the business infrastructure. This assessment can help to determine different biases and conflicts of interest. It can also allow one to identify which predictive modelling techniques are best leveraged in the scope of the business problem with regard to outliers or anomalies present within the data.

Learning bias occurs when the actionable information that was gathered for model deployment interferes with the evaluation of its benefits. (Unterkalmsteiner, et, al. 2011: 7-8) The only solution for this would have to be allocating correct responsibility that matches the expertise for those that are doing the evaluation, and have it be a third party, or have everyone working on the project on the same level of expertise.

Participant bias occurs when the data gathered from the participants lacks volume and depth. This occurs frequently during poorly conducted surveys. For example, thesis surveys where samples and results are taken only from the university where the author studies at. This creates a skewed perception of the reality. The author works off this data because it is the only primary source of data available to them, and then bases their whole thesis on this data. The solution to this is not easily solved. However a

method can be applied where data is collected according to randomly generated allocation algorithm that allows for a diverse range of samples to be taken.

Over reliance on complicated mathematical algorithms that are not fully comprehended by the business entity should be avoided unless otherwise guided by third party solutions and expertise. These algorithms involve neural networks and deep learning, which steps into the field of artificial intelligence. The focus of the data should bring the business closer to its objectives and enhance functions in a way that can be quantified. "Bringing in technology too soon can guide the solution to a technology, and the actual business problem might be forgotten or not fully answered." (Eunice. Biddle and Christensen 2020) Methodologies that use quality data and simple models based on high volume of data, that keep their scope narrow will avoid most of the limitations and drawbacks of NLP.

Developing clear business objectives will always assist in clarifying this framework. At each stage and process, the business is should reflect on how this new information can this help to solve the scope of the problem. Evaluative efforts should be stressed because it helps to frame what data does for the business, and whether it can be a competitive advantage or a risk to fail. The evaluation stage illustrated in figure 1 will help business to take into account the external environment it operates in. These include barriers to entry, threat of new entrants, bargaining power of buyers/suppliers, and threat of substitute products. (Porter 1979) Evaluation of these forces will increase the likelihood of creating data-based solutions that will add lasting business value.

The risk of integrating such analytic technologies to solve business problems is that businesses can find correlation that does not result in quantifiable success. Constant evaluation of a clearly established quantifiable criterion of success will help to mitigate this. CRISP DM process has the strength of a feedback loop where the business can move between different stages, offering flexibility when new knowledge or data is uncovered. "Feedback Loops exist whenever the output of a system becomes one of the inputs in the next cycle. Feedback is how systems learn — if the system is capable of perceiving its environment, that feedback helps the system understand whether it's under control and satisfying the required selection tests." (Kaufman, J. 2012) When preparing and analysing data, there often is greater understanding of the customer or

business which leads to revising of different stages of the decision making process. This will reshape the final decision making process and help the business leverage the data to a greater degree. With these biases and success metrics clearly understood, a business can expect to develop sustainable competitive advantage. Moreover, one can expect to gain greater understanding of the customer, as well as the inner workings of one's own business.

As customer understanding deepens, this knowledge and content will find the customers. This is done by providing the necessary knowledge and content collected going back to the customers through the different frameworks and methods described. An informed customer will be inclined to give more feedback if they know that the feedback is being properly utilized to create value. This should help business to develop ethical and effective strategies. Moreover it can be a component for a solid unique selling point that the competition cannot beat or mimic. This framework creates the correct foundations for leveraging NLP. Nevertheless, those that wish to integrate NLP will require the understanding of the context and perspective in which it works best through use cases and comparative studies.

## 4.2 NLP Integration into Customer Service

As organizations start designing and building NLP applications, it's essential to ensure implementation partners have the expertise required to conduct thorough assessment. This assessment should take into account the different stages of CRISP DM. Moreover, it should be considered whether NLP is compatible with the company's existing strategies and objectives. Whether the organization is only in the initial phase of evaluating various NLP technologies, or has already identified the preferred solution for implementation, keeping a clear business strategy to penetrate the market depends on having a framework to realize this. It is impossible to cover all the details of research in topic models, semantic analysis, and machine learning, but this section introduces the common techniques and use cases of models used. High level concepts such as measuring innovation, cross-language connection, and sentiments can be made with a simple dynamic topic model. This can be constructed by separating a body of text into sections and estimating the probability distribution over words for each topic in each section. "Building and validating custom topic

models is a powerful tool, but requires significant investment in coding and debugging, and may not be able to take advantage of computational optimizations available for simpler models." (Boyd-Graber 2017: 112-114)

Topic modeling is one of the most effective solutions to understanding content. This is in part due to advantage of stacking multiple models and functions to gain greater insights to assist in the problem and decision making. (Accenture 2019) One case where is where businesses are able to capture context regarding customer service feedback, and then applying multiple forms of analytics to take real-time and targeted action. The result is improvements to the consistency and quality of customer services that "delivers measurable business impact." (Opinionlab) Manual analysis of vast amounts of data is time consuming and would not provide the results that NLP provides in terms of actionable information in a short amount of time. Topic modeling and opinion mining is compatible with the framework process described in this paper, where data is prepared, a model is trained with this data, a scored document is provided, and evaluation of the performance is made. The measurement of semantics as similarity in patterns and latent variables (factor analysis) has been enhanced by the use of statistical methods. (Hecking, Leydesdorff 2018) A combination of several of these statistical tools such as Bayesian analysis in combination with NLP can be beneficial for developing reasoned justification. This is done by drawing on evidence based conclusion, probabilities, on the hypothesis in question. This should be leveraged only once the quality data is available. These tools in combination may help to provide greater depth for actionable information.

The objective of topic modeling is similar to that of "co-word mapping" and "semantic maps" techniques. These techniques allow large quantities of texts to be organized into context. Facebook, Twitter, Instagram, Tripadvisor, etc. all have large volumes of data available for business to leverage to their own advantage. Databases have been used by scholars and researchers to develop virtual agents for government infrastructure. This is currently being implemented successfully in Finland for social service, urban strategy, and healthcare. Analysis of customer feedback is done in a rapid feedback system with artificial intelligence. Automatic keyword utilization with text analysis and artificial intelligence is used with the goal to enhance LinkedEvents experience platform and provide events that interest the city's residents even better. The Linked Events

interface categorizes events using keyword naming. To support this, the project is developing automation. Artificial intelligence, or NLP algorithms, are also used to analyze the quantitative feedback of the rapid feedback system's health stations and oral health care, and to classify the feedback according to time, day of the week and season. The lessons learned in this project can be utilized in the analysis of large data sets. Furthermore, the project compares customer feedback with service availability information. (Meloni, Rautio 2019)

Another example showcases of NLP integrated to customer services is an intelligent recommender for cultural and leisure services. The aim is to create an intelligent recommender for the city's cultural services that learns to respond to user needs and interests based on past online behavior and data sources. The customer understanding, accuracy and “emotional intelligence” of the referral service also increase with use. Furthermore Finland is also using text analytics help to implementation of the urban strategy. Make it easier for content producers with the objective to use text analytics to get a better overall picture of how we as a city, can communicate the implementation of the urban strategy. Text analytics can be utilized in the city’s many different channels so that artificial intelligence can identify which theme each content is related to. (Meloni, Rautio 2019)

The utilization for topic modeling as a technique to summarize large amounts of documents in the format of a limited number of words or graphic illustration is not always justified because of the validity problems which are inherent to this methodology. The probabilistic character of the results may easily lead to misunderstandings outside the context of the production of these models. (Hecking, Leydesdorff 2018) However, businesses and programming communities are enabling topic modeling and sentiment analysis with user-friendly interfaces. These are easily applicable for non-experts, and it is essentially a sorted list of words with their probabilities. It is useful to label things to have a unique ID to refer to and compare with, as this is the most efficient way to process large amounts of unstructured data to have actionable information for business to contrast, compare, and utilize to their own advantages.

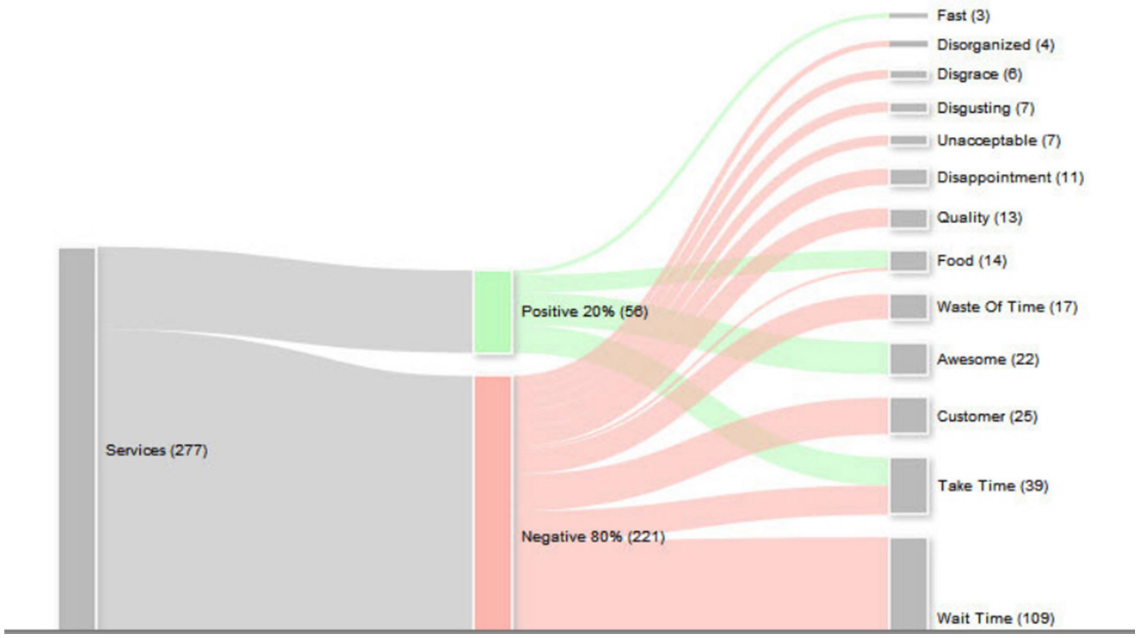


Figure 2: Diagram illustrating sentiment feed-back (Zaino 2014)

The diagram illustrated in figure 2 covers the sentiment analysis of 277 different inputs of data. It tells a story through positive and negative bars that reveal which has the greatest proportion of comments, as well as the various subtopics in each thread. Companies that have lost sight of the importance of human interaction and often make it too difficult for consumers to get the right level of help and service that they need. Sentiment analysis is one tool that modern customer service agents will identify as a solution to gauge customer satisfaction, as illustrated in Figure 2. Business will benefit from time-savings advantage and the ability to show their information more precisely and succinctly to stakeholders involved. This can help the business to identify what part of the service needs to be evaluated in order to find the business problems, and use analytics to enhance it.

In its simplest form, sentiment analysis categorizes a sentiment as a binary positive or negative value. It could also quantify the sentiment (e.g. -1 to +1). "We have multiple pre-defined notions of how to represent a quantity of interest. Sentiment is sometimes represented as a continuous positive or negative value" (Boyd-Graber 2017: 112-114) It can categorize it at a more granular level, as shown above, disappointment, unacceptable, awesome, quality, etc. 'Oftentimes, companies with the best intentions find themselves in an insights vacuum, sentiment analysis provides some answers into

what the most important issues are, from the perspective of customers.' (Dumbleton 2018) Understanding of customer reviews and feedback including business and consumer interaction will provide greater insights into understand the customers wants, goals, and beliefs.

With these methods, decisions can be made based on effectively piloted models, rather than businesses past experience and intuition. One is able to use sentiment analysis to focus on the customer feedback where the sentiment is strongly negative. Moreover, one can look at customer comments which have strongly positive sentiment to find out what prompted this reaction, and adjust accordingly. When used in combination with topic modeling, one can further narrow down this information to find precisely which topics are talked about with positive/negative sentiment, providing actionable insights for businesses.

From the perspective of reliable decision making, interpreting and evaluation of the models in the context of adding value to society and the customer is critical. These NLP models can help to provide an equivalent to catalogue systems with statistical probability distribution, but can be observed to contain a higher degree of objectivity because the generation is computer-assisted with inputs being quality information from a variety of sources.

# 5 Analysis and Results

### 5.1.1 Developing Sustainable Competitive Advantage Using NLP

In order for companies to survive, investment into the field of NLP can be a critical element in determining customers changing, wants, goals, beliefs. Most companies tend to stick to their values and only adopt new technology where there is a need to or when one finds that the benefits outweighs the risks. "We're at an awkward inflection point where some companies are doing an amazing job of being on the forefront of customer experience technology, and others are still struggling with the basics. Companies are slowly making progress toward more seamless and simpler customer

experiences." (Morgan 2016) Analyzing unstructured data using topic modeling and sentiment analysis enables high value business outcomes. This can help to get a USP and a more successful business, and help to exceed and/or match customer satisfaction. All industries can use NLP tools such as text clustering (topic modeling) and analytics (sentiment analysis) to innovate, address quality or safety issues, or improve customer service. (Teradata 2018)

In driving competitive advantage, operational excellence achieved through matching customers' expectations might help the firm succeed in short term objectives, but other firms will surpass through innovation or exceeding customer expectations. Understanding content through the framework and tools presented in this thesis can help to create a strategy to differentiate from the competition. This kind of strategy is continuously being implemented in large and small scale business such as Utopia Analytics, Facebook, Google, Amazon, Banks, and government bodies. Dex uses a novel approach to NLP which utilizes a similar framework to CRISP DM.


Figure 3: Flowchart 'Four easy steps' (Dex 2020)

The consultation phase 1 in figure 3 provides the initial business and data understand of CRISP DM. This is where KPI are established through defining objectives and what the business intends to achieve. One case study by Dex shows the evidence of this in action. The requirement of the client was to increase the capacity for businesses to be able to integrate various data science tasks through develop and deployment of text mining tool kits. With these tools, businesses are easily able to develop client database warehouse with personalized information concerning customer preferences and expectations. These were used to build models to find relationships between people. The research phase, parallel to the data preparation phase of CRISP DM prepared and processed annual reports of more than 50,000 documents. "The team has implemented new approaches to the process of extracting the necessary information

from a large amount of unstructured data of financial documents. Automation of the extraction process accelerated the finding of 90 parameters from tables and texts, 15 times in comparison with manual processing. This also allowed for a reduction in budgeting by 2.5 times. (Jenkins 2020)

The deployment phase highlights the effective aspect being able to layer frameworks and stacking functions on top of existing business infrastructure. Dex has created frameworks to develop solutions which allowed recognizing parts of contracts and their structure. The objective was to create a framework that could be applied to a wide variety of legal documents for the purpose of identification and extraction of text from images. This assisted in finding potential loopholes or advantages relating to their clients cases. The project was completed during a 6 week period. As a result automated systems took over routine manual work and saves $60,000 per year. (Flynt 2020) Automation excels in where humans tend to falter in repeated judgment jobs. 'Any slack is perpetually being optimized out of the system.' (Dzieza 2020). This is where a business is readily able to find lasting value and automate effectively.

Quality controls through evaluation as well as flexible infrastructure are two factors which enabled Dex to build a wide range of automated business processes. The development phase in the figure 3 flow chart emphasizes quality control. These allow clients to increase revenue and reduce cost. The use for this was developed for a scalable system to predict the behavior of the financial markets. A solution for analysis channels of communication developed. This was used to find relationships, building causal links between significant world events. Scalable architecture was used was a large database was constructed from a wide range of news outlets. In the course of this 1 month project, a lot of issues were discovered. The issue with the similarity of semantic units was a challenge. This is because the same words can suggest different meanings depending on the context.

Another challenge was that it necessary to balance and distribute the unstructured data for the deployment phase. Hadoop ecosystem is a useful ecosystem to implement to solve this challenge. This was done in order to process the data as quickly as possible. As a result Dex was able to increase the company's revenue by 4% in 1 month. Moreover, they have developed a solution that has allowed for the discovery of hundreds of thousands of customers for companies over a 24/7 period. (Dex 2020)

Businesses are doing a better job of retaining their customers through finding imaginative ways to exceed customer expectations with NLP. They use new information to 'mass-customize' their offerings to individuals. (Kotler 2000: 6-9) A fully automated cloud solution was established to control lead generation. It enabled the function of being able to upload data of various formats concerning lists of potential customers. The project period was 2 months. The solution increased sales by 40% and allowed the client to find new points of sales growth. Due to rapid growth, the client created new teams in new locations and implemented a new sales model based on insights obtained from the platform. The client was also able to manage integration with third-party services for convenience. (Dex 2020) As a result competitors have found it increasingly difficult to acquire new customers.

The data for processing and analysis was taken from the world's largest news agencies. Since the client's IT department regularly works with NLP, the Requirements had to be able to process complex unique data sets and be easily integrated into existing infrastructure the task was to increase ease of use and enhance personalization for the working environment of NLP developers.

"Almost all available tools for working with code and NLP were found, collected, categorized and packed into a single technical solution, which allowed to speed up the development process and significantly reduce costs for the company's IT department. Our team solved the issues of integrity, compatibility, GPU driver related problems, memory consumption optimization and picked an optimal set of tools for the toolchain. As a result, the customer received a preconfigured machine learning and deep learning solution for natural language processing applications. The toolkit was able to collect, map, transform, translate, encrypt, analyze, visualize and extract knowledge from text. The toolkit made developers' jobs easier, as they no longer have to look for necessary tools. The solutions saved 66% of time for new project deployment – 15 minutes instead of 45." (Reilly 2020)

Insights created with topic modeling can be used to improve business decisions and enhance customer experience. Another use case is shown with Utopia Analytics. Finnish startup Utopia Analytics has created a globally applicable solution to moderate hate speech, harassment and other inappropriate content. "Automatically in real-time with AI, Utopia Analytics frees staff for more complex tasks, helping service providers

to save costs and to implement their moderation policy in a consistent and effective way." (Sallinen 2019) This start up is provided funding by the Finnish government and has built a globally applicable solution that attracts foreign direct investments and scalable SaSS. "By sustaining higher quality content through online discussions, Utopia AI Moderator is able to enhance customer services by making platforms and websites more attractive to both users and advertisers. This increases active engagement metrics, and improves KPIs such as conversion rates." (Sallinen 2019) While NLP is a relatively new enterprise technology, it's being enhanced every day. The Machine learning algorithms supporting NLP are improving all the time, with industry giants like Google, Microsoft, and Amazon all investing to improve accuracy and effectiveness.

Fully personalized solutions, easy to implement tools and services developed by Utopia Analytics stand out as a USP in the global marketplace. Moreover, Utopia and other AI driven data-based start-ups benefits from funding, networking, and grants, provided by business programs, universities such as Aalto, and government infrastructure. With this growing ecosystem of NLP solutions, enterprises have increasing flexibility to select appropriate models and toolsets in open data communities. Businesses might have difficulties choosing the solution and adopting new technologies which best fits their objectives, however, text analytic tools such as topic modeling and sentiment analysis provides excellent ROI for minimal cost.

These analyses can be further expanded with other models not discussed in this thesis such as, LDA models, named entity extraction models, and churn prediction models. Moreover, Textual data and these models describe can be used to improve the accuracy of already available models in financial organizations, like credit risk models. (Krstić, et al. 2019). Vehicle manufacturers are using data integration through NLP to improve customer service and deliver an elevated customer experience. By learning what customers like and dislike about current products, companies can improve their design, such as adding new features to a car to enhance the driving experience. (Teradata 2018). "Companies with call centers can identify common complaints by monitoring customer calls. Financial advisors can review emails to ensure compliance. Utility companies can identify safety issues with a utility grid. Text clustering also allows retailers to quickly understand what product lines their customers are excited about on social media. Retailers can observe text clusters to see specific issues, which

can reveal customers wants, goals, and beliefs. Understanding customer sentiments empowers companies to make more informed decisions" (Teradata 2018). This can help in creating new products or refining the old, help to enact new policies by government bodies, and develop solutions that meet customers evolving needs. NLP is the critical factor in enabling sustainable competitive advantage if the limitations and biases concerning data preparation and frameworks are taken into consideration. Businesses that are unwilling to adopt new systems and technologies will result in circumstances where time, safety, and ability to match customer satisfaction in comparison to the competition are adversely affected.

### 5.1.2 Biases and Limitations of Analysis

There are additional functions that can be used to bolster the effectiveness of NLP that involve statistical science which the author has not discussed to keep the scope of the thesis narrow and concise. By addition of Bayesian inference and injecting it into sentiment and topic analysis within the framework of CRISP DM, models can become more robust when used to update the probability of a hypothesis as more evidence and information comes available. The limitation of this thesis is that there is insufficient secondary research where a longitudinal study of two businesses could be used to verify the effectiveness of NLP.

Since the objectives and frameworks of companies are rarely identical, moreover the methods to gather data to utilize NLP for enterprises differ greatly, the quality of a comparative case study would have been ineffective in answering the main research question. To contrast and compare a comparative case study of two companies, one that uses NLP, and one that does not, and use the financial statements and their evolution over time would have been the next step in the authors synthesis of a masters or higher level dissertation and illustrate the main research question in greater depth.

The authors own experience and evidence aggregated in this paper can be criticized to be only scratching the surface of what knowledge and expertise is required in deploying effective NLP models.

The uses of machine learning algorithms that can be used in conjunction with NLP have not been discussed in a quantitative case study. These are what top business entities are leveraging to enhance business functions with high success rates, using neural networks and adaptive algorithms which are modeled to mimic natural organism and higher level adaptive modes of decision making. The justification for not constructing primary research is that the objective of this thesis was to keep the scope of building a framework and using one or two simple tools to show how business are using NLP in a business context to enhance decision making. One of the primary challenges of deploying effective NLP models is making them more simple and accessible to everyone. This thesis has been able to assist in this endeavor.

# 6 Conclusion

To design a data project framework that will continue to provide accurate results is what the leading data scientists strive for. Businesses are complex systems with the inherent risk that eventually the system and models will fail. The more complex a system is and the longer it operates the higher likelihood of encountering uncertainties and changes from the environment it operates in. Creating an automated ecosystem with a clear framework will help to develop safer and more accurate systems.

The value of information lies in the value of the decisions they inform. This is what NLP helps to generate, greater value in the information to create accurate decision making and enhance business functions. Having clear, measurable objectives with check-ins every week, month, quarter, or all of the above will aid in refocusing the scope of business objectives and allow for evaluative readjustments. Precise KPI's will aid in tracking business performance. This will help to determine whether the actionable information derived from NLP brings any success.

NLP is a tool. Tools require a supportive environment. A supportive environment means clear understanding of inputs and outputs in a suitable framework. It requires a definitive objective which can be evaluated and adapted according to changes within the environment. This will help to derive deeper understanding and create lasting value. It will create a business which can leverage resources effectively and generate

valuable actionable information. NLP has its limitations and advantages. Limitations and advantages considered and offset, businesses can expect to have great success with integrating NLP into their business strategy. In the scope where automated technology will not be able to quantify or act upon, leaders and managers should make intuitive decisions based on their own experience and all the knowledge they have at their disposal.

## References

Accenture. 2019. Top Natural Language Processing Applications in Business [Online] Available at: https://www.accenture.com/_acnmedia/pdf-106/accenture-unlocking-value-unstructured-data.pdf [Accessed on 19 December 2019]

Ahimbisibwe, et, al. 2015. Pg. 7-33 A Contingency Fit Model Of Critical Success Factors for Software Development Projects. Journal of Enterprise Information [Online] Available at: https://www.researchgate.net/publication/328980085_Ahimbisibwe_A_Cavana_RY_Daellenbach_U_2015_A_Contingency_Fit_Model_Of_Critical_Success_Factors_for_Software_Development_Projects_Journal_of_Enterprise_Information_Management_Emerald_28_1_7-33 [Accessed 10 January 2020]

A. Thakur November 14, 2018 How Chatbots Work and why you should care. [Online] Available at: https://www.boost.ai/articles/how-chatbots-work-and-why-you-should-care [Accessed on 28 April 2019]

Amazon 2020. Discover insights and relationships in text. [Accessed 12 April 2020]

Boyd, Bergh, et al. February 2013. Constructs in Strategic Management. [Online] Available at: https://www.researchgate.net/publication/258174101_Constructs_in_Strategic_Management. [Accessed 13 4 2020]

Boyd-Graber. 2017. Applications of Topic Models Pgs. 112-114 [Online] Available at: https://mimno.infosci.cornell.edu/papers/2017_fntir_tm_applications.pdf Accessed 14 April 2020

C, Shannon. 1949. A Mathematical Theory of Communication. [Accessed 4 January 2020]

Dixon, Hayley. 7 January 2019. Data now as important a commodity as oil [Online] Available at: https://www.telegraph.co.uk/news/2019/01/07/data-now-important-commodity-oil-leading-qc-says/ [Accessed on 26 November 2019]

Dex (2020) Easy steps to having an awesome AI. Digital Image [Online] Available at: https://dex.re/ [Accessed on 14 April 2020]

Dex (2020) Dex Technology case studies [Online] Available at: https://dex.re/case-studies/ [Accessed on 15 April 2020]

Dretske, Fred. 1981 Knowledge and the Flow of Information [Accessed on 12 April 2020]

Dumbleton, Rob. 19 December 2018. Sentiment analysis [Online] Available at: https://getthematic.com/insights/sentiment-analysis/

Dzieza, Josh. February 27 2020. How hard will the robots make us work? [Online] Available at: https://www.theverge.com/2020/2/27/21155254/automation-robots-unemployment-jobs-vs-human-google-amazon

Efraim , Lee, et al. 2000. Electronic Commerce: A Managerial Perspective. Prentice Hall

Eunice. Biddle and Christensen. 2020. Translate a business problem into an AI and data science solution [Online] Available at: https://www.ibm.com/garage/method/practices/discover/business-problem-to-ai-data-science-solution [Accessed on 14 January 2020]

Feldman, et al. Jan 2007. The text mining handbook: Advanced approaches in analyzing unstructured data. [Online] Available at: https://www.researchgate.net/publication/200504395_The_text_mining_handbook_Advanced_approaches_in_analyzing_unstructured_data [Accessed on 19 March 2020]

Flynt, Williams. 2020. Identification and extraction of specific text blocks from images [Online] Available at: https://dex.re/portfolio/identify-and-extract-specific-text-blocks-from-images/ [Accessed on 14 April 2020]

Finkelstein, et al. 2009. Strategic Leadership: Theory and Research on Executives, Top Management Teams, and Boards

Forney, et al. 2019. An Agile Success Estimation Framework for Software Projects [Online] Available at: https://pqdtopen.proquest.com/doc/2287020617.html?FMT=ABS

Gartner. March 30 2011: 5. Customer 360 Summit.. [Online] Available at: https://www.gartner.com/imagesrv/summits/docs/na/customer-360/C360_2011_brochure_FINAL.pdf [Accessed on 24 November 2019]

Grant, Kenneth 2007: 173-180 Tacit Knowledge Revisited – We Can Still Learn from Polanyi. [Online] Available at : http://scholar.google.fi/scholar_url?url=http://www.ejkm.com/issue/download.html%3FidArticle%3D101&hl=en&sa=X&scisig=AAGBfm2KG2Z1LT6MYkV-wN9Dm0Gm0PwotQ&nossl=1&oi=scholarr [Accessed 15 April 2020]

Griffiths, A. and Wall, S. 2011. Economics for business and management. Harlow, England: Financial Times/Prentice Hall.

Gupta, Pankaj. 2019. Neural Information Extraction From Natural Language Text. [Online] Available at: https://www.researchgate.net/publication/336739252_PhD_Thesis_Neural_Information_Extraction_From_Natural_Language_Text/citation/download [Accessed on 9 April 2020] [Accessed 11 April 2020

Haponik, Arthur. July 31 2019. Reduce Operating Costs and Improve Efficiency Using AI. [Online] Available at: https://addepto.com/reduce-operating-costs-and-improve-efficiency-using-ai/ [Accessed on 4 April 2020]

Hecking, Leysdesdorff. 4 June 2018. Topic Modelling of Empirical Text Corpora: Validity, Reliability, and Reproducibility in Comparison to Semantic Maps. [Online] Available at: https://arxiv.org/abs/1806.01045 [Accessed 12 April 2020]

Dietz, Widdershoven. September 27 1991. Speech Acts or Communicative Action. [Online] Available at: https://dl.eusset.eu/bitstream/20.500.12015/2569/1/00055.pdf [Accessed 13.4.2020]

Jenkins, Tommy. 2020. Text mining from annual reports for financial organization. [Online] Available at: https://dex.re/portfolio/text-mining-from-annual-reports-for-financial-organization/ [Accessed in 15 April 2020]

Holmberg, Milka. 30 Oct. 2019 Boost Customer Experience with Conversational AI. [Online] Available at: <https://www.boost.ai/articles/building-a-better-event-experience-with-conversational-ai> [Accessed 28 February 2020]

IBM. CRISP Process, Digital Image, [Online] Available at:https://www.ibm.com/support/knowledgecenter/SS3RA7_15.0.0/com.ibm.spss.crispdm.help/crisp_overview.htm [Accessed 4 January 2020]

Kivisto, Mikko. 15 Apr 2018. Conference Call.

Kleinberg, Jon. 5 June 2016. Human Decision Making in Machine Learning Processes. Center for Data Science, NYU, [Online] Available at: <cds.nyu.edu/human-decision-making-machine-learning-processes>.

Kaufman, J. [2012 The personal MBA: Master the Art of Business. London: Portfolio Penguin.

Kotler, Philip. January 2000: 6 Marketing Management: The Millennium Edition. Chapter 1. [Online] Available at: https://www.researchgate.net/publication/235362523_Marketing_Management_The_Millennium_Edition [Accessed on 12 April 2020]

Krstić, Živko, et al. 2019. Visualization of Big Data Text Analytics in Financial Industry: A Case Study of Topic Extraction for Italian Banks. [Online] Available at: https://www.researchgate.net/publication/337903467_Visualization_of_Big_Data_Text_Analytics_in_Financial_Industry_A_Case_Study_of_Topic_Extraction_for_Italian_Banks?fbclid=IwAR33YYLZ0R73IHfyO-lPBfGPYEssI6hsAyJbQET41wh-u_xIa8j29ebjmbo [Accessed on 11 April 2020]

Miguel, J., S. Caballé, and F. Xhafa. 2017. Intelligent Data Analysis for E-Learning. Enhancing Security and Trustworthiness in Online Learning Systems. Amsterdam: Academic Press [Accessed 14 April 2020]

Leggett, Kate. Forrester Research. 2020. [Online] Available at: https://go.forrester.com/research/ [Accessed on 12 March 2020]

Lin, et, al. 2016: 18-23. A Hyperconnected Manufacturing Collaboration System Using the Semantic Web and Hadoop Ecosystem System. Procedia [Accessed January 26 2020]

Mackay. 1985. Advances in Psychology: A Theory of The Representation, Organization And Timing of Action With Implications For Sequencing Disorders Volume 23, Pgs. 267-308 [Accessed on 12 April 2020]

Mankins, March 01 2017. Great Companies Obsess Over Productivity, Not Efficiency. [Online] Available at: <https://hbr.org/2017/03/great-companies-obsess-over-productivity-not-efficiency> [Accessed on Nov 13 2019]

McKinsey April 2017: 9-11 Smartening up with Artificial Intelligence. [Online] Available at: https://www.mckinsey.com/~/media/McKinsey/Industries/Semiconductors/Our%20Insights/Smartening%20up%20with%20artificial%20intelligence/Smartening-up-with-artificial-intelligence.ashx [Accessed 20 October 2019]

Menon, Pradeep. 1 July 2017. Data Science Simplified. [Online] Available at: https://becominghuman.ai/data-science-simplified-principles-and-process-b06304d63308 [Accessed 10 April 2020]

Meloni, Rautio. 9 December 2019 Helsingin kaupunki aloittaa seitsemän tekoälykokeilua [Online] Available at: https://www.hel.fi/uutiset/fi/kaupunginkanslia/helsingin-kaupunki-aloittaa-seitseman-tekoalykokeilua [Accessed 14 April 2020]

Morgan, Blake. 18 April 2016. The Evolution Of Customer Service. [Online] Available at: https://www.forbes.com/sites/blakemorgan/2016/04/18/the-evolution-of-customer-service/#3592a9de2442 [Accessed 10 April 2020]

Nagel and Teodoridis. 11 September 2019. Jack of all trades and master of knowledge: The role of diversification in new distant knowledge integration [Online] Available at: https://onlinelibrary.wiley.com/doi/full/10.1002/smj.3091?af=R [Accessed on April 6 2020]

Neisser, Ulric. 2014: Cognitive Psychology. Classic Edition. Psychology Press.

Nesi, P., G. Pantaleo, and G. Sanesi. 2015. A hadoop based platform for natural language processing of web pages and documents" Journal of Visual Languages and Computing 31: 130-138 [Accessed 10 April 2020]

NIST, SEMATECH 2012. e-Handbook of Statistical Methods, [Online] Available at: http://www.itl.nist.gov/div898/handbook [Accessed on 26 February 2020]

Provost and Fawcett. 2013. Data Science for Business. O'Reilly Media. [Accessed on 10 April 2020]

O'Shaughnessy, John. "Competitive Marketing: A Strategic Approach".

Opinionlab. Context to Act. [Online] Available at: https://www.opinionlab.com/. [Accessed 13 4 2020]

Sallinen, Sami. 24 January 2019. Utopia Analytics' global moderation tools clean up the Internet. [Online] Available at: https://www.businessfinland.fi/en/whats-new/cases/2019/utopia-analytics-global-moderation-tools-clean-up-the-internet/ [Accessed on April 10, 2020]

Särkiö, Ilkka. 13 February 2019: 60. Topic modelling of Finnish Internet discussion forums as a tool for trend identification and marketing applications [Online] Available at: https://sal.aalto.fi/publications/pdf-files/tsar19_public.pdf

Scarborough, N.E. 2011. Essentials of Entrepreneurship and Small Business Management. New Jersey: Pearson Education Limited.

Schneider, Christie. October 16, 2017. 10 reasons why AI-Powered, automated customer service is the future. [Online] Available at: <https://www.ibm.com/blogs/watson/2017/10/10-reasons-ai-powered-automated-customer-service-future/>.

Searle. 1969. Speech Acts: An Essay in the Philosophy of Language. Cambridge

Siegel, Eric. 2016. Predictive Analytics. Wiley

Sloman, J and Garratt, D. 2013. Essentials of Economics. Harlow: Pearson Education Limited.

Slaats, Diego. 2013: 16. The digital transformation of customer services [Online] Available at: https://www2.deloitte.com/content/dam/Deloitte/nl/Documents/consumer-business/deloitte-nl-the-digital-transformation-of-customer-services.pdf [Accessed on 14 Jan, 2020]

S. Russell, P. Norvig. 1995 Artificial Intelligence: A Modern Approach Upper Saddle River, New Jersey.

RG Tian. 2011. Anthropological Approach to Consumer Science. High Plains Applied Anthropologist No. 2, Vol. 21. [Accessed on 7 January 2020]

Trujillo, Erika What Is Automated Customer Service? A Guide for Growth That Helps People Do More, Not Less. [Online] Available at: https://www.groovehq.com/blog/automated-customer-service [Accessed on 28 February 2020]

Unterkalmsteiner, et al. 2011: 7-8. Evaluation and Measurement of Software Process Improvement [Accessed on 10 April 10 2020]

Spalek, 10 October, 2018: 117. Data Analytics in Project Management. [Online] Available at: https://www.academia.edu/37828229/Data_Analytics_in_Project_Management [Accessed on 10 April 2020]

Teradata 2018. Text Analysis Quickly Reveals Customer Sentiment to Spark Innovation [Online] Available at: https://www.teradata.com/Resources/Case-Studies/Text-Analysis-Quickly-Reveals-Customer-Sentiment-to-Spark-Innovation [Accessed on 3 April 2020]

Teradata 2017. Deliver Innovation by Understanding Customer Sentiments [Online] Available at: http://assets.teradata.com/resourceCenter/downloads/CaseStudies/EB9859.pdf [Accessed on 3 April 2020]

Teradata 2017. Growing customer base with Network Experience Analytics [Online] Available at: http://assets.teradata.com/resourceCenter/downloads/CaseStudies/Growing_customer_base_with_Network_Experience_Analytics.pdf [Accessed on 3 April 2020]

Teradata 2017. Using Teradata Business Analytics Solutions to Deliver an Enhanced Customer Experience [Online] Available at: https://www.teradata.com/Resources/Case-

Studies/Using-Teradata-Business-Analytics-Solutions-to-Deliver-an-Enhanced-Customer-Experience [Accessed on 3 April 2020]

Yang, Liu. 2018. Deep Learning in Natural Language Processing. Springer.

Parasuraman, et, al. 1991 Pg. 39. Understanding Customer Expectations of Service. [Online] Available at: https://www.researchgate.net/profile/Valarie_Zeithaml/publication/225084143_Understanding_Customer_Expectations_of_Service/links/02e7e53cd16b876407000000/Understanding-Customer-Expectations-of-Service.pdf [Accessed on April 3 2020]

Porter. E, Michael. 1979. How Competitive Forces Shape Strategy. [Online] Available at: https://hbr.org/1979/03/how-competitive-forces-shape-strategy [Accessed on 15. April 2020]

Reilly. 2020, Brendan NLP toolkit for AI software platform. [Online] Available at: https://dex.re/portfolio/nlp-toolkit-for-ai-software-platform/

Verspoor, Karin, et al. 17 August 2012. Natural Language Processing. A corpus of full-text journal articles is a robust evaluation tool for revealing differences in performance of biomedical natural language processing tools. [Online] Available at: https://bmcbioinformatics.biomedcentral.com/articles/10.1186/1471-2105-13-207 [Accessed on March 3 2020]

Zaino, Jennifer. September 22 2014. [Digital image] Turn Your Text Analytics Findings Into Story-Telling Visualizations [Online] Available at: https://www.dataversity.net/wp-content/uploads/2014/09/SA-Chart.jpg [Accessed 13 4 2020]

Zhou, L. and D. Zhang. December 2002 "NLPIR: A Theoretical framework for applying natural language processing to information retrieval". Journal of the American Society for Information Science and Technology 54(2): 115-123