Exploring text and data mining for recent consumer and sensory issues and their implication in food trends

A Dissertation
submitted in partial fulfilment
of the requirements for the Degree of
Master of Food Innovation

at
Lincoln University
by
Ziyang Chen

Lincoln University
2021

Abstract of a Dissertation submitted in partial fulfilment of the requirements for the Degree of Master of Food Innovation.

Exploring text and data mining for recent consumer and sensory issues and their implication in food trends

by

Ziyang Chen

Meat consumption has caused several problems in terms of overusing freshwater, underground water contamination, land degradation, and animal welfare. To mitigate these problems, replacing animal meat products with alternatives such as plant-, insect-, algae-, yeast-fermented-based protein, and cultured meat is an available strategy. To enhance the commercial success of alternative protein products, understanding the sensory profiles and acceptability from consumers is necessary. In traditional sensory tests, conducting descriptive sensory evaluation is expensive and time-consuming. To overcome these drawbacks, text mining and natural language processing are introduced as a novel approach to obtain sensory attributes and rapidly develop a descriptive lexicon. In this study, the application of text mining and natural language processing in alternative protein profiles was explored by analysing alternative proteins' attributes and descriptive words from n=20 academic papers (that described the recent information of alternative proteins). From 2018 to 2021, plant- and insect-based proteins are the centres of alternative proteins research. Insect-based protein was less popular than plant-based proteins because of food neophobia and psychological barrier. Adults were more likely to accept insect-based protein products. The emotional profile analysis showed that there was no significant association between emotions and protein categories in this study. Our research showed that applying text mining and natural language processing can benefit the descriptive sensory evaluation, which means that it can rapidly obtain and analyse an large amount of data rapidly, thus overcoming traditional lexicon development techniques.

Keywords: alternative proteins, text mining, natural language processing, descriptive test

Acknowledgements

The codes of scraping PDF documents were written by Cristhiam Gurdian. Dr Damir Torrico and Dr Chetan Sharma gave suggestions for writing this article. Furthermore, Dr Damir Torrico helped to justify the whole article as well as Dr Chetan Sharma helped to justify the Materials and Methods, and Results and Discussions.

Table of Contents

ABSTRACT .... **II**

ACKNOWLEDGEMENTS .... **III**

TABLE OF CONTENTS .... **IV**

LIST OF TABLES .... **V**

LIST OF FIGURES .... **VI**

CHAPTER 1 LITERATURE REVIEW **1**

1.1 PLANT-BASED PROTEINS 2
1.2 INSECT-BASED PROTEINS 3
1.3 ALGAE PROTEINS 3
1.4 YEAST FERMENTED PROTEINS 5
1.5 CULTURED MEAT 5
1.6 SENSORY SCIENCE 6
1.7 TEXT MINING AND NATURAL LANGUAGE PROCESSING 8

CHAPTER 2 INTRODUCTION **10**

CHAPTER 3 MATERIALS AND METHODS **12**

3.1 SELECTION OF PAPERS 12
3.2 PROCESSING OF PAPERS OR TEXT 12
3.3 TEXT MINING 12
3.3.1 Web Scraping 12
3.3.2 PDF scraping and text processing 14
3.3.3 TXT scraping and natural language processing 14
3.4 STATISTICAL ANALYSIS 16

CHAPTER 4 RESULTS AND DISCUSSION **17**

4.1 WORD FREQUENCY 17
4.2 RELEVANCE BETWEEN DIFFERENT WORDS 20
4.3 EMOTION ANALYSIS 23
4.4 COMPARISON WITH OTHER TEXT MINING WORKS AND THE LIMITATION OF THIS STUDY 27
4.5 IMPROVEMENT OF THE CODES 28

CHAPTER 5 CONCLUSION **29**

REFERENCES **30**

APPENDIX A PDF DOCUMENT TEXT MINING CODES AND EXPLANATION PRODUCED BY CRISTHIAM GURDIAN **36**

APPENDIX B TXT DOCUMENT TEXT MINING CODES RETRIEVED FROM INTERNET **41**

APPENDIX C THE LIST OF SCIENTIFIC REPORTS THAT ANALYSED BY NATURAL LANGUAGE PROCESSING **45**

APPENDIX D THE FREQUENCY OF WORD IN TEXT MATRIX (TOP 50) **48**

APPENDIX E THE RELEVANCE BETWEEN KEY WORDS AND OTHER WORDS **49**

List of Tables

Table 1. The relevance between key words and other words 22
Table 2. Independence test in correspondence analysis between the protein types and the emotions 24
Table 3. Emotion analysis of each academic paper 24

List of Figures

Figure 1. A part of the text captured from the website by a crawler 14
Figure 2 Basic workflow of NPL 16
Figure 3. Top 50 most frequent words 18
Figure 4. Word cloud of text matrix 18
Figure 5. The proportion of each emotion in the text matrix 26
Figure 6. The counted times of each sentiment in the text matrix 26
Figure 7. The symmetric plot of correspondence analysis 27

Chapter 1

Literature Review

Meat is one of the most important food types all around the world. Over the past two decades, the global consumption of meat has increased by 60% approximately, reaching 380 million tons in 2018 (Department of Agriculture, Water and the Environment of Australia, 2020). Due to the population increase, and meat product requirement trends, meat products' consumption would be double in 2050 (Fiorentini, Kinchla & Nolden, 2020). Although meat production is increasing significantly, it has still not met the growing population's demand to access this product. Usually, developed countries can provide enough meat products and protein resources for their citizens. However, in many developing countries, such as the countries in Africa, Asia or South America, most children do not have sufficient access to this type of protein because the meat is not affordable (Klunklin & Savage, 2018b). Under this situation, if the traditional strategy was followed, there is a continuing pressure to increase the amount of livestock to feed the world.

Producing such an enormous amount of meat and its products has been a heavy burden for the environment, leading to increases in greenhouse gas emissions and nitrates leaching problems. Beef production causes 14.5% of human-related greenhouse gas emissions annually (Fiorentini et al., 2020). In New Zealand, 50% of greenhouse gases are generated by agriculture, half of which comes from the dairy industry (Foote, Joy & Death, 2015). The nitrate concentrations in 30% of groundwater exceeded the standards in agricultural areas (Foote et al., 2015), which would lead to algal blooms (Marsh, 2012). Waste and digesting gas from dairy cattle are the main reasons for causing the above problems (Foote et al., 2015). Moreover, animal husbandry requires a large number of resources from nature, such as freshwater and land. A total of 30% of available freshwater is used to process meat products (including feeding, butchering, packaging) per year (Fiorentini et al., 2020). The area in which cattle live may be compacted, leading to land degradation. As a result, this area may not be suitable to grow crops due to a reduced yield (Nawaz, Bourrie & Trolard, 2013). Furthermore, negative animal welfare has been an emerging problem because the density of livestock is increased. It results in less living space for livestock and also leads to enhance the possibility of microorganism infection. For instance, to protect livestock from microorganisms, antibiotics are needed, resulting in antimicrobial resistance (Giacomelli, Salata, Martini, Montesissa & Piccirillo, 2014). When the antimicrobial resistance is increased, livestock is prone to suffering from diseases. Besides, confined cattle production is not environmentally friendly. Therefore, providing enough meat products or protein resources to meet the entire world's meat consumption demand with limited natural resources has recently been one of the most prominent challenges (Fiorentini et al., 2020).

To fill the gap between production and consumption, using alternative proteins to replace meat products is now considered a potential wholesome solution. It is an available strategy (Fiorentini et al., 2020) because it requires fewer resources from the environment but provides more amount of proteins in volume for humans. For instance, providing the same amount of protein that comes from plants could save 35% - 50% of natural resources in terms of land and freshwater compared to animal protein (Fiorentini et al., 2020). This novel strategy uses plants, insects, microorganisms, or growing animal meat cells in laboratories to produce proteins for replacing animal meat (De Koning et al., 2020; FAIRR, 2019). This alternative can save resources and provide vegetarian options (De Koning et al., 2020) to supply the demand for protein worldwide. There are five main types of alternative protein, and they can be divided into two groups. Group 1 can be gained from the natural environment directly, including plant-based, insect-based, or algae-related alternative proteins. Group 2 is human-made proteins, which means that they can be produced by yeast fermentation, and growing meat from animal muscle cells in a laboratory (FAIRR, 2019).

1.1 Plant-based Proteins

Plant-based alternative proteins have been known and accepted by the public over the past few years. Some meat analogues have been increasingly consumed on the market (De Koning et al., 2020; Fiorentini et al., 2020). Compared with other types of alternative proteins, plant-based proteins are more popular in Western countries, and around the world. These proteins are usually extracted from grains or crops, such as soybean, legumes, and seed (FAIRR, 2019; Fiorentini et al., 2020). For instance, tofu is one of the most famous plant-based protein products made from soy, which has been provided to and accepted by Western culture since the 1960s (Fiorentini et al., 2020). Mung bean and its products also can provide a high amount of protein at low prices (El-Moniem, 1999). The plant-based alternative protein is anchored with 'health', 'environmentally friendly', 'green', and other positive concepts by consumers; thus, it is accepted widely all around the world compared to other alternative proteins. However, some disadvantages are also found when plant-based protein was applied to food products. For example, by using mung bean protein isolate, the modified product would be perceived as bitter and astringent (Ares, Barreiro, Deliza & Gámbaro, 2009), which may be caused by the presence of polyphenolic in mung bean (El-Moniem, 1999). Furthermore, with the addition of plant proteins such as pea, the colour of the meat analogue might be altered (Cosson et al., 2020). Recently, mushrooms were used by some companies (e.g., McDonald) to make beef meat analogues. It is in great demand and worth investing in, and many companies have started to improve their plant-based products' sensory characteristics to taste similar to real animal meat products (Fiorentini et al., 2020).

1.2 Insect-based Proteins

Insects have been introduced as high protein food, and insect-related protein has gained more traction due to its advantages in resources usage which requires less water and soil to raise (De Koning et al., 2020). However, most of the consumers in Western countries still have not accepted entomophagy or consume its products yet (De Koning et al., 2020) due to psychological barriers. Even in East Asia, such as China, consuming insects proactively is a challenging practice in the culture. In south-west China, some ethnic cultures have insect-based cuisine; for example, cockroaches are deep-fried inside a dough that is made of wheat powder (Lv, 2017). furthermore, in some cases, cockroaches are applied in Chinese traditional medicine. Despite this, consuming cockroaches is still quite challenging for the public. Usually, for both Western and Eastern (around the world) groups, they refuse to consume insects because of neophobia and disgusting feelings (De Koning et al., 2020). These negative psychological effects would also result in that consumers prefer plant-based proteins and their products rather than insect-based alternatives (Gómez-Luciano, de Aguiar, Vriesekoop & Urbano, 2019). Interestedly, the acceptance and liking score of insect-based products were higher than plant-based protein when sensory attributes were the only thing to be regarded, which means participants did not know they were consuming insect protein (Schouteten et al., 2016). Although cultural barriers affect the preference of insect-based alternative proteins in the short term, these products would tend to be more acceptable for the public in the long term due to a more frequent exposure over time through advertisement, education, and marketing (De Koning et al., 2020). Thus, insect-based alternative protein is a potential alternative product in the marketplace.

1.3 Algae Proteins

Seaweed and algae are other types of alternative proteins that exist naturally in the environment. Some specific microalgae species can provide a similar protein level as meat and legume (Bleakley & Hayes, 2017). Furthermore, algae and seaweed have a higher yield (9.5 tons per 104 m^2 per year and 5 tons per 104 m^2 per year on average, respectively) compared with traditional plant-based alternative protein sources (1 ton per 104 m^2 per year on average) (Van Krimpen, Bikker, Van der Meer, Van der Peet-Schwering & Vereijken, 2013). They can also provide an advantage in resources' usage. Typically, plant-based alternative proteins and traditional meat proteins require enormous amounts of land and freshwater usage, as mentioned earlier. In contrast, seaweed and algae need neither freshwater nor land to grow and save plenty of resources for humans to use for other activities. The polyphenols and pigments inside seaweed and algae are good for human health, enhancing the nutritional and commercial value of algae-based alternative proteins. At this moment, algae-based proteins are not popular in the marketplace, with almost non-existing commercial products and minimal research. This is the result of several factors' interaction, including the right of harvesting, season, geographic location, commercialisation costs, and the technology of isolating protein from algae (Bleakley & Hayes, 2017).

There is a limited amount of sensory research of algae-based alternative protein. Sălăgean, Pop, Catrinoi & Nagy, (2015) studied the effect of adding brown algae on sausages. In their study, 10% and 15% brown algae added to sausages were made and evaluated. Both 10% and 15% of algae sausages were evaluated as pleasurable and pleasant by participants regarding the overall liking score. The 10% algae sausage had better performance in terms of taste and flavour than the 15% algae sausage. In comparison, the 15% algae sausage performed higher in nutritional value and had a greater influence on the physicochemical characteristics. The researchers concluded that with algae's addition, sausage quality was improved regarding the sensory and nutritional value by adding algae (Sălăgean et al., 2015).

Despite these findings, plant-based alternative proteins are still more popular products than algae in some cases. In Parniakov et al., (2018) study, the effect of different additions, including soy, pea, broad bean, lentils, and two algae species (Spirulina and Chlorella) on chicken Rotti were compared in terms of the sensory characteristics. In their results, Spirulina and Chlorella chicken Rotti had the lowest acceptability scores and sensory preference among the modified chicken Rotti treatments. The reason was the dry texture and particular characteristical tastes that algae provided to the product (Parniakov et al., 2018). Thus, although algae is a potential alternative protein source that can compete with plant-based alternative proteins due to natural resource usage, it still requires further research due to its shortcomings in some product types.

1.4 Yeast Fermented Proteins

Yeast-based protein is not only used in the food industry but also applied in several industries including chemistry, pharmaceutics, cosmeceutical, and detergent (Vieira Gomes, Souza Carmo, Silva Carvalho, Mendonça Bahia & Parachin, 2018). For instance, insulin can be produced by yeast in the pharmaceutical industry (Baeshen et al., 2014). In the food industry, the key point of this technology is to produce an animal protein analogue (for example, whey) by programming yeast through fermentation (FAIRR, 2019), thus, this can produce protein more efficiently rather than raising livestock. Although yeast's protein production has been introduced since the 1980s (Vieira Gomes et al., 2018), it is still considered a novel technology, and there is very limited data about its sensory characteristics. This technique's cost may be the main reason that limits its widespread application in the food industry. Compared to other alternative proteins that were mentioned in the sections above, yeast-produced protein requires more resources (money, time, machine, and professional handlers). Besides, the yield of using yeast to produce protein still needs to be improved, and more research has to be done to understand its properties. Rodríguez-Limas, Tannenbaum & Tyo (2015) introduced a technique to obtaining a higher yield of yeast proteins. Despite these improvements, price is still a significant disadvantage for yeast-produced alternative proteins than its plant-based and insect-based counterparts.

1.5 Cultured Meat

Cultured meat, or clean meat, which is grown in a laboratory from a specific animal muscle cell, is a technique that does not require feeding livestock (FAIRR, 2019; Rolland, Markus & Post, 2020). It can be described as 'cultured, in vitro, synthetic, artificial, and laboratory-grown or factory-grown meat' based on Verbeke, Sans & Van Loo (2015). This idea was first introduced by Post, (2012), and it has been well developed over the past years. In 2015, Verbeke and Post predicted that cultured burger would be marketable in 2020 with a price of 65 US$ per kilogram (Verbeke et al., 2015). The price was still significantly high in 2018, which was up to 800 US$ according to Future Meat Company (González & Koltrowitz, 2019). However, after one year, the price decreased significantly, only costing 100 euros per kilogram, with the projection of achieving 10 euros per kilogram by 2021 (González & Koltrowitz, 2019). Furthermore, the types of meat (not only beef) were also explored, including pork, chicken, and fish (Rolland et al., 2020). The data of sensory evaluation of cultured meat is still very limited. However, recent research showed that psychological factors played an important role in the acceptance and preference of this product. Based on a consumer test in 2015, 43% of participants were willing to try cultured meat; however, 51% of participants are still hesitant (Verbeke et al., 2015). In contrast, 58% of participants were happy to pay more on cultured meat in 2020, as well as more participants preferred to consume (higher acceptance) cultured meat rather than traditional meat. This may be the result of further information that the public received over the last couple of years. Usually, the public would reject novel products or techniques they are unfamiliar with, especially animal products (Rolland et al., 2020). In this case, there were two main concerns from the consumer; the first concern was of the moral order, and the second was health (Verbeke et al., 2015). Some participants felt disgusted when they faced cultured meat caused by cultural connotations or not matching the 'appropriate' food concepts for them (Rolland et al., 2020). Furthermore, research indicated that vegetarians would be more likely to relate this product to unhealthy, unnatural, or other negative concepts (Verbeke et al., 2015). With the development of cultured meat for the past five years, today, more and more consumers are willing to try cultured meat or even pay more for cultured meat instead of traditional meat (Rolland et al., 2020). This product has also an advantage in taste, which was not different from animal meat, such as beef, significantly (Tucker, 2014). This opinion is also supported by Rolland et al. (2020) and Verbeke et al. (2015) research, which indicated that the more participants knew the information of cultured meat, the higher participants would rate its acceptance.

1.6 Sensory Science

Sensory evaluation is not a latterly emerging subject but has been evolving during the past few years. Humans have known to evaluate food since ancient time (Meilgaard, Civille & Carr, 1991). Odd foods were discarded, and delicious foods were maintained in humans' daily lives over the past centuries. However, as more chemical and nutritional research have been conducted on different foods, consumers nowadays are willing to sacrifice the hedonic sensory experiences to gain better nutritional and healthier values (Klunklin & Savage, 2018a; Park, Choi & Kim, 2015). Usually, developing a better nutritional or functional product would lead to poorer sensory experiences because of the usage of novel ingredients (which sometimes have off-tastes and undesirable sensory characteristics) or the reduction of taste-related elements (such as the reduction of sugar and fat).

For instance, to enrich the protein concentration of a biscuit, adding protein from a novel protein source such as mussel can be a viable option to improve its nutrition. Due to the mussel addition, the fortified biscuit's texture would tend to be harder (Klunklin & Savage, 2018a). This might be the result of reducing moisture content due to the increased proportion of protein, which leads to changes in texture (Mancebo, Rodriguez & Gómez 2016).

Furthermore, reducing sugar concentrations for increasing the healthy value in a product may result in lower sensory quality and more increased off-taste. Under this situation, developing a balance point between sensory and nutrition is required for products to achieve commercial success, and consumer sensory evaluation is the key to achieving this goal. Thus, sensory science has always played an important role in product development, increasing the likelihood of success in the commercialised novel products (Verbeke et al., 2015). Consumer sensory evaluation can help product developers to modify their products according to the feedback from consumers. For instance, Tucker (2014) investigated the acceptance and preference of alternative proteins (including insect-based protein and cultured meat) in a group of New Zealander using consumer tests and focus groups in 2014. Few of the participants (most of them were female) carried some psychological barriers against insect alternative protein products, stating that they would never consume insects. In contrast, most of the participants had a positive overall view of the alternative proteins. Although entomophagy can be an intimidating issue for New Zealanders, if positive information is presented (such as "it is good for your health"), or if insects are processed into familiar products (such as hamburgers), some participants might change their opinions about this practice. In the same study, the sensory characteristics of cultured meat were accepted; however, most New Zealanders in the group rejected choosing this product because of the perception that this meat was artificial (Tucker, 2014). These findings were also supported by De Koning et al., (2020), Rolland et al., (2020), and Verbeke et al., (2015). According to sensory and consumer research, introducing entomophagy and cultured meat to the general public is still challenging today. Companies in New Zealand can develop strategies to commercialise their cultured meat or insect-based meat based on this study. They can provide their insect patty to hamburger stores because the participants in the sensory research mentioned it was acceptable when the insect was processed as a part of hamburgers. What is more, they can introduce their cultured meat product as 'healthy' and 'green' as plant-based meat claims through advertising because the main factor which affected consumer acceptance was health concern based on the sensory research. Hence, this is one of the ways that sensory research and evaluation can contribute to product development and commercialisation.

As a key part of sensory science, developing a descriptive lexicon for food products requires an enormous amount of data, which means that the sensory test may need to be repeated several times (Heymann, King & Hopfer, 2014). What is more, to gain high-quality data, participants would need to be trained for several months (Hamilton & Jacob, 2020). Both repeating sensory tests and training pannel consume plenty of time and money (Hamilton & Jacob, 2020). For example, to gain the sensory profile and preference of whisky, a traditional approach is to hold sensory evaluation and focusing group to achieve this goal. It costs time and space for finding participants and holding the evaluation. If the company wants to obtain high-quality data, they have to train the participants, which would cost time and money in large amounts. Otherwise, the untrained participants may not be able to evaluate the sensory profiles of whisky correctly because they may not be familiar with the attributes of whisky. To gain reliable data more efficiently and economically is a goal that many companies aim to achieve. One more challenge in the descriptive sensory evaluation is that different participants may use different words to describe the same or similar attributes, characteristics, and emotions. It would cause problems when the researcher analyses the data, spending more time to group the descriptive words. Thus, exploring an alternative method to improve the descriptive lexicon's developing process for a product is required (Hamilton & Jacob, 2020).

1.7 Text Mining and Natural Language Processing

Under this situation, text mining and natural language processing were introduced (Bakhtin, Khabirova, Kuzminov & Thurner, 2020; Bécue-Bertaut, Mónica, Álvarez-Esteban & Pagès, 2008; Zong, Hamilton & Jacob, 2020; Wang & Du, 2010) as options to obtain quick and reliable information. Text mining is an artificial intelligence technology that can read enormous text and reform that information into a structured and justified form, which is suitable for analysing key trends using machine learning algorithms (Linguamatic, 2020). Text mining relies on natural language processing to achieve this goal. Helping researchers to collect information for answering a specific question, text mining is widely used in knowledge-driven organisations, biomedical science, biological science, autoimmune disease research, toxicogenomic and protein docking (Badal, Kundrotas & Vakser, 2015; Cohen & Hunter, 2008; Gorr, Wennblom, Horvath, Wong & Michie, 2012; Leaman, Wei, Allot & Lu, 2020; Lee, Liu, Kelly & Tong, 2014; Linguamatic, 2020; Rebholz-Schuhmann, Kirsch & Couto, 2005). In food science, especially sensory evaluation, the number of research applying text mining is still very limited. It might be the result that there are several challenges regarding using text mining in sensory analysis. A limited number of database and scattered data in terms of food sensory would be one of the first challenges in this field.

Hamilton & Jacob (2020) developed a descriptive lexicon for whiskies with two prominent websites: WhiskyCast and Whiskyadvocate by applying text mining and natural language processing. They ran codes to do the natural language processing (including created word cloud and other visual analysed data). They achieved their goal after gaining a large data set (2309 reviews and 4298 reviews, respectively) regarding the sensory attributes and prices from these two websites. However, their findings were limited to these two databases (or websites). Information can be collected easily using several codes because all this information was contained in these sources. If the data were scattered, the collection of this information would need to reach hundreds of different websites, which could be a problem because the codes used to grab texts from the website would need to be modified several times to fit the different website structures (Hamilton & Jacob, 2020). Furthermore, hundreds of URL (Uniform Resource Locator) and CSS (Cascading Style Sheets) selectors are required to be provided to the controller each time the web structure changes. In this case, a website providing abundant alternative protein information was not found. Thus, this was one factor that limited the amount of analysed data.

When the specific database is not found, social media data would be a choice. Some consumers would post their sensory experience of products on social media, which is valuable free information for sensory research. However, this type of data might be subjected to restrictions (such as being not open accessible). Taking Facebook as an example, most of the data on Facebook is not open access, which means that web crawlers could not grab them (Tao, Yang & Feng, 2020). What is more, in some regions or countries, using a web crawler to collect information on the internet can be forbidden (Hamilton & Jacob, 2020). Therefore, applying text mining and natural language processing in sensory science is limited by these factors, making it just available in some specific cases. For its broader application in the sensory field, developing more databases or exploring other ways to improve this technique is required.

Chapter 2

Introduction

Several environmental problems were caused by the increased meat consumption and related industry, including increasing greenhouse gas emissions, nitrates leaching, land compaction, over-consumption of water, and antimicrobial resistance (Fiorentini et al., 2020; Foote et al., 2015; Giacomelli et al., 2014; Nawaz et al., 2013). Thus, to meet the increasing requirement of meat consumption in a more environmentally friendly manner, replacing traditional meat with alternative protein is a potential solution. There are five main kinds of alternative proteins, including plant-based, insect-based, algae-related, fermented by yeast, and culture meat (or in-vitro meat) (FAIRR, 2019). Many companies have started to explore the possibility of replacing animal meat products with these five types of alternative proteins (Fiorentini et al., 2020). To increase the likelihood of successfully commercialising novel products like these, therefore, sensory evaluation plays an important role in product development to justify the product according to feedback from the consumer (Verbeke et al., 2015).

As a key part of sensory science, developing descriptive lexica for food products requires a large data set, which means that the sensory tests may need to be repeated over time (Heymann, King & Hopfer, 2014). What is more, for the evaluated product's specific sensory experiences, open-ended questions are used commonly instead of closed-ended questions. Open-ended questions were used to investigate the factor that drives consumers' liking score on products regarding the sensory attributes (Lawless & Heymann, 2010; Spinelli et al., 2017). It can be an alternative or complementing approach to map consumers' preferences according to their descriptive comments based on texts (Spinelli et al., 2017). Widely applied in food industries, such as wine and meat, it can indicate the descriptive words which are related to the products from the feedback of participants (Spinelli et al., 2017). Consumers can be freer and provide all the product's opinions by using their language rather than focusing on the researcher's aspects compared to determinate questions (Spinelli et al., 2017). However, due to its characteristics and high freedom of word choice, the rough text would tend to be harder for the analysis and time consuming (Delarue & Lawlor, 2014; Spinelli et al., 2017). The analysed text matrix may lead up to thousands of words, and one word may have different meanings in different sentences. This kind of text's basic workflow is followed by text segmentation, sentence tokenisation, lemmatisation, and stemming to group similar words together (Bakhtin et al., 2020; Spinelli et al., 2017).

All these works are barely possible to be finished by using manual operation; thus, under this situation, an automatic approach (algorithms) shows its significant advantage in time-saving. Recently, text mining and natural language processing were introduced to help researchers obtaining sensory data easier and faster from the internet instead of using repeating sensory tests (Bakhtin, Khabirova, Kuzminov & Thurner, 2020; Bécue-Bertaut, Mónica, Álvarez-Esteban & Pagès, 2008; Zong, Hamilton & Jacob, 2020; Wang & Du, 2010). This automation can decrease the time and money spent on research. What is more, it can read an enormous amount of sensory data and reform that information into a structured and justified form that is suitable for further analyses (Linguamatic, 2020). With this technique, sensory research can be held more quickly relatively than the traditional sensory tests. For the past decades, to save time and money in descriptive analysis, researchers have developed several types of rapid descriptive analysis. Simultaneously, all of these methods diminished some of the outstanding properties of descriptive analysis on varied levels. With the combination of algorithms and descriptive analysis, the limitation of human processing data has been broken. This research aimed to use text mining and natural language processing to explore alternative protein's sensory attributes based on the data collected from the internet and scientific reports. Although this has a limitation in terms of the number of scientific papers (n=20 articles), it still provides a prototype of applying text mining on future consumer tests.

Chapter 3

Materials and Methods

3.1 Selection of papers

To obtain the data, one of the most important things is that it is accessible. All of these 20 papers are accessible for hypertext-markup-language (HTML) and portable-document-format (PDF) form, which means that they can be scraped by web crawler as well as PDF text mining command in R (Version 1.3.1093) (R Core Team, 2019) after downloading. Thus, an alternative approach can be developed if the first scarping method does not work. Furthermore, all these 20 articles are in the period from 2018 to 2021. They are recent studies, hence can provide the latest information and trend of the alternative proteins.

3.2 Processing of papers or text

All the work was done in a statistical computing language called R (Version 1.3.1093) (R Core Team, 2019). The packages applied in R were *rvest* and *xml2* (for web scraping), *pdftool* (for PDF document scraping), *tm* (for text mining), *SnowballC* (for text stemming), *RColorBrewer* (for colouring bar chat and word cloud), *syuzhet* (for emotion analysis and classification), *ggplot2* (for plotting charts) and *wordcloud* (for developing word-cloud). The pictures of some results were cut by screenshot in portable-graphics-format (PNG) document type in order to improve the pixel of the image.

3.3 Text Mining

3.3.1 Web Scraping

Although grabbing information from a website manually is available in some cases (Hamilton & Jacob, 2020; Ickes, Lee & Cadwallader, 2017), applying a web crawler would be more advantageous because it saves time. In this case, almost all the data was collected from scientific reports which were in .pdf document type. This simple web crawler just scraping a single page to show the basic guideline for web scraping.

Codes:

```
‘

library(xml2)

library(rvest)

read_html("https://www.dataquest.io/blog/web-scraping-in-r-rvest/")

web=read_html("https://www.dataquest.io/blog/web-scraping-in-r-rvest/")

web %>% html_nodes("body span, p, ul, li") %>% html_text

’
```

The first step was that loaded the packages which supported web scarping. In this case, *xml2* (R code) and *rvest* were loaded by the first and second lines. Applying ‘*read_html()*’ command and typing the URL into the brackets to capture this page’s source file in the third line. After this, CSS (Cascading Style Sheets) information (in the .html document) was used to locate the text which was needed to be scraped on the page. Normally, these elements of the website could be reached by opening the developing tool in the browser. Typed the CSS information into the brackets in the ‘*html_nodes()*’ command, and all of the text on this webpage was scraped and illustrated in the R console. A part of it was indicated in figure 1.

```
 [62] "Before we can start learning how to scrape a web page, we need to understand how a web page itself is structure
d."


 [63] "From a user perspective, a web page has text, images and links all organized in a way that is aesthetically pleas
ing and easy to read. But the web page itself is written in specific coding languages that are then interpreted by our w
eb browsers. When we're web scraping, we'll need to deal with the actual contents of the web page itself: the code befor
e it's interpreted by the browser."
 [64] "The main languages used to build web pages are called Hypertext Markup Language (HTML), Cascasing Style Sheets (C
SS) and Javascript. HTML gives a web page its actual structure and content. CSS gives a web page its style and look, inc
luding details like fonts and colors. Javascript gives a webpage functionality."


 [65] "In this tutorial, we'll focus mostly on how to use R web scraping to read the HTML and CSS that make up a web pag
e."


 [66] "Unlike R, HTML is not a programming language. Instead, it's called a markup language – it describes the content a
nd structure of a web page. HTML is organized using tags, which are surrounded by <> symbols. Different tags perform dif
ferent functions. Together, many tags will form and contain the content of a web page."


 [67] "The simplest HTML document looks like this:"
```

Figure 1. A part of the text captured from the website by a crawler

3.3.2 PDF scraping and text processing

Because of the multiple problems mentioned above, .pdf documents were used to scrape in this study. To scrap the .pdf document had a similar workflow as web scraping. The codes applied in this study were indicated in Appendix A, and it was written by Cristhiam Gurdian from Louisiana State University, the United States of America. The first step is to download academic articles that are suitable for the research topic. As the description in Appendix A, the codes would only work if the working directory was set to the folder to which the PDF files were downloaded. After the directory was set successfully, the codes can be run to do Natural Language Processing (text segmentation, sentence tokenisation, lemmatisation, and stemming). When this step was done, it meant that this text matrix was ready to be analysed. Straight afterwards, word count and other visual data can be produced by applying packages in the R program such as, *syuzhet*, *ggplot2*, and word cloud which were used in this study. These codes achieved word counts of keywords in the text. To go further and be more specific, works were done by the codes illustrated in Appendix B.

3.3.3 TXT scraping and natural language processing

In order to gain more specific data regarding sensory attributes of alternative protein, text from the academic paper was selected. The introduction, materials and methods, conclusion, and references sections were eliminated, and only the results and discussions part were generated. The text from academic papers was copied and pasted into the TXT document. There were 20 analysed papers, and each result and discussion section of the papers would be pasted in a new TXT document alone. After that, all the text in those TXT documents would be collected and created a new TXT document as the main analysed text matrix in this study. Thus, 20 TXT documents that contain the text from 20 academic papers and one TXT document named 'Main Text Matrix' that contained all of the text which was inside of 20 TXT documents were generated. In total, 21 TXT documents needed to be analysed. The Main Text Matrix was to investigate the whole picture of these twenty academic papers in terms of the sensory attributes of alternative proteins. All of these documents would be captured and processed to Natural Language Processing text segmentation, sentence tokenisation, lemmatisation, and stemming) by the codes (shown in Appendix B) before producing any visual data.

The frequency of each word occurring in the Main Text Matrix would be counted, indicated as a table and bar chart. By doing this, the relationship between words and alternative proteins could be developed preliminarily. Sentiment analysis and emotion classification would be done by the package called *syuzhet* (R code). The occurring times of sentiment would be counted and indicated in a bar chart, while the proportion of each emotion in the matrix would be calculated as a percentage and illustrated in the bar chart. The emotion classification of 20 TXT documents was run individually in order to gain the proportion of emotions data in each paper. The types of alternative proteins mentioned in each article were also be indicated, thus, the emotion of each type of alternative protein can be explored. Word cloud would also be produced during the analysis. It is an intuitive image showing the frequency of words in the matrix. Based on the result of the word frequency, which was analysed above, the association of these words was investigated. This process can show the vocabularies around the terms which were aimed at, as well as how strongly they were related. More specific and reliable details of alternative protein can be collected by following the word association data.

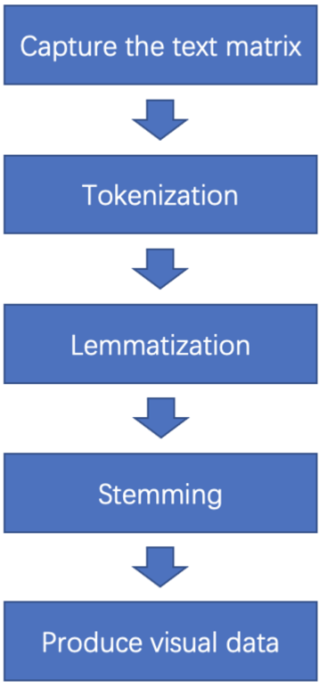

Figure 2 Basic workflow of NPL

3.4 Statistical Analysis

To obtain the visual relationship between emotions and the types of alternative proteins, the statistical analyses were correspondence analysis and k-proportion test processed by software XLSTAT (Version 2018.1.1.62926) in Excel while $p<0.05$ for significant analysis.

Chapter 4

Results and Discussion

Overall, the frequency of words showed in the Main Text Matrix was analysed and indicated in Figure 3 as a bar chat. The specific data of word count has been attached in Appendix D. Word cloud was generated to show the word frequency more intuitively and illustrated in Figure 4. The most frequent word would be placed in the centre of the word cloud as well as the words with higher frequency would tend to be bigger while the words with lower frequency would be smaller. The proportion of each emotion in the text matrix and the times of sentiments showed in the text matrix were indicated in Figure 5 and Figure 6 as bar charts, respectively. The proportion of emotions in each paper (20 articles in total) were generated and showed in table 3. Last but not least, the relevance between keywords and other words was analysed and attached in Appendix E. All the words showed in the tables, figures and appendixes were their root form. For instance, '*consum*' would represent 'consumer', 'consume', 'consumes', 'consuming', 'consumed', and 'consumption'. Thus, when the frequency of consum was 264 times, it means all the words which were related to this root (in this case, they were consumer, consume, consumes, consuming, consumed, and consumption) appeared 264 times in total.

4.1 Word frequency

The roots of the word including *ffs*, *meat*, *protei*n, *product*, *food*, and *consum* were at a very high level of frequency with 697, 531, 432, 404, 356, and 264 times, respectively. Indeed, it is normal to read these words (*meat*, *protein*, *product*, *food*, and *consum*) in the academic papers which were studying alternative proteins. *Ffs* was the highest word root showed in the text matrix, while its meaning was not clear. When the twenty academic papers were analysed individually, *ffs* was also indicated in their word cloud. Hence, the assumption was that *ffs* represented a by-product or a type of compounds of the alternative proteins, which would occur in every alternative proteins' product.

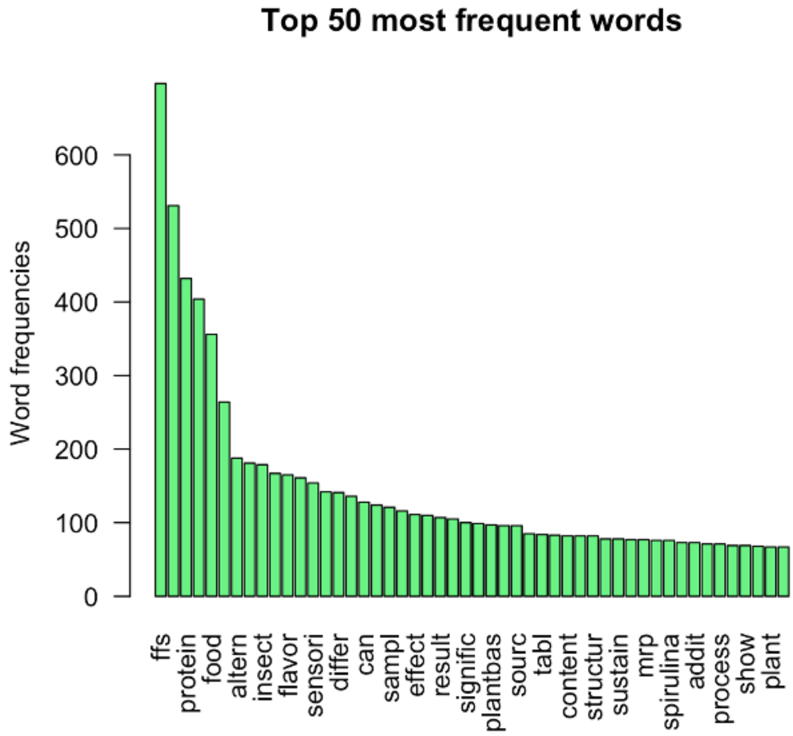

Figure 3. Top 50 most frequent words

Figure 4. Word cloud of text matrix

The other words that were related to the type of alternative protein were *insect* (179 times), *plantbas* (97 times), *pea* (82 times), *spirulina* (76 times), and *plant* (67 times) were also indicated as top 50 frequent words in the matrix. Due to the word root *plantbas* and *plant* had a similar meaning, they could be summarised together and represented the plant types of alternative protein. Thus, the claim could be made that among five types of research focused on insect-based protein the most (179 times) and followed by the plant-based alternative protein (164 times). This result was conflicting with Appendix D. Appendix D indicated that plant-based protein was the topic of 15 academic papers (Agbemafle, Hadzi, Amagloh, Zotor & Reddy, 2020; Chiang, Hardacre & Parker, 2020; Cosson et al., 2020; De Koning et al., 2020; Fiorentini et al., 2020; García-Segovia, Igual & Martínez-Monzó, 2020; Grahl et al., 2018; Grasso, Hung, Olthof, Verbeke & Brouwer, 2019; Kaleda et al., 2020; Kamani, Meera, Bhaskar & Modi, 2019; Martin, Lange & Marette, 2021; Possidónio, Prada, Graça & Piazza, 2021; Sha & Xiong, 2020; Stephan, Ahlborn, Zajul & Zorn, 2018; Yuliarti, Kovis & Yi, 2021) while insect-based protein was only mentioned in 9 academic papers (Agbemafle, Hadzi, Amagloh, Zotor & Reddy, 2020; Altmann, Neumann, Velten, Liebert & Mörlein, 2018; Ardoin, & Prinyawiwatkul, 2020; Chow, Riantiningtyas, Sørensen & Frøst, 2021; De Koning et al., 2020; García-Segovia, Igual & Martínez-Monzó, 2020; Grasso, Hung, Olthof, Verbeke & Brouwer, 2019; Mishyna, Chen & Benjamin, 2020; Possidónio, Prada, Graça & Piazza, 2021). Thus, plant-based protein was the hottest topic in this study. This limitation might be the result of the number of academic papers because the result of text mining would tend to be more correct when the text matrix is bigger (Hamilton & Jacob, 2020). In the small text matrix, for instance, in this study, a slight difference might not be able to be investigated, while a significant difference could still be found. The difference in frequency between plant-based protein and insect-based protein was only 15 times, which was a very small difference. In contrast, there was no word root which was associated with cultured meat and yeast fermented protein that is shown in the top 50 frequency word list, which means that they were not important among these 20 articles. Hence, based on the frequency of the word, the claim can be made that plant-based protein (164 times) and insect-based protein (179 times) were the hottest topics in this study, while algae-related (76 times) protein was with lower focusing and the cultured meat and yeast fermented protein were the least. Instead of soy, pea was the only plant word indicated in the top 50 frequency words. Based on this result, the assumption can be made that the attention of researchers has been shifted to pea in terms of producing plant-based protein in recent studies (from 2018 to 2021, which was the published year range of twenty papers; Cosson et al., 2020; García-Segovia et al., 2020; Kaleda et al., 2020; Martin et al., 2021; Sha & Xiong, 2020; Stephan et al., 2018; Yuliarti et al., 2021). In their research, pea was the investigated plant while two papers made the comparison with soy, and in Cosson et al., (2020) study, they mentioned that pea protein has become more commonly applied in the food product as a plant-based alternative protein. The same approach could also be applied to the word root spirulina. When the application of algae alternative protein was explored, spirulina was the most comment algae, which was used in products based on this study in recent years.

Thus, based on the frequency of the word in the text matrix, key points of the matrix can be investigated. For instance, in this study, plant-based and insect-based proteins were found out that they were in the centre of the researcher's attention, while the cultured meat and yeast fermented protein were studied less compared to plant and insect protein recently. Pea has become the most comment plant regarding plant-based protein application research, while spirulina was the most popular algae in algae alternative protein research.

The word roots which may indicate the attributes of alternative proteins such as *differ*, *accept*, *increas* (the root of increase), *like*, and *posit* (the root of positive), were also illustrated in Appendix D with 141 times, 124 times, 116 times, 110 times and 73 times respectively. To analyse this type of word, the assumption could not be made easily only based on the word roots' frequency because word roots would be counted whether it was positive or negative in the article. Taking the word root '*differ*' as an example, in the text, either time of significant different or no significant difference would be counted as word root '*differ*'. The proportion of significant difference and no significant difference in the text matrix was unknown. Thus, the assumption could not be made that there was a difference between traditional meat product and alternative protein product due to the word root differ occurred high frequently as simple as above. The same rule was also suitable for the word root *accept*, *increas*, and *like* because they may represent not acceptable, not increased, and not like. Although the antonym of these words could be written as unacceptable, decrease and dislike, like participants using their own descriptive word, it is still possible that different authors have their own writing style. Hence, analysing the relevance between keywords and other words can be used to support data analysis for word frequency and improve the reliability of the assumption made based on the text mining data.

4.2 Relevance between different words

Word frequency would indicate a whole picture of the text matrix, while neither positive nor negative statements were still unclear. In table 1, a part of the relevance between keywords and other words were shown. For the full result of the relevance, Appendix E can be referred to. It is easy to notice that the word ‘insect’ had a high association with words including ‘willing’, ‘neophobia’, ‘cockroach’, ‘disgust’, ‘novel’, and ‘bit’. The sensory profile and the acceptance of insect alternative protein were illustrated in these words. ‘Willing’ and ‘neophobia’ had a similar coefficient which means that they may be associated with each other. The claim can be assumed that insect neophobia would affect the willingness of trying insect alternative protein. Several articles supported this finding. In De Koning et al., (2020) research, they found that food neophobia would affect the willingness to consume insect protein and would impact plant-based protein. It would cause a negative influence on accepting the terms of entomophagy and sensory appeal (Ardoin & Prinyawiwatkul, 2020; Chow et al., 2021; Grasso et al., 2019). The word ‘cockroach’, ‘disgust’, and ‘novel’ also showed a high and similar relation between them and ‘insect’. It can be concluded that ‘cockroach’ was a hot topic regarding insect alternative protein because it was mentioned in Chow et al., (2021) and García-Segovia et al., (2020) studies. ‘Disgust’ and ‘novel’ were the significant descriptive word for the insect alternative protein based on this result. Indeed, entomophagy was novel in Western cultures and disgusting was a common emotion appeared in participants while consuming insect was introduced (Ardoin & Prinyawiwatkul, 2020; De Koning, et al., 2020). Furthermore, insect bread was evaluated as disgust by participants (García-Segovia et al., 2020), and the observation of the disgust emotion contributed to the rejection of entomophagy more greatly than food neophobia (Chow et al., 2021).

As mentioned above, to make an assumption that was only based on word frequency of the words such as ‘accept’, ‘like’ and ‘expect’ were not critical and reliable. All of these three words were surrounded by negative words in high association level: *Dont* (0.45) was related to *accept*; *Negat* (0.29) which was the word root of negative was related to *like*; *Disappoint* (0.35) and *reject* (0.29) were associated with *expect*. Thus, alternative proteins had still not been accepted/liked/expected at some level (the negative words were not 100% related to the keywords) that could be made as an assumption. To figure out which type of alternative proteins impacting acceptance negatively, all data need to be considered. Firstly, plant-based alternative protein, insect-based alternative protein, and algae protein were the text matrix’s main objectives based on the result of word frequency. According to the relevance of insect with other words, it could be assumed that insect protein would cause a negative effect on acceptance. Furthermore, in the relevance analysis of word like, *spirulinarel* (the word root of spirulina, an alga) was highly related to *like*. Last but not least, there was no negative word shown in the relevance of *plantbas*. Hence, it could be assumed that insect-based protein was the part that has not been accepted by consumers among alternative proteins while the plant-based and algae protein was better accepted. The acceptance of cultured meat and yeast fermented protein was unknown because the information of them was limited. Indeed, plant-based alternative products were acceptable for participants normally while most of the negative comments about the acceptance of alternative proteins were associated with insect-based protein and cultured meat. According to the sensory evaluation result from n=71 participants, plant-based (soy) meat analogues were as acceptable as beef samples regarding visual appearance (Fiorentini et al., 2020; Gómez, Ibañez & Beriain, 2019). The sausage made from wheat and soy isolate was not significantly different from the traditional sausage in terms of texture (Fiorentini et al., 2020; Kamani, Meera, Bhaskar & Modi, 2019). Other research also showed that there was no significant difference between the plant-based (soy) meat patty and all-beef patty regarding overall liking score (Fiorentini et al., 2020; Wong, Corradini, Autio & Kinchla, 2019). Furthermore, De Koning et al., (2020) and Gómez-Luciano et al., (2019) claimed that consumers prefer to adopt plant-based alternative protein rather than insect-based protein based on his research. The same result was also indicated in other papers that were used in this text mining analysis. Usually, consumers refuse to consume insects because of neophobia and disgusting feelings (De Koning et al., 2020). Although sometimes participants might accept the insect protein product after educating or convincing, the first impression of consuming insect was disgusting and unadoptable the most of times (Ardoin & Prinyawiwatkul, 2020; Chow et al., 2021; De Koning, et al., 2020; García-Segovia et al., 2020; Tucker, 2014).

In table 1, a part of the descriptive lexicon was developed regarding plant-based and pea alternative protein. The word *plantbas* was related to health, *insectbas*, and Asia. The consumer would tend to agree what consume plant protein was healthier than meat as well as it has been proved scientifically (Cosson et al., 2020; De Koning et al., 2020; Fiorentini et al., 2020; García-Segovia et al., 2020; Martin et al., 2021). Many plant-based protein products such as tofu, were first introduced in Asia (De Koning et al., 2020; Fiorentini et al., 2020). This might be able to explain the high relevance between plant-based protein and Asia. According to the table, plant-based protein and insect-based protein were compared in high frequency. Hence, it was not surprising to find that there was high relevance between plant-based and insect-based proteins. Due to the plant-based protein was significantly high in the word frequency, and pea was the only type of plants that were shown in the top 50 words frequency, pea was also analysed as a keyword. As the result shown in table 1, pea was in high relevance with mushroom (0.51) and lupin (0.54) because there were two articles in the text matrix that compared pea protein to mushroom and lupin (Cosson et al., 2020; Stephan et al., 2018). Because of this, it was difficult to judge whether the following descriptive words were related to pea, mushroom, or lupin. Indeed, in the article, pea was described as green, beany, fresh, and grassy attributes while lupin was evaluated as beany/green, mushroom/earthy, nutty, and other attributes

(Cosson et al., 2020). In other words, the attributes in terms of dusty and earthy were not related to pea protein (Cosson et al., 2020).

Table 1. The relevance between keywords and other words

1. Insect (%)				
Willing (0.38)	Neophobia (0.37)	Cockroach (0.34)	Disgust (0.33)	Novel (0.33)
2. Accept (%)				
Dont (0.45)	Adult (0.31)	Barrier (0.29)	Elder (0.29)	
3. Like (%)				
Tomato (0.41)	Spirulinarel (0.30)	Negat (0.29)		
4. Plantbas (%)				
Health (0.26)	Insectsbas (0.26)	Asia (0.25)		
5. Expect (%)				
Disappoint (0.35)	Novel (0.30)	Reject (0.29)		
6. Pea (%)				
Lupin (0.54)	Mushroomi (0.51)	Dusti (0.49)	Green (0.45)	Nut (0.40)
Earthi (0.38)	Bitter (0.34)			

Based on the relevance analysis, it could be also found that the acceptance of alternative protein was also related to age. Adult (0.31) and elder (0.29) were more likely to accept alternative proteins (Grasso et al., 2019). This may result in food neophobia mitigated while the age increases (Ardoin & Prinyawiwatkul, 2020; Chow et al., 2021; De Koning et al., 2020; Grasso et al., 2019). All of these findings were contributed by word frequency and relevance analysis as a part of text mining, and it was necessary to consider both while assumptions were made.

4.3 Emotion analysis

The emotions analysis for the whole text matrix were illustrated in Figure 5 and 6, while the analysis of the emotions of each paper was shown in table 3 and the result of correspondence analysis was indicated as a symmetric plot in Figure 7. Overall, there no significant relevance between protein types and emotions in alpha=0.05 level because Chi-square observed value was lower than Chi-square critical value in correspondence analysis according to table 2. Hence, the null hypothesis (The rows and the columns of the table are independent) was accepted.

Table 2. Independence test in correspondence analysis between the protein types and the emotions

Chi-square (Observed Value)	50.734
Chi-square (Critical Value)	66.339
DF	49
p-value	0.405
alpha	0.05

Table 3. Emotion analysis of each academic paper

NO.	Alternative protein types	Proportions of each emotion in the text (%)							
		Trust	Joy	Anticipation	Sadness	Fear	Disgust	Anger	Surprise
1	Plant	30	18	14	8	10	7	7	6
2	Plant, insect	34	17	16	9	7.5	7.5	9	0
3	Plant, insect	23	14	17	15	11	6	7	7
4	Plant	22	13.5	17	15	11	6.5	7.5	7.5
5	Plant, insect	40	21	13	6.5	5.5	4	5	5
6	Insect, algae	27.5	17.5	19	8	5	11	6	6
7	Plant, insect, cultured meat	32	20	14	11	10	4	4	5
8	Insect	23	16	14	11	11	9.5	11	4.5
9	Plant	35	17	29	4	4	3.5	0	7.5
10	Plant	28	14	16	16.5	13	1	4	7.5
11	Plant	19	14.5	13.5	12.5	13.5	8.5	7.5	11
12	Plant	34	12	26	14	4	4	6	0
13	Plant, algae	41	16	8.5	10.5	7	5	4	8
14	Plant	26	14	13	12	5.5	12.5	10	7

15	Algae	32	23	14	10	5	8	5	3
16	Plant, insect, algae, cultured meat	36	20	15	10	7	4	6	2
17	Plant	30	15	15.5	11.5	10	5	6.5	6.5
18	Insect	31	16.5	20	9.5	2.5	6	5	9.5
19	Plant	32	27.5	11.5	6	11.5	0	0	11.5
20	Insect	24	16	11	9	14.5	14	8	3.5

'Trust' dominated the main proportion of text matrix emotions (table included 'trust', 'joy', 'anticipation', 'sadness', 'dear', 'disgust', 'anger', and 'surprise') making up to 34%. However, because of the tone of the academic papers, 'trust' was not positive neither negative in the test. 'Joy' had the-second-high proportion in the text matrix, which might indicate that researchers were optimistic for the future of the alternative proteins. Consumers had well accepted plant-based protein due to its health worth and relatively pleasant sensory attributes (Cosson et al., 2020; De Koning et al., 2020; Fiorentini et al., 2020; García-Segovia et al., 2020; Martin et al., 2021). For the insect-based protein, although cultural barriers affect the preference of insect-based alternative proteins in the short term, these products would tend to be more acceptable for the public in the long term due to a more frequent exposure over time through advertisement, education, and marketing (De Koning et al., 2020). Algae-based protein, cultured meat, and yeast-fermented protein were an advantage in their resource usages such as requiring less soil and freshwater (Rodríguez-Limas et al., 2015; Van Krimpen et al., 2013; Verbeke et al., 2015). The negative emotions, including 'sadness', 'fear', and 'disgust', represented 10%, 8.5%, and 7%, respectively. These emotions in the text matrix might be affected by the considering of alternative protein drawbacks. For example, the text about food neophobia would lead to fear and the descriptive word disgusting would cause disgust in the text matrix emotion analysis.

Figure 7 indicated the relationship between emotions and the categories of alternative proteins. In the symmetric plot, insect and insect, algae were separated from other categories as well as associated with disgust and anger. Besides this, there was no noticeable relationship shown. This result was the same as the independence test.

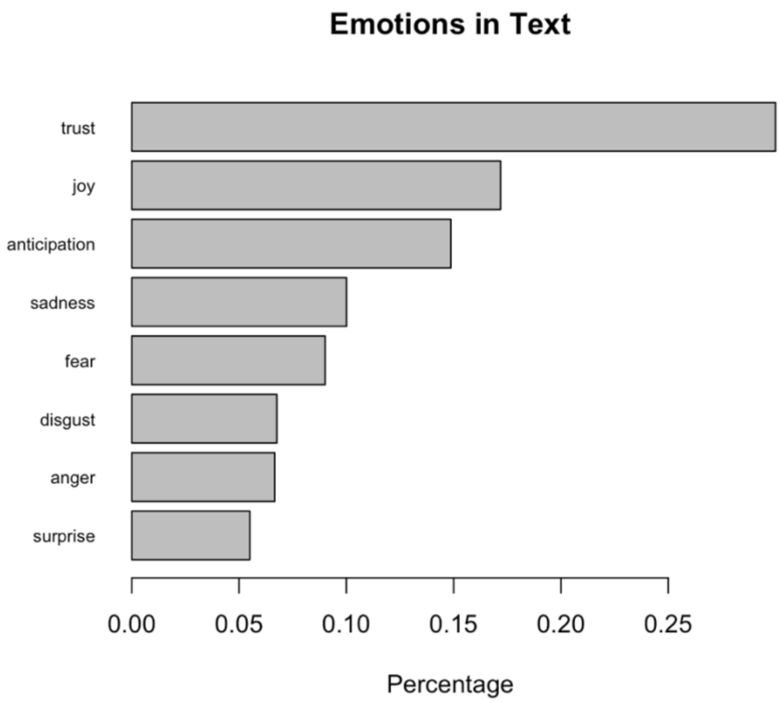

Figure 5. The proportion of each emotion in the text matrix

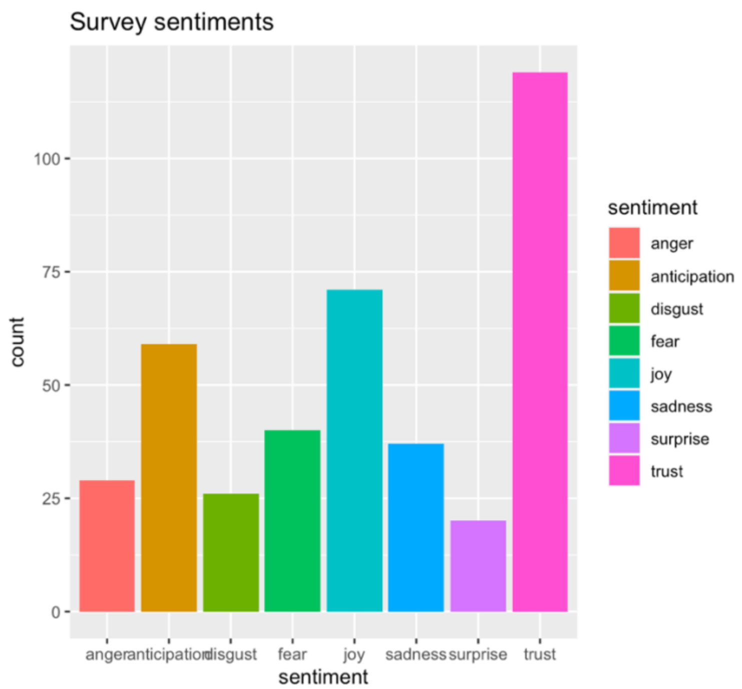

Figure 6. The counted times of each sentiment in the text matrix

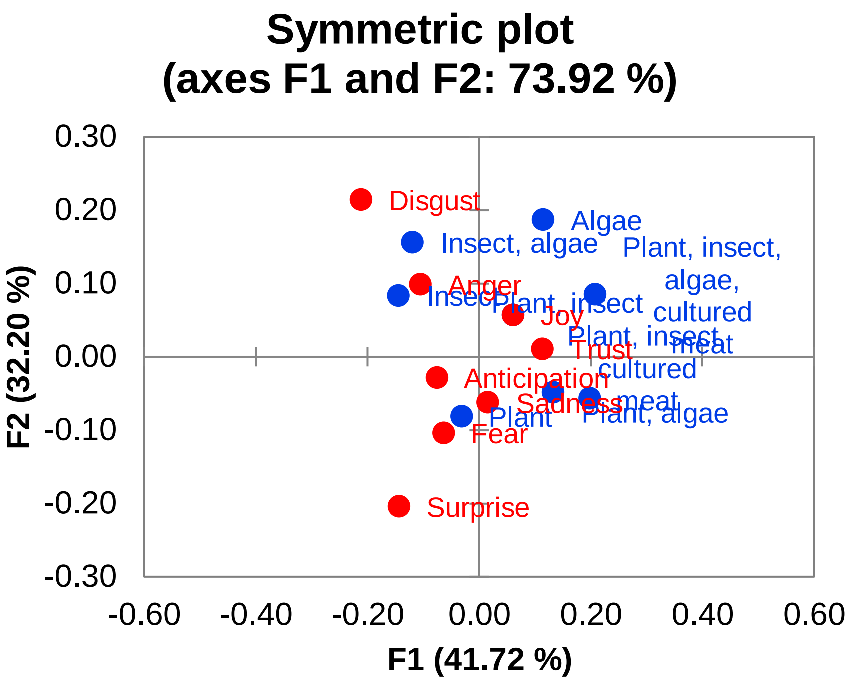

Figure 7. The symmetric plot of correspondence analysis

4.4 Comparison with other text mining works and the limitation of this study

In Bakhtin et al., (2020) research, they analysed over 30 million documents to try to figure out the core research topics and trends in agriculture and food production. The data was collected from several databases, media, websites, and organisations. Based on their research, using fertilisers and chemical agents in farming were the major issues which were studied in food security. Embryo DNA, gene editing, and CRISPR/Cas9 were becoming the centre of genetic research instead of gene modification which had been popular for years. In the future, edible insect, industrial meat production, and industrial food systems would be the focus of extensive research. Furthermore, there is a higher relevance between food security and biological hazards, fungicides, and pesticides (Bakhtin et al., 2020). The approaches applied in their research were only text clustering and word frequency. Indeed, with the enormous amount of text, it was unnecessary to apply word association to justify the result. The error would be mitigated or even almost eliminated with that huge amount of text in text mining. Compared to our study, it was a significant advantage in text scope, while the website providing great alternative protein information was not found in our study leading to a limited context. Their research could be claimed as big data analysis, which was a robust analysis approach for finding a relationship between terms through an enormous amount of text without knowing the reason.

In another study, social media, website, and databased papers were collected to study which was in terms of food safety, dietary pattern characterisation, consumer opinion, product development, food knowledge discovery, and food supply chain management by text mining (Tao et al., 2020). There were 57 papers used in total, which was similar to our study. The approaches which were similar to our study were used in their study including word frequency, word association analysis, and sentiment analysis. Furthermore, they also explored the application of other novel text mining analysing approaches such as text classification, text clustering, and topic modelling (Tao et al., 2020). Compared to our research, on the one hand, although more papers were investigated in their study, the number of scientific reports of our study was higher than in their study, which means that we have qualifier data resource (Tao et al., 2020). The result would tend to be more scientific and the words had been justified in advance in the original paper, leading to a more specific description. On the other hand, a higher proportion of social media and internet data is beneficial for the development of the lexicon. The descriptive lexicon developed in our study was very limited, which could not show the whole sensory picture of alternative proteins. One of the reasons was that we did not have a description of alternative proteins from consumers. In the scientific reports, more words would be used to discuss the mechanism rather than attributes. What is more, with the text from consumers, emotion analysis can be applied to analyse consumer. It could show the attitude of consumers on alternative proteins more directly. There was another drawback of their research that the visual plot was not provided in their article. With the visual plot, the result of their research would be indicated more intuitively rather than text.

4.5 Improvement of the codes

Both the scraping PDF and TXT codes had a drawback: they could not locate specific text sections. Specifically, the codes only could scrape the whole document instead of sections in the document. It caused a problem that to make the aimed section be scraped individually, it could only be achieved by copying the sections and pasting them as a new document. Furthermore, for the TXT document's codes of scraping, the aimed document needed to be selected manually every time. Thus, the locating command and looping command could be further explored if they were available.

Chapter 5

Conclusion

In conclusion, this study analysed n=20 scientific reports to explore the application of text mining in sensory research. According to the result of word frequency, the plant-based and insect-based alternative proteins were the centres of alternative protein research. What is more, pea was focused the most rather than soy among all plants. With support from word association analysis, the insect-based protein was related to terms such as neophobia, cockroach, disgust, and novel while plant-based protein was associated with health and Asia. Furthermore, the insect-based protein contributed most of the negative comments in the text matrix. Correspondence analysis showed that there was no significant difference between emotions and protein categories. To develop the descriptive lexicon of alternative proteins, the text matrix's scope needs to be improved. Our study indicated that the application of text mining in sensory research is helpful when the text matrix is enormous. To improve the codes to achieve less manual work and obtain more reliable results, further research is required.

References

Agbemafle, I., Hadzi, D., Amagloh, F. K., Zotor, F. B., & Reddy, M. B. (2020). Nutritional, Microbial, and Sensory Evaluation of Complementary Foods Made from Blends of Orange-Fleshed Sweet Potato and Edible Insects. *Foods*, *9*(9), 1225.

Altmann, B. A., Neumann, C., Velten, S., Liebert, F., & Mörlein, D. (2018). Meat quality derived. from high inclusion of a micro-alga or insect meal as an alternative protein source in poultry diets: A pilot study. *Foods*, *7*(3), 34.

Ardoin, R., & Prinyawiwatkul, W. (2020). Product appropriateness, willingness to try and. perceived risks of foods containing insect protein powder: A survey of US consumers. *International Journal of Food Science & Technology*, *55*(9), 3215-3226.

Ares, G., Barreiro, C., Deliza, R., & Gámbaro, A. (2009). Alternatives to reduce the bitterness, astringency and characteristic flavour of antioxidant extracts. *Food Research International*, *42*(7), 871-878.

Badal, V. D., Kundrotas, P. J., & Vakser, I. A. (2015). Text mining for protein docking. *PLoS computational biology*, *11*(12), e1004630.

Baeshen, N. A., Baeshen, M. N., Sheikh, A., Bora, R. S., Ahmed, M. M. M., Ramadan, H. A., ... & Redwan, E. M. (2014). Cell factories for insulin production. *Microbial cell factories*, *13*(1), 1-9.

Bakhtin, P., Khabirova, E., Kuzminov, I., & Thurner, T. (2020). The future of food production–a text-mining approach. *Technology Analysis & Strategic Management*, *32*(5), 516-528.

Bécue-Bertaut, Mónica, Álvarez-Esteban, R., & Pagès, J. (2008). Rating of products through scores and free-text assertions: Comparing and combining both. *Food Quality and Preference, 19*(1), 122–134. https://doi.org/10.1016/j.foodqual.2007.07.006.

Bleakley, S., & Hayes, M. (2017). Algal proteins: extraction, application, and challenges concerning production. *Foods*, *6*(5), 33.

Chiang, J. H., Hardacre, A. K., & Parker, M. E. (2020). Effects of Maillard-reacted beef bone hydrolysate on the physicochemical properties of extruded meat alternatives. *Journal of food science*, *85*(3), 567-575.

Chow, C. Y., Riantiningtyas, R. R., Sørensen, H., & Frøst, M. B. (2021). School children cooking and eating insects as part of a teaching program–Effects of cooking, insect type, tasting order and food neophobia on hedonic response. *Food Quality and Preference*, *87*, 104027.

Cohen, K. B., & Hunter, L. (2008). Getting started in text mining. *PLoS Comput Biol*, *4*(1), e20.

Cosson, A., Delarue, J., Mabille, A. C., Druon, A., Descamps, N., Roturier, J. M., ... & Saint-Eve, A. (2020). Block protocol for conventional profiling to sensory characterise plant protein isolates. *Food Quality and Preference*, *83*, 103927.

De Koning, W., Dean, D., Vriesekoop, F., Aguiar, L. K., Anderson, M., Mongondry, P., ... & Boereboom, A. (2020). Drivers and inhibitors in the acceptance of meat alternatives: The case of plant and insect-based proteins. *Foods*, *9*(9), 1292.

Delarue, J., & Lawlor, B. (Eds.). (2014). Rapid sensory profiling techniques: Applications in new product development and consumer research.

Department of Agriculture, Water and the Environment of Australia. (2020). *Meat consumption*. Retrieved from https://www.agriculture.gov.au/abares/research-topics/agricultural-outlook/meat-consumption

El-Moniem, G. M. A. (1999). Sensory evaluation and in vitro protein digestibility of mung bean as affected by cooking time. *Journal of the Science of Food and Agriculture*, *79*(14), 2025-2028.

FAIRR, (2019). Meat alternatives and the alternative protein market. Retrieved from https://www.fairr.org/article/alternative-protein-market/

Fiorentini, M., Kinchla, A. J., & Nolden, A. A. (2020). Role of Sensory Evaluation in Consumer Acceptance of Plant-Based Meat Analogs and Meat Extenders: A Scoping Review. *Foods, 9(9), 1334.*

Foote, K. J., Joy, M. K., & Death, R. G. (2015). New Zealand dairy farming: milking our environment for all its worth. *Environmental management*, *56*(3), 709-720.
From https://www.linguamatics.com/what-text-mining-text-analytics-and-natural-language-processing

García-Segovia, P., Igual, M., & Martínez-Monzó, J. (2020). Physicochemical properties and consumer acceptance of bread enriched with alternative proteins. *Foods*, *9*(7), 933.

Giacomelli, M., Salata, C., Martini, M., Montesissa, C., & Piccirillo, A. (2014). Antimicrobial resistance of Campylobacter jejuni and Campylobacter coli from poultry in Italy. *Microbial Drug Resistance, 20*(2), 181-188.

Gómez, I., Ibañez, F. C., & Beriain, M. J. (2019). Physicochemical and sensory properties of sous vide meat and meat analog products marinated and cooked at different temperature-time combinations. *International Journal of Food Properties*, *22*(1), 1693-1708.

Gómez-Luciano, C. A., de Aguiar, L. K., Vriesekoop, F., & Urbano, B. (2019). Consumers' willingness to purchase three alternatives to meat proteins in the United Kingdom, Spain, Brazil and the Dominican Republic. *Food quality and preference*, *78*, 103732.

González, A., & Koltrowitz, S., (2019). The $280,000 lab-grown burger could be a more palatable $10 in two years. Retrieved from https://www.reuters.com/article/us-food-tech-labmeat/the-280000-lab-grown-burger-could-be-a-more-palatable-10-in-two-years-idUSKCN1U41W8

Gorr, S. U., Wennblom, T. J., Horvath, S., Wong, D. T., & Michie, S. A. (2012). Text-mining applied to autoimmune disease research: the Sjögren's syndrome knowledge base. *BMC musculoskeletal disorders*, *13*(1), 119.

Grahl, S., Palanisamy, M., Strack, M., Meier-Dinkel, L., Toepfl, S., & Mörlein, D. (2018). Towards. more sustainable meat alternatives: How technical parameters affect the sensory properties of extrusion products derived from soy and algae. *Journal of Cleaner Production*, *198*, 962-971.

Grahl, S., Strack, M., Mensching, A., & Mörlein, D. (2020). Alternative protein sources in Western diets: Food product development and consumer acceptance of spirulina-filled pasta. *Food Quality and Preference*, *84*, 103933.

Grasso, A. C., Hung, Y., Olthof, M. R., Verbeke, W., & Brouwer, I. A. (2019). Older consumers' readiness to accept alternative, more sustainable protein sources in the European Union. *Nutrients*, *11*(8), 1904.

Hamilton, M. L., & Jacob, L. (2020). Fast and automated sensory analysis: Using natural language processing for descriptive lexicon development. *Food Quality and Preference*, 103926.

Heymann, H., King, E. S., & Hopfer, H. (2014). Classical Descriptive Analysis. In P. Varela & G. Ares (Eds.), Novel Techniques in Sensory Characterization and Consumer Profiling (pp. 9–40). CRC Press.

Ickes, C. M., Lee, S.-Y., & Cadwallader, K. R. (2017). Novel creation of a rum flavor lexicon through the use of web-based material. *Journal of Food Science, 82*(5), 1216–1223.

Kaleda, A., Talvistu, K., Tamm, M., Viirma, M., Rosend, J., Tanilas, K., ... & Tammik, M. L. (2020). Impact of fermentation and phytase treatment of pea-oat protein blend on physicochemical, sensory, and nutritional properties of extruded meat analogs. *Foods*, *9*(8), 1059.

Kamani, M. H., Meera, M. S., Bhaskar, N., & Modi, V. K. (2019). Partial and total replacement of meat by plant-based proteins in chicken sausage: Evaluation of mechanical, physico-chemical and sensory characteristics. *Journal of food science and technology*, *56*(5), 2660-2669.

Kamani, M. H., Meera, M. S., Bhaskar, N., & Modi, V. K. (2019). Partial and total replacement of meat by plant-based proteins in chicken sausage: Evaluation of mechanical, physico-chemical and sensory characteristics. *Journal of food science and technology*, *56*(5), 2660-2669.

Klunklin, W., & Savage, G. (2018a). Physicochemical properties and sensory evaluation of wheat-purple rice biscuits enriched with green-lipped mussel powder (Perna canaliculus) and spices. *Journal of Food Quality, 2018*.

Klunklin, W., & Savage, G. (2018b). Addition of defatted green-lipped mussel powder and mixed spices to wheat–purple rice flour biscuits: Physicochemical, in vitro digestibility and sensory evaluation. *Food Science & Nutrition*, *6*(7), 1839-1847.

Lawless, H. T., & Heymann, H. (2010). *Sensory evaluation of food: principles and practices*. Springer Science & Business Media.

Leaman, R., Wei, C. H., Allot, A., & Lu, Z. (2020). Ten tips for a text-mining-ready article: How to improve automated discoverability and interpretability. *Plos Biology*, *18*(6), e3000716.

Lee, M., Liu, Z., Kelly, R., & Tong, W. (2014). Of text and gene–using text mining methods to uncover hidden knowledge in toxicogenomics. *BMC systems biology*, *8*(1), 93.

Linguamatic, (2020). What is Text Mining, Text Analytics and Natural Language? Retrieved

Lv, B. L., (2017). *Deep-fry cockroach*. Retrieved from https://www.zhihu.com/column/p/30572346

Mancebo, C. M., Rodriguez, P., & Gómez, M. (2016). Assessing rice flour-starch-protein mixtures to produce gluten free sugar-snap cookies. *LWT-Food Science and Technology*, *67*, 127-132.

Marsh, D. (2012). Water resource management in New Zealand: jobs or algal blooms? *Journal of environmental management*, *109*, 33-42.

Martin, C., Lange, C., & Marette, S. (2021). Importance of additional information, as a complement to information coming from packaging, to promote meat substitutes: A case study on a sausage based on vegetable proteins. *Food Quality and Preference*, *87*, 104058.

Meilgaard, M. C., Civille, G. V., & Carr, B. T. (1991). Sensory evaluation techniques (2nd ed.). Boca Raton, FL: CRC Press.

Mishyna, M., Chen, J., & Benjamin, O. (2020). Sensory attributes of edible insects and insect-based foods–Future outlooks for enhancing consumer appeal. *Trends in Food Science & Technology*, *95*, 141-148.

Nawaz, M. F., Bourrie, G., & Trolard, F. (2013). Soil compaction impact and modelling. A review. *Agronomy for sustainable development*, *33*(2), 291-309.

Park, J., Choi, I., & Kim, Y. (2015). Cookies formulated from fresh okara using starch, soy flour and hydroxypropyl methylcellulose have high quality and nutritional value. *LWT-Food Science and Technology, 63(1), 660-666*.

Parniakov, O., Toepfl, S., Barba, F. J., Granato, D., Zamuz, S., Galvez, F., & Lorenzo, J. M. (2018). Impact of the soy protein replacement by legumes and algae-based proteins on the quality of chicken rotti. *Journal of food science and technology*, *55*(7), 2552-2559.

Possidónio, C., Prada, M., Graça, J., & Piazza, J. (2021). Consumer perceptions of conventional and alternative protein sources: A mixed-methods approach with meal and product framing. *Appetite*, *156*, 104860.

Post, M. J. (2012). Cultured meat from stem cells: Challenges and prospects. *Meat science*, *92*(3), 297-301.

R Core Team. (2019). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. Retrieved from https://www.r-project.org.

Rebholz-Schuhmann, D., Kirsch, H., & Couto, F. (2005). Facts from text—is text mining ready to deliver? *PLoS Biol*, *3*(2), e65.

Rodríguez-Limas, W. A., Tannenbaum, V., & Tyo, K. E. (2015). Blocking endocytotic mechanisms to improve heterologous protein titers in Saccharomyces cerevisiae. *Biotechnology and bioengineering*, *112*(2), 376-385.

Rolland, N. C., Markus, C. R., & Post, M. J. (2020). The effect of information content on acceptance of cultured meat in a tasting context. *Plos one*, *15*(4), e0231176.

Sălăgean, C. D., Pop, C., Catrinoi, M., & Nagy, M. (2015). Influence of the Brown Marine Algae on the Physicochemical and Sensory Characteristcs of the Sausages. *Bulletin UASVM Food Science and Technology*, *72*(2), 193-199.

Schouteten, J. J., De Steur, H., De Pelsmaeker, S., Lagast, S., Juvinal, J. G., De Bourdeaudhuij, I., ... & Gellynck, X. (2016). Emotional and sensory profiling of insect-, plant-and meat-based burgers under blind, expected and informed conditions. *Food Quality and Preference*, *52*, 27-31.

Sha, L., & Xiong, Y. L. (2020). Plant protein-based alternatives of reconstructed meat: Science, technology, and challenges. *Trends in Food Science & Technology*.

Spinelli, S., Dinnella, C., Masi, C., Zoboli, G. P., Prescott, J., & Monteleone, E. (2017). Investigating preferred coffee consumption contexts using open-ended questions. *Food Quality and Preference*, *61*, 63-73.

Stephan, A., Ahlborn, J., Zajul, M., & Zorn, H. (2018). Edible mushroom mycelia of Pleurotus sapidus as novel protein sources in a vegan boiled sausage analog system: functionality and sensory tests in comparison to commercial proteins and meat sausages. *European Food Research and Technology*, *244*(5), 913-924.

Tao, D., Yang, P., & Feng, H. (2020). Utilisation of text mining as a big data analysis tool for food science and nutrition. *Comprehensive Reviews in Food Science and Food Safety*, *19*(2), 875-894.

Tucker, C. A. (2014). The significance of sensory appeal for reduced meat consumption. *Appetite*, *81*, 168-179.

Van Krimpen, M. M., Bikker, P., Van der Meer, I. M., Van der Peet-Schwering, C. M. C., & Vereijken, J. M. (2013). *Cultivation, processing and nutritional aspects for pigs and poultry of European protein sources as alternatives for imported soybean products* (No. 662). Wageningen UR Livestock Research.

Verbeke, W., Sans, P., & Van Loo, E. J. (2015). Challenges and prospects for consumer acceptance of cultured meat. *Journal of Integrative Agriculture*, *14*(2), 285-294.

Vieira Gomes, A. M., Souza Carmo, T., Silva Carvalho, L., Mendonça Bahia, F., & Parachin, N. S. (2018). Comparison of yeasts as hosts for recombinant protein production. *Microorganisms*, *6*(2), 38.

Wong, K. M., Corradini, M. G., Autio, W., & Kinchla, A. J. (2019). Sodium reduction strategies through use of meat extenders (white button mushrooms vs. textured soy) in beef patties. *Food science & nutrition*, *7*(2), 506-518.

Yuliarti, O., Kovis, T. J. K., & Yi, N. J. (2021). Structuring the meat analogue by using plant-based derived composites. *Journal of Food Engineering*, *288*, 110138.

Zong, C. M., Wang, H. B., & Du, X. J. (2010). The Application of Data Mining in the Food Manufacturing Industry. In *Applied Mechanics and Materials* (Vol. 20, pp. 856-860). Trans Tech Publications Ltd.

Appendix A

PDF document text mining codes and explanation produced by Cristhiam Gurdian

```
#' ---

#' title:  Text mining trial

#' author:  Cristhiam Gurdian (cgurdi3@lsu.edu)

#' date:   2021-January-06

#' ---


install.packages("pdftools")

library(pdftools)

#set working directory to the folder that contains the pdf files

#remove from the pdf names greek characters or symbols because the lappy function will not work with those


#create a vector of PDF file names using the list.files function.

#The pattern argument says to only grab those files ending with “pdf”:

files <- list.files(pattern = “pdf$”)#only works if you have your working directory set to the folder where you downloaded the PDF files

files #The “files” vector contains all the PDF file names. We’ll use this vector to automate the process of reading in the text of the PDF files.
```

#The pdftools function for extracting text is pdf_text.

#Using the lapply function, we can apply the pdf_text function to each element in the “files” vector and create an object called “text”.

text <- lapply(files, pdf_text) #This creates a list object with three elements, one for each document.

length(text)

#Each element in “text” is a vector that contains the text of the PDF file.

lapply(text, length) #The length of each vector corresponds to the number of pages in the PDF file.

####USING TEXT MINING PACKAGE FOR TEXT ANALYSIS

#First load tm package and then create a corpus, which is a database for text.

#instead of working with the “opinions” text” object we created earlier, we start over.

install.packages("tm")

library(tm)

corp <- Corpus(URISource(files),

readerControl = list(reader = readPDF))#The Corpus function creates a corpus. The first argument to Corpus is what we want to use to create the corpus.

#In this case, it’s the vector of PDF files. To do this, we use the URISource function to indicate that the files vector is a URI (Uniform Resource Identifier) source.

we’re telling the Corpus function that the vector of file names identifies our resources.

#The second argument, readerControl, tells Corpus which reader to use to read in the text from the PDF files (readPDF, a tm function).

#The readerControl argument requires a list of control parameters, one of which is reader, so we enter list(reader = readPDF).

#Finally we save the result to an object called “corp”.

#Now that we have a corpus, we can create a term-document matrix, (TDM) that stores counts of terms for each document.

#The tm package provides a function to create a TDM called TermDocumentMatrix.

```
library(SnowballC)

text.tdm <- TermDocumentMatrix(corp,
                               control =
                                list(removePunctuation = TRUE,
                                     stopwords = TRUE,
                                     tolower = TRUE,
                                     stemming = TRUE,
                                     removeNumbers = TRUE,
                                     bounds = list(global = c(3, Inf))))
```

#The first argument is our corpus. The second argument is a list of control parameters.

#clean up the corpus before creating the TDM. Remove punctuation, stopwords (eg, the, of, in, etc.), convert text to lower case, stem the words,

#remove numbers, and only count words that appear at least 3 times. We save the result to an object called "text.tdm".

```
inspect(text.tdm[1:10,])#first 10 terms
```

#pdf_text function may preserve the unicode curly-quotes and em-dashes used in the PDF files.

```
#manually use the removePunctuation function with tm_map, both functions in the tm package.

#removePunctuation function has an argument called ucp that when set to TRUE will look for unicode punctuation.

corp <- tm_map(corp, removePunctuation, ucp = TRUE)

#re-create the TDM, this time without the removePunctuation = TRUE argument.

text.tdm <- TermDocumentMatrix(corp,
                    control =
                     list(stopwords = TRUE,
                       tolower = TRUE,
                       stemming = TRUE,
                       removeNumbers = TRUE,
                       bounds = list(global = c(3, Inf))))

inspect(text.tdm[1:10,])#first 10 terms


#findFreqTerms function to find words that occur at least 100 times:

findFreqTerms(text.tdm, lowfreq = 100, highfreq = Inf)


#To see the counts of those words we could save the result and use it to subset the TDM.

#we have to use as.matrix to see the print out of the subsetted TDM.

ft <- findFreqTerms(text.tdm, lowfreq = 100, highfreq = Inf)

as.matrix(text.tdm[ft,])
```

```
#To see thews

ft.tdm <- as.matrix(text.tdm[ft,])

sort(apply(ft.tdm, 1, sum), decreasing = TRUE)
```

Appendix B

TXT document text mining codes retrieved from Internet

```
library("SnowballC")

library("RColorBrewer")

library("wordcloud")

library("syuzhet")

library("ggplot2")

library("tm")


text <- readLines(file.choose())

TextDoc <- Corpus(VectorSource(text))


toSpace <- content_transformer(function (x , pattern ) gsub(pattern, " ", x))

TextDoc <- tm_map(TextDoc, toSpace, "/")

TextDoc <- tm_map(TextDoc, toSpace, "@")

TextDoc <- tm_map(TextDoc, toSpace, "\\|")


TextDoc <- tm_map(TextDoc, content_transformer(tolower))

TextDoc <- tm_map(TextDoc, removeNumbers)

TextDoc <- tm_map(TextDoc, removeWords, stopwords("english"))

TextDoc <- tm_map(TextDoc, removeWords, c("s", "company", "team"))
```

```
TextDoc <- tm_map(TextDoc, removePunctuation)

TextDoc <- tm_map(TextDoc, stripWhitespace)

TextDoc <- tm_map(TextDoc, stemDocument)


TextDoc_dtm <- TermDocumentMatrix(TextDoc)

dtm_m <- as.matrix(TextDoc_dtm)

dtm_v <- sort(rowSums(dtm_m),decreasing=TRUE)

dtm_d <- data.frame(word = names(dtm_v),freq=dtm_v)

head(dtm_d, 50)


barplot(dtm_d[1:50,]$freq, las = 2, names.arg = dtm_d[1:50,]$word,

    col ="lightgreen", main ="Top 50 most frequent words",

    ylab = "Word frequencies")


set.seed(1234)

wordcloud(words = dtm_d$word, freq = dtm_d$freq, min.freq = 5,

     max.words=100, random.order=FALSE, rot.per=0.40,

     colors=brewer.pal(8, "Dark2"))


findAssocs(TextDoc_dtm, terms = c("insect","flavor","like"), corlimit = 0.25)


syuzhet_vector <- get_sentiment(text, method="syuzhet")

head(syuzhet_vector)
```

```
summary(syuzhet_vector)


bing_vector <- get_sentiment(text, method="bing")

head(bing_vector)

summary(bing_vector)


afinn_vector <- get_sentiment(text, method="afinn")

head(afinn_vector)

summary(afinn_vector)


d<-get_nrc_sentiment(text)

td<-data.frame(t(d))

td_new <- data.frame(rowSums(td[2:253]))

names(td_new)[1] <- "count"

td_new <- cbind("sentiment" = rownames(td_new), td_new)

rownames(td_new) <- NULL

td_new2<-td_new[1:8,]

quickplot(sentiment, data=td_new2, weight=count, geom="bar", fill=sentiment,
ylab="count")+ggtitle("Survey sentiments")


barplot(

  sort(colSums(prop.table(d[, 1:8]))),

  horiz = TRUE,
```

```
  cex.names = 0.7,

  las = 1,

  main = "Emotions in Text", xlab="Percentage"

)
```

Appendix C

The list of scientific reports that analysed by Natural language processing

Number	Title	Source
1	Role of Sensory Evaluation in Consumer Acceptance of Plant-Based Meat Analogs and Meat Extenders: A Scoping Review	*Food*
2	Drivers and Inhibitors in the Acceptance of Meat Alternatives: The Case of Plant and Insect-Based Proteins	*Food*
3	Nutritional, Microbial, and Sensory Evaluation of Complementary Foods Made from Blends of Orange-Fleshed Sweet Potato and Edible Insects	*Food*
4	Impact of Fermentation and Phytase Treatment of Pea-Oat Protein Blend on Physicochemical, Sensory, and Nutritional Properties of Extruded Meat Analogs	*Food*
5	Physicochemical Properties and Consumer Acceptance of Bread Enriched with Alternative Proteins	*Food*
6	Meat Quality Derived from High Inclusion of a Micro-Alga or Insect Meal as an Alternative Protein Source in Poultry Diets: A Pilot Study	*Food*
7	Consumer perceptions of conventional and alternative protein sources: A mixed-methods approach with meal and product framing	*Appetite*
8	School children cooking and eating insects as part of a teaching program - Effects of cooking, insect	*Food Quality and Preference*

	type, tasting order and food neophobia on hedonic response	
9	Importance of additional information, as a complement to information coming from packaging, to promote meat substitutes: A case study on a sausage based on vegetable protein	*Food Quality and Preference*
10	Structuring the meat analogue by using plant-based derived composites	*Journal of Food Engineering*
11	Effects of Maillard-reacted beef bone hydrolysate on the physicochemical properties of extruded meat alternatives	*Journal of Food Science*
12	Edible mushroom mycelia of Pleurotus sapidus as novel protein sources in a vegan boiled sausage analog system: functionality and sensory tests in comparison to commercial proteins and meat sausages	*European Food Research and Technology*
13	Towards more sustainable meat alternatives: How technical parameters affect the sensory properties of extrusion products derived from soy and algae	*Cleaner Production*
14	Block protocol for conventional profiling to sensory characterise plant protein isolates	*Food Quality and Preference*
15	Alternative protein sources in Western diets: Food product development and consumer acceptance of spirulina-filled pasta	*Food Quality and Preference*
16	Older Consumers' Readiness to Accept Alternative, More Sustainable Protein Sources in the European Union	*Nutrients*
17	Plant protein-based alternatives of reconstructed meat: Science, technology, and challenges	*Trends in Food Science and Technology*

18	Product appropriateness, willingness to try and perceived risks of foods containing insect protein powder: A survey of US consumers	*International Journal of Food Science and Technology*
19	Partial and total replacement of meat by plant-based proteins in chicken sausage: evaluation of mechanical, physico-chemical and sensory characteristics	*Journal of Food Science and Technology*
20	Sensory attributes of edible insects and insect-based foods - Future outlooks for enhancing consumer appeal	*Trends in Food Science and Technology*

Appendix D

The frequency of word in text matrix (top 50)

Word	Frequency	Word	Frequency
Ffs	697	plantbas	97
meat	531	also	96
protein	432	sourc	96
product	404	compar	85
food	356	tabl	84
consum	264	expect	83
altern	188	content	82
studi	181	pea	82
insect	179	structur	82
tast	167	present	78
flavor	165	sustain	78
Use	161	howev	77
sensori	154	mrp	77
textur	142	high	76
differ	141	spirulina	76
attribut	136	posit	73
can	128	addit	73
accept	124	mayb	71
sampl	121	process	71
increas	116	develop	69
effect	111	show	69
like	110	analogu	68
result	107	plant	67
evalu	105	substitut	67
signific	100		

Appendix E

The relevance between key words and other words

```
$meat
      altern     product   substitut    labgrown     barrier        lack         use       addit       skill
        0.62        0.51        0.50        0.44        0.42        0.37        0.36        0.36        0.36
    consumpt     process         red       studi      includ       reduc      analog         one         way
        0.35        0.35        0.34        0.34        0.34        0.34        0.33        0.33        0.33
        diet       clean         can      toward    concentr        meal       graea      bryant        also
        0.33        0.32        0.31        0.31        0.30        0.30        0.30        0.30        0.29
        term     particip         due     benefit environment    address       chang      impact         may
        0.29        0.29        0.29        0.29        0.29        0.29        0.28        0.28        0.28
     contain        help       lower        tofu   animalfre       enjoy       monik        pose        unfa
        0.28        0.28        0.28        0.28        0.28        0.28        0.28        0.28        0.28
      assess         mrp       natur        fish       organ      seitan       frame     attract     perceiv
        0.27        0.27        0.27        0.27        0.27        0.27        0.27        0.27        0.26
      compar      differ     ingredi      replac    behavior       label        hoek      improv      nutrit
        0.26        0.26        0.26        0.26        0.26        0.26        0.26        0.25        0.25
    research     percept    siegrist       white       mimic
        0.25        0.25        0.25        0.25        0.25

$protein
       sourc   animalbas   singlecel      altern       vitro    structur       found       studi   crosslink
        0.60        0.41        0.41        0.40        0.40        0.39        0.38        0.37        0.37
      extent      gluten        dont         use       wheat environment     contain       dairi     various
        0.37        0.36        0.35        0.34        0.34        0.34        0.33        0.33        0.33
     product     without     fibrous     belgium     process       intak  seafoodbas     consist     techniqu
        0.31        0.31        0.31        0.31        0.30        0.30        0.30        0.29        0.29
    although       freez         lin     meatbas       deriv       grasp       salad     student     develop
        0.29        0.29        0.29        0.29        0.29        0.29        0.29        0.29        0.28
    properti       uniqu       major        know         egg      nutrit     sustain         due    research
        0.28        0.28        0.28        0.28        0.28        0.27        0.27        0.27        0.27
        wide  viscoelast        land         eat     analogu      manner       aquat   biodivers businessusu
        0.27        0.27        0.27        0.27        0.27        0.27        0.27        0.27        0.27
       emiss    freshwat        ghge   greenhous    nitrogen   terrestri    threaten       chang      impact
        0.27        0.27        0.27        0.27        0.27        0.27        0.27        0.26        0.26
        also        form       total       muscl      requir  furthermor    meatlik   manufactur      buffer
        0.26        0.26        0.26        0.26        0.26        0.26        0.26        0.26        0.26
     solubil         soy        vari      nugget       ideal         com   agreement      expand      specul
        0.26        0.25        0.25        0.25        0.25        0.25        0.25        0.25        0.25
    globular        mung
        0.25        0.25

$consum
    product   research    consumpt   substitut     percept        food       studi     barrier      altern      prefer
       0.47       0.38        0.37        0.37        0.36        0.35        0.34        0.34        0.33        0.33
     specif     market        hoek         may      method    identifi    behavior     pathway         can         way
       0.33       0.31        0.31        0.30        0.30        0.30        0.30        0.30        0.29        0.29
     target     qualit         tri      toward         one       howev        find         new     develop      impact
       0.29       0.29        0.29        0.29        0.28        0.28        0.28        0.28        0.27        0.27
    suggest      adopt  understand      rather         buy        diet       pasta       dutch      appeal    approach
       0.27       0.27        0.27        0.27        0.27        0.27        0.27        0.27        0.26        0.26
        aim      singl     attitud        mani      academ      freeli     reliabl    roininen     alterna        tive
       0.26       0.26        0.26        0.26        0.26        0.26        0.26        0.26        0.26        0.26
    acknowl        edg      tailor   wellknown  disappoint    familiar        test       sourc     willing
       0.26       0.26        0.26        0.26        0.26        0.25        0.25        0.25        0.25
```

```
$insect
      ball    oatmeal   children       food    caparro      edibl   mealworm     ashmor  childhood     dunlop
      0.57       0.57       0.54       0.47       0.46       0.45       0.44       0.42       0.42       0.42
familiaris       gust      haidt       hood   mccauley     medium   realfood   tinguish     troduc      rozin
      0.42       0.42       0.42       0.42       0.42       0.42       0.42       0.42       0.42       0.40
    megido    willing neophobia     respons    barbera   selfcook        may      first   recognis     chitin
      0.39       0.38       0.37       0.36       0.36       0.35       0.34       0.34       0.34       0.34
      choi   tortilla  grasshopp       cite  cockroach     covari        eli      gmuer       guth   lensvelt
      0.34       0.34       0.34       0.34       0.34       0.34       0.34       0.34       0.34       0.34
   nuessli  steenbekk        ver      hedon    disgust        tri      novel      might      studi    version
      0.34       0.34       0.34       0.33       0.33       0.33       0.33       0.32       0.31       0.30
       bit     author        fao        eat     living     visibl        dis     implic      later     second
      0.29       0.29       0.29       0.29       0.29       0.29       0.29       0.29       0.29       0.28
    explor      moder       also       rate   previous    stimuli    belgian      among   influenc     measur
      0.28       0.28       0.27       0.27       0.27       0.27       0.27       0.26       0.26       0.26
     futur    western     matric       chip   encourag    decosta    continu
      0.26       0.26       0.26       0.26       0.26       0.26       0.25
```

```
$flavor
             mask         lemonbasil              extrud              lemon              tomato              spice
             0.62               0.59                0.55               0.54                0.49               0.47
      vegetablebas            content              enhanc           confound       contentflavor                dim
             0.46               0.44                0.43               0.42                0.42               0.42
          equidist          nonlinear            smallest               spir   spirulinaextrud               tens
             0.42               0.42                0.42               0.42                0.42               0.42
             ulina              sampl             beetging               odor       chickenflavor            crablik
             0.42               0.41                0.41               0.39                0.39               0.39
          katayama             narrow                oili             oyster           oysterlik              shred
             0.39               0.39                0.39               0.39                0.39               0.39
         shredshap             shrimp               strip          stripshap              volunt             wilson
             0.39               0.39                0.39               0.39                0.39               0.39
             basil         complement  cuellarbermfadez           foreseen             geosmin             glutam
             0.39               0.39                0.39               0.39                0.39               0.39
 methylisoborneol                mib      milovanoviucu            propriat        spirulinarel           steepest
             0.39               0.39                0.39               0.39                0.39               0.39
             spici            chicken             control             intens              soybas            beefbon
             0.38               0.38                0.37               0.37                0.37               0.37
          unflavor              pasta              linear         mustyearthi               meati               four
             0.37               0.37                0.37               0.36                0.35               0.35
            earthi              aroma              pepper               bake               salti              shape
             0.35               0.34                0.34               0.34                0.34               0.34
            compar             season            descript             analog                test              beefi
             0.33               0.33                0.33               0.32                0.32               0.32
           beeflik              crush             fortifi           fullmeat              grassi          justright
             0.32               0.32                0.32               0.32                0.32               0.32
             modif             pastri              peanut                ppc                puff               roll
             0.32               0.32                0.32               0.32                0.32               0.32
             smoke               stuf              versus             overal             resembl             provid
             0.32               0.32                0.32               0.31                0.31               0.31
             lipid          peanutbas                term              formul              spread         substanti
             0.31               0.31                0.30               0.30                0.30               0.30
           perform              level                 aim            pathway               musti                use
             0.29               0.29                0.29               0.29                0.29               0.28
             addit               beef                best               minc                 fri             assess
             0.28               0.28                0.28               0.28                0.28               0.27
               tsp               base                 bit           encapsul              former             liquid
             0.27               0.27                0.27               0.27                0.27               0.27
            pocket             qualit             quantit             though              method              scale
             0.27               0.27                0.27               0.27                0.26               0.26
           success              three                show            insight
             0.26               0.25                0.25               0.25
```

```
$sensori
       breslin            keast            evalu              use          bitter            test
          0.46             0.46             0.45             0.41            0.41            0.39
        assess         properti           mechan           overal           sampl           howev
          0.39             0.38             0.38             0.37            0.37            0.37
      descript           tomato            addit           compar        panelist         analysi
          0.37             0.36             0.35             0.35            0.35            0.34
        result             base         quantifi   multidimension         control          sensat
          0.34             0.34             0.34             0.34            0.33            0.33
        sensit           bakker            busch         clifford         considin         danguin
          0.33             0.33             0.33             0.33            0.33            0.33
     delahunti          heymann       kallithraka            kfchn  neurophysiolog            pend
          0.33             0.33             0.33             0.33            0.33            0.33
        precis          prolong         reitmeier             sall         septier           singh
          0.33             0.33             0.33             0.33            0.33            0.33
         thoma  torrespenaranda            trice            bajec       bartoshuk          bolhui
          0.33             0.33             0.33             0.33            0.33            0.33
     capsaicin           carney          cessari        chemesthet         ciceral           duffi
          0.33             0.33             0.33             0.33            0.33            0.33
          file           hansen        hollowood             hort             iii           irrit
          0.33             0.33             0.33             0.33            0.33            0.33
          kidd           lanier npropylthiouracil          picker         pothesi           proxi
          0.33             0.33             0.33             0.33            0.33            0.33
       pungenc           sociat       swaincampbel           taster            wish          wright
          0.33             0.33             0.33             0.33            0.33            0.33
       product            agent             past         dinehart          method      understand
          0.32             0.32             0.32             0.32            0.31            0.31
          oral          percept        character           cognit         persist             cog
          0.31             0.31             0.31             0.30            0.30            0.30
        fatigu         guichard            nitiv             omit        beetging          analog
          0.30             0.30             0.30             0.30            0.30            0.29
      attribut              can         research          complex             hay        prescott
          0.29             0.29             0.29             0.29            0.29            0.29
    instrument            score           affect         interest        interact         pathway
          0.28             0.28             0.28             0.28            0.28            0.28
        combin           differ          beefbon         unflavor          liquid            reed
          0.27             0.27             0.27             0.27            0.27            0.27
          webb            studi         fullmeat             mask       physiolog      lemonbasil
          0.27             0.26             0.26             0.26            0.26            0.26
          show            limit           better          qualiti             spi            prop
          0.25             0.25             0.25             0.25            0.25            0.25

$textur
     properti     physiochem         nugget         analog            use        techniqu            ratio
         0.49           0.47           0.43           0.37           0.37            0.37             0.37
      analysi            soy     instrument          howev          optim         emulsifi           capabl
         0.36           0.35           0.35           0.35           0.35            0.35             0.35
         judg     microscopi           scan       yuliarti       structur         analogu             form
         0.35           0.35           0.35           0.35           0.34            0.34             0.33
          gum          freez          layer         gluten        develop             show            capac
         0.33           0.33           0.33           0.32           0.31            0.31             0.31
        emuls        fibrous          ideal      mouthfeel   hydrocolloid             icgn iotacarrageenan
         0.31           0.31           0.31           0.30           0.29            0.29             0.29
       analyt       creation         extrus           soup       composit             test           result
         0.29           0.29           0.29           0.29           0.29            0.28             0.28
        anoth         molecul          wheat           name      crosslink            desir             term
         0.28           0.28           0.28           0.28           0.28            0.27             0.27
       affect            tvp        generat         includ          uniqu           sausag            sampl
         0.27           0.27           0.27           0.27           0.27            0.27             0.26
       method        control      viscoelast        thicken            ppi              gel        synergist
         0.26           0.26           0.26           0.26           0.26            0.26             0.26
      process       commerci
         0.25           0.25
```

```
$accept
      vitro          eat        sourc         dont      belgium      control      sustain    singlecel
       0.49         0.47         0.46         0.45         0.43         0.41         0.41         0.41
 seafoodbas       analog        sampl        studi      influenc       compar        found       specul
       0.41         0.40         0.39         0.39         0.39         0.37         0.37         0.37
   particip      meatbas         know        grasp        salad      student          may        score
       0.35         0.35         0.35         0.35         0.35         0.35         0.34         0.34
   determin         four      without       overal       manner      product         show         food
       0.34         0.34         0.34         0.33         0.33         0.32         0.32         0.32
      choic        meati       result         term        scale       extent        adult      surpris
       0.32         0.32         0.31         0.31         0.31         0.31         0.31         0.30
    contain     fullmeat     although    agreement      analysi        hedon      perform       provid
       0.30         0.30         0.30         0.30         0.29         0.29         0.29         0.29
     enhanc        smoke      beefbon     unflavor      barrier       liquid     research      concept
       0.29         0.29         0.29         0.29         0.29         0.29         0.29         0.29
      motiv        older      competit       eventu healthpromot industryscal        kilo      lowcost
       0.29         0.29         0.29         0.29         0.29         0.29         0.29         0.29
        owe singlecellbas    slaughter    transport     vitrobas         test        addit        adequ
       0.29         0.29         0.29         0.29         0.29         0.28         0.28         0.28
      level      conduct  vegetablebas  environment        price          can     colleagu      suggest
       0.28         0.28         0.28         0.28         0.28         0.27         0.27         0.27
     altern        spici        agent          use       reveal       affect      explain        spice
       0.27         0.27         0.27         0.26         0.26         0.26         0.26         0.26
   descript   unfamiliar     previous       visibl      respond
       0.26         0.26         0.26         0.26         0.25

$increas
        decreas          concentr               mrp       carotenoid sharimaabdullah          appear         chickpea
           0.46              0.44              0.44              0.39            0.39            0.38             0.38
           fact             flour               due            lowest           addit           world            chang
           0.34              0.33              0.32              0.32            0.29            0.29             0.28
         import             score               tvp      flavorenhanc        hydrolyz           oppos            moist
           0.27              0.27              0.26              0.26            0.26            0.26             0.26
            led        colorimetr              erti              mrps       placement          broken            creas
           0.26              0.26              0.26              0.26            0.26            0.26             0.26
         melton              oliv           polymer               sds          solubi         solvent          stanley
           0.26              0.26              0.26              0.26            0.26            0.26             0.26
         teract          twocompon       watersolubl            fabric          hirata          intact           altern
           0.26              0.26              0.26              0.26            0.26            0.26             0.25

$like
            pasta            pathway           beetging            credenc               leav           outstrip
             0.61               0.47               0.42               0.42               0.42               0.42
            sound             tomato         lemonbasil           conceptu         mustyearthi            ceptual
             0.42               0.41               0.41               0.40               0.38               0.38
           compar          experienc             overal               cata                met              basil
             0.33               0.32               0.31               0.30               0.30               0.30
       complement cuellarbermfadez           foreseen            geosmin             glutam methylisoborneol
             0.30               0.30               0.30               0.30               0.30               0.30
              mib      milovanoviucu           propriat        spirulinarel           steepest                tsp
             0.30               0.30               0.30               0.30               0.30               0.29
            negat             higher               mask              bland         netherland              femal
             0.29               0.27               0.27               0.27               0.27               0.27
            stand
             0.25

$plantbas
     composit         impl       makeup satisfactori    smallscal        struc     vestment       nutrit
         0.49         0.46         0.46         0.46         0.46         0.46         0.46         0.43
        freez       nugget     techniqu      develop       imposs     diversif      pietsch      prevail
         0.43         0.42         0.41         0.40         0.38         0.37         0.37         0.37
         rear        round          sur     surround        unsus       mentat          ppn    substitut
         0.37         0.37         0.37         0.37         0.37         0.37         0.37         0.36
      process      structur        mimic    animalbas      generat       altern      regular        uniqu
         0.34         0.34         0.34         0.33         0.32         0.32         0.32         0.32
  necessarili         home      product        littl         issu      meatlik      analogu        studi
         0.32         0.32         0.31         0.31         0.31         0.31         0.30         0.29
       recent         date      ingredi      strategi        natur      potenti   constraint         rais
         0.29         0.29         0.28         0.27         0.27         0.27         0.27         0.27
          use         vari         anim       health        everi  insectsbas       stigma      resourc
         0.26         0.26         0.26         0.26         0.26         0.26         0.26         0.26
      envisag         face       terest          uni         vers       dekker        xiong         amin
         0.26         0.26         0.26         0.26         0.26         0.26         0.26         0.26
     barzegar       berman   carcinogen        cheng       dioxid       excess        jiang    kamankesh
         0.26         0.26         0.26         0.26         0.26         0.26         0.26         0.26
         khan        kuhnl       nadeem     predispos      preserv       rahman        sahar thylcellulos
         0.26         0.26         0.26         0.26         0.26         0.26         0.26         0.26
     titanium        truli characterist      includ      conduct       specif         asia      tainabl
         0.26         0.26         0.25         0.25         0.25         0.25         0.25         0.25
      rheolog
         0.25
```

```
$expect
       assimil             deliza              macfi             actual            discrep            predict
          0.67               0.67               0.67               0.63               0.59               0.57
        dispar            distast              gativ           kehkfnen               kole            latitud
          0.56               0.56               0.56               0.56               0.56               0.56
         mojet             sawyer     schifferstein               sect         sensorylik               wors
          0.56               0.56               0.56               0.56               0.56               0.56
    disconfirm              brain               frog           cardello           contrast             experi
          0.55               0.55               0.55               0.51               0.43               0.43
      hypothet             pectat            unclear              pasta             chines             magnif
          0.43               0.43               0.43               0.40               0.39               0.39
      assimila               chil              chocol        contradict               dren           feedback
          0.39               0.39               0.39               0.39               0.39               0.39
         posur             pothet              spons             switch     tioncontrast             upward
          0.39               0.39               0.39               0.39               0.39               0.39
       pathway expectationexperi               lamb             adjust                neo             phobia
          0.38               0.37               0.36               0.35               0.35               0.35
    disappoint           familiar            credenc               leav           outstrip              sound
          0.35               0.33               0.33               0.33               0.33               0.33
        presum             effect               thus            perform             german                now
          0.32               0.31               0.31               0.31               0.31               0.31
         novel        crosscultur               food             reject                tan            ceptual
          0.30               0.30               0.29               0.29               0.29               0.29
          lead          predictor                gap            fischer            stieger              small
          0.28               0.28               0.28               0.28               0.28               0.27
         model              palat                enc            tuorila             postul          irrespect
          0.27               0.27               0.27               0.27               0.27               0.27
      somewhat              trast            obvious           conceptu                met              dutch
          0.27               0.27               0.27               0.27               0.27               0.26
        resist                via
          0.25               0.25

$pea
    starchi       lupin       grain     chamber       broth        bott        defi   mushroomi      nition
       0.54        0.54        0.53        0.53        0.53        0.52        0.51        0.51        0.51
        ope        ubol        vara    varaubol     astring     powderi       dusti       green     definit
       0.51        0.51        0.51        0.51        0.50        0.50        0.49        0.45        0.44
      beani         pod         nut      earthi       fvalu        resp     spectiv    attribut      crimin
       0.42        0.41        0.40        0.38        0.38        0.38        0.38        0.37        0.37
     tribut       attri        bute      almond  newmankeul       solut      bitter        puls   australia
       0.36        0.36        0.36        0.36        0.36        0.35        0.34        0.33        0.32
     canada cholesterol      fortif       manag      repair     satieti   sportmind         usa       musti
       0.32        0.32        0.32        0.32        0.32        0.32        0.32        0.32        0.32
    describ    protocol       block   discrimin       among      phenol     display       devel   dimension
       0.31        0.31        0.31        0.30        0.29        0.29        0.29        0.29        0.29
      multi    panelist     classic     consist       degre      launch         six        alum    granular
       0.29        0.28        0.28        0.27        0.27        0.27        0.27        0.27        0.27
   threeway       nutti  manufactur         hay      ambigu       arcot      brothi  chandrahio       drake
       0.27        0.26        0.26        0.26        0.26        0.26        0.26        0.26        0.26
     ejszo  kaczmarska      litera lopetcharat  malcolmson        pars         ska        soil szymkiewicz
       0.26        0.26        0.26        0.26        0.26        0.26        0.26        0.26        0.26
  troszyucu       woodi         wou        sour
       0.26        0.26        0.26        0.25
```

```
$spirulina
         lemonbasil        mustyearthi               mask              pasta              lemon             though
               0.64               0.61               0.54               0.53               0.49               0.48
spirulinasoyextrud             promis            biomass           beetging            content             extrud
               0.48               0.47               0.47               0.44               0.43               0.43
             tomato            peratur                tem              grahl               fill              futur
               0.43               0.42               0.42               0.42               0.42               0.39
               turn               step         expenditur               ject             someth              stood
               0.39               0.39               0.39               0.39               0.39               0.39
               usag              basil         complement   cuellarbermfadez           foreseen            geosmin
               0.39               0.39               0.39               0.39               0.39               0.39
             glutam   methylisoborneol                mib       milovanoviucu            propriat        spirulinarel
               0.39               0.39               0.39               0.39               0.39               0.39
           steepest                bao             foundat           intensif                ren              zheng
               0.39               0.39               0.39               0.39               0.39               0.39
               take          microalga               even               part            pathway               firm
               0.38               0.37               0.37               0.36               0.36               0.34
             earthi             intens               serv              musti           confound      contentflavor
               0.34               0.33               0.33               0.33               0.33               0.33
                dim           equidist          nonlinear           smallest               spir    spirulinaextrud
               0.33               0.33               0.33               0.33               0.33               0.33
               tens              ulina              input            justifi         netherland            insight
               0.33               0.33               0.32               0.32               0.32               0.30
         palanisami            account              still               gain            investig               goal
               0.29               0.29               0.29               0.29               0.28               0.28
          standalon              share               deep              screw              speed              entir
               0.28               0.28               0.28               0.28               0.28               0.28
               soon                sus         preprocess               need               left           conceptu
               0.28               0.28               0.28               0.27               0.27               0.27
               odor              small             energi             accord          experienc              impli
               0.26               0.26               0.26               0.26               0.26               0.26
                key             adjust           pronounc             acolor               alga            foveral
               0.26               0.26               0.26               0.26               0.26               0.26
          influenti             omusti       beheshtipour            central                clt        flavorflavor
               0.26               0.26               0.26               0.26               0.26               0.26
              franc              gould     khosravidarani              learn             leitch             mobini
               0.26               0.26               0.26               0.26               0.26               0.26
        mortazavian        sohrabvandi             yeoman               mere
               0.26               0.26               0.26               0.26
```

```
$posit
        rate     familiar      present     contrast         meal        famil       iariti        intro
        0.41         0.39         0.39         0.39         0.36         0.36         0.36         0.36
        posi         rule          may      product        ethic multisensori      apprais       pectat
        0.36         0.36         0.35         0.34         0.33         0.32         0.32         0.32
         wtt      particip      context         rich        negat      individu       correl       ground
        0.32         0.31         0.31         0.31         0.29         0.29         0.28         0.28
      expens       actual      assimil       deliza        macfi      perceiv         food        novel
        0.28         0.28         0.28         0.28         0.28         0.27         0.27         0.27
        corr        iness       tiness       dispar      distast        gativ     kehkfnen         kole
        0.27         0.27         0.27         0.27         0.27         0.27         0.27         0.27
     latitud        mojet       sawyer schifferstein       sect   sensorylik         wors          are
        0.27         0.27         0.27         0.27         0.27         0.27         0.27         0.27
     gembaro        provi        these       verain      cracker        shake         hous   disconfirm
        0.27         0.27         0.27         0.27         0.27         0.27         0.26         0.25
        chip         late        grown         basi        decid         duce         lose      hypothet
        0.25         0.25         0.25         0.25         0.25         0.25         0.25         0.25
     unclear      contact          est       exceed
        0.25         0.25         0.25         0.25
```

```
$plant
       hydrolys         success             use         meatlik           muscl          nutrit         structur
           0.39            0.38            0.36            0.36            0.35            0.34            0.34
       composit        hardacr          parker            bean          improv           chand           cooker
           0.33            0.33            0.33            0.32            0.31            0.31            0.31
            eyr gokulakrishnan           malav           ribos         silcock          talukd        tasteless
           0.31            0.31            0.31            0.31            0.31            0.31            0.31
           form            loss         techniqu          select          chiang           freez             impl
           0.30            0.30            0.29            0.29            0.29            0.29            0.29
         makeup    satisfactori        smallscal           struc        vestment         loveday             vari
           0.29            0.29            0.29            0.29            0.29            0.29            0.28
            aim         develop            bone   maillardreact        challeng            date          analogu
           0.28            0.27            0.27            0.27            0.27            0.27            0.27
            ppn       waterbind         product          materi
           0.27            0.27            0.26            0.26
```