Baltic J. Modern Computing, Vol. 12 (2024), No. 1, 50-81
https://doi.org/10.22364/bjmc.2024.12.1.04

Development of A New Conceptual Framework for Better Understanding of the Food Consumer: An Interdisciplinary Big Data Approach

Maija KĀLE

Faculty of Computing, University of Latvia, 19 Raina Blvd., Riga, LV-1586, Latvia

maija.kale@lu.lv

ORCID 0000-0002-6951-9009

Abstract: The author's motivation for this study is to create a new application of big data to food. While the amount of food data increases, many important topics have been studied within separate fields or too narrow disciplines, due to the interdisciplinary nature of this type of data. The author is of the opinion that these drawbacks can be eliminated by analysing food big data using computing methods and by taking interdisciplinary approach to develop research questions. The fact that computer science has a significant role in the analysis of food consumption data is evidenced by the development of a new field of science—food computing, which is the framework of this study. Since there are still many unresolved tasks in the collection and analysis of food-related data, the use of social media and other type of big data is an important introduction. The author, by combining the methodological frameworks of cognitive, social and computer sciences, has developed several methods for using big data, based on the natural language processing and by applying the following methodologies: a) sentiment analysis of affective reactions about food and multi-sensory eating experience; b) comparative analysis of different national cuisines, by applying the topic modelling methodology; and c) bigram analysis to trace ways how food consumers talk about healthy food. Results have been summarised in a new conceptual framework by illustrating its application in practice; in the course of application, several researches, have been conducted using various computer science methodologies. As a combination of theories and methodologies of social, cognitive, and computer sciences, the new conceptual framework offers methods and a direction for studying food consumers. By using computer science methods in the analysis of big data and thus improving the understanding of food consumption, the author hopes to improve the efficiency of public health policies resulting in better public health and higher quality of life.

Keywords: Food Computing, Big Data, Interdisciplinarity, Recipes, Topic Modelling, NLP, Healthy Food, Multisensory Research

1. Terms and Definitions

Affective reactions are the sentiment and emotions related to food. In this paper, affective reactions to food are considered to be important driving factors for choosing food.

Food computing is an interdisciplinary field that includes food-related research in computer science (Min et al., 2019a). To understand the basis of various issues related to food, food computing takes advantage of the possibilities offered by the Internet revolution: social networks, mobile data networks, and the Internet of things, that allows users to share food images, recipes, cooking videos and to make food diaries, thus easily creating rich food data sets (Min et al., 2019a).

Gastrophysics is a non-formal term of a multi-dimensionally integrated food science—it is one of the new research fields that studies the choice, perception, and multi-dimensional flavour experience of food consumers, and the goal of which is to reduce the proportion of unhealthy food ingredients by raising the understanding of consumer choices (Spence, 2017).

Hedonistic experiences are positive affective reactions to eating.

Big data is sets of collected data that are so complex that their processing requires modern technologies, like artificial intelligence. Data can be acquired from various and different sources. Often, this data is homogenous, e. g. GPS communication data from millions of mobile phones that is used to reduce traffic problems; but also heterogeneous data can be used, like medical histories and data on app use. Technologies allow collecting such data very quickly, almost in real-time, and analyse it to gain new perspectives.

Multi-sensory experiences are impressions that build our perceptions of the external world. By focusing on specific events, we gain multi-sensory experiences from how we see, hear, feel, taste and smell (Velasco and Obrist, 2020).

Super-consumer: the choice of food must be evaluated in a context, and the author focuses on food consumers or super-consumers (comparatively wealthy and digitally active consumers) who embody the modern change in people's habits and everyday life (Spence et al., 2016).

Tweet: a microblogging post on the social network Twitter (now X, but referred to as Twitter throughout this paper), with a character count of 140 (until 2017) or 280 (since 2017).

2. Introduction

The choice of food and food consumption plays a great role in the public health. Obesity, type II diabetes, and cardiovascular diseases are only some of the health problems brought by the specifics of modern diet (Min et al., 2019a; Mai and Hoffmann, 2015).

More than 39% of adults aged 18 years and over were overweight in 2016, and 13% were obese. There has been triple increase in the spread of obesity in the world from 1975 (WHO, 2021). The spread of obesity is therefore considered to be a pandemic already (The Lancet Gastroenterology and Hepatology, 2021). For these reasons, it is vital to understand the factors behind food choice and the wider aspects of food consumption.

While the impact of food on individual's health is an issue discussed on a global scale by food policy makers and nutritionists, another discourse has emerged in relation to food consumption, namely, the impact of food on the health of the planet by paying attention to the conservation of biodiversity, generation of methane and carbon dioxide during the production processes, and creation of other types of pollution affecting the global ecosystems and causing climate change (Grivins et al., 2020). A third of the global carbon dioxide emissions is generated by food systems: the biggest share comes from using land for agriculture (around 71% of total emissions, according to estimates), while the food supply chains—transport, consumption, retail, and other related processes—account for 29% (Crippa et al., 2021). Better understanding of eating behaviours can help building policy that meets the needs of both people and the planet (Persson et al., 2021).

Food computing is an interdisciplinary field which, in its wider meaning, includes food-related research undertaken by computer science. The possibilities offered by the Internet revolution, e.g. social networks, mobile networks, the Internet of things, create big data sets as modern technologies allow users to easily share food images, recipes, cooking videos or to make food diaries (Min et al., 2019a). Progress of food computing largely promotes the development of new systems for individual food recommendations. Such food recommendation systems must understand the users and not only their individual behaviours but also the impact their social groups have on their eating choices. Thanks to new, technology-based solutions, this interdisciplinary field is becoming increasingly important in the academic research dedicated to food consumption, public health problems and also the global environmental wellbeing.

The term 'food computing' means using of computer science methods—natural language analysis, computer vision, machine learning, and others—in analysing various food-related data. The main attention is paid to big data and the possibilities to summarise and analyse the data to understand the perception of food and to draw recommendations for forecasting food choices.

Big data comes with many hopes, although there is still a lack of regulation for using it. According to the European Parliament, big data is large amounts of data generated quickly by many different sources (European Parliament, 2021), which indicates an unprecedented scale (amount), fragmentation (rich diversity) and increase in real-time (speed). While the use of big data is being tested academically more often, there are many untapped possibilities for it to create new research fields in food-related research branches. For example, the term 'big data' is rarely used in connection with food safety, mainly because data and information on food safety is disseminated across food, health, and agricultural research fields (Marvin et al., 2017). Thus, interdisciplinary-related challenges can also be seen here.

It is possible to identify several researches where big data is used to study food safety or other corrective aspects connected with food consumption, but there are very few researches studying the positive aspects associated with food consumption (Obrist et al., 2019). In addition, the COVID-19 pandemic, during which this study was prepared, has additionally urged researchers to study changes in eating habits and practices (Laguna et al., 2020). Fewer studies have been conducted using rich data to identify affective reactions to food, including multi-sensory experience and elements important for multi-sensory perception of food (for example, the atmosphere, texture, colour, smell, and others). The author thinks that understanding of the role and impact of the multi-sensory experience in food consumption is critical for improving public health.

In the view of the above, definition of multi-sensory experience is that multi-sensory experiences are impressions of our perceptions of the external world. By focusing on specific events, multi-sensory experiences are gained from how we see, hear, feel, taste and smell (Velasco and Obrist, 2020). At the same time, gastrophysics, which is an informal term of food science integrated with multi-sensory experiences, is one of the modern research fields studying the consumers' food choices, food perception and multi-sensory flavour experience and aiming to reduce the proportion of unhealthy food ingredients by raising the awareness of consumers' behaviours (Spence, 2017). Instead of focusing on the taste, ingredients or nutritional value only, gastrophysics is trying to reveal the impact of 'everything else' that is related to food choice and consumption—be it the eating environment and atmosphere, tableware, names of dishes, cutlery, etc.—as it all forms the experience of taste (Spence, 2017).

The author thinks that understanding of the role and impact of the multi-sensory experience in food consumption is critical for improving public health. The research fields mentioned before (food computing, gastrophysics, and analysis of multi-sensory experiences) serve as the key theoretical basis of the author's new conceptual framework for big data analysis, the aim of which is to understand how people perceive, choose and consume food. Development of a new conceptual framework for the analysis of food perception is an innovative and necessary approach that eliminates the shortcomings of a too narrow research perspective and allows conducting of an in-depth study of this food-related interdisciplinary research field.

Table **1** presents the conceptual framework for the analysis of food perception and consumption, using the big data approach.

In the coming sub-chapters, the new conceptual framework is explained and substantiated by illustrating each of its key elements with an example of results presented in publications by the author and co-authors. The goal of the paper is to illustrate this new conceptual framework, as well as to illustrate the application of this conceptual framework by generating a collection of papers in food computing, based on the new methods. Thus, not only a theoretical framework is offered that serves as a basis for developing research questions, but also the author, using publications by herself and co-authors, illustrates implementation of this conceptual framework in practice, which is related to development of new and innovative scientific papers.

Based on the observation that the amount of data is constantly increasing but there is no methodological approach for the analysis of this data, the author's motivation is to build a new conceptual research framework that would include the diverse approaches used by food computing, cognitive, and social sciences for studying food perception and consumption. While the awareness of the need for an interdisciplinary science is visible on different governance levels (Allmendinger, 2015), it is also recognised at the same time that interdisciplinary research is still problematic. However, computer science is an area capable of promoting the interdisciplinary characteristics, and development of such field as food computing is an example of that.

Table 1: Conceptual framework for the analysis of food perception and consumption, using the big data analysis approach

Collecting information on the context	
Analysis of social media users (awareness of restrictions in the social and economic context).	Specifics of different national cuisines, analysis of recipe data.
Frequency determination and analysis of the timescale	
Analysis of multi-sensory experiences, development of food-related lexicon that serves searches in social networks.	Analysis of the timescale and frequency, which indicates potential new context variables (seasonality, weather conditions, days, weeks,
Identification of emotionally affective reactions	
Analysis of sentiment, which allows drawing conclusions on affective reactions in case of widely discussed topics.	
Adding of additional context variables	
In the context of food analysis: seasonality, weather conditions, trends in the society (e.g.,in relation to planet sustainability).	

Finally, the purpose of this work is to illustrate the usefulness of the data offered by small resource languages like the Latvian language, by using the Latvian Twitter Eater Corpus (LTEC) —a unique resource intended for analysing the food-related tweets in Latvian. This could serve as an example for other small resource languages and promote better understanding of the differences in food descriptions depending on the language used (Fenko et al., 2010).

Based on the research motivation, goals and tasks, this paper answers the following research question: **what methods should be applied, when using big data, to understand food consumers better?**

To answer this research question, the author aims to:

- develop a new interdisciplinary research framework that promotes interdisciplinary understanding and development of methodology, by merging computer, social and cognitive sciences;
- provide an example of the new research framework in action by drafting methodologies and publishing scientific papers that include big-data based analysis of multi-sensory experiences and food perception;
- contribute significantly to the development of food computing as a field by illustrating the research possibilities in analysing big data created within social networks (specifically, Twitter);
- propose further research directions that could be useful for the academic society working in the field of food consumer studies and perception studies;
- contribute to the development of the public health policy by promoting data-based approach in order to advance healthier food consumption by the society and thus higher quality of life of food consumers.

The new conceptual framework consists of four main elements:

1. collection of context knowledge on the topic;
2. detecting of frequency rates;
3. identification of affective reactions and analysis of the timescale;
4. adding of additional context variables.

The new conceptual framework for using big data in the analysis of food consumption can serve as an example for the development of other perception-related studies, for example, by studying the perception of various aspects connected with health. Following the example of this new conceptual framework, analysis of the perception of healthiness in the society can be carried out, as health is a topic in which fake facts, myths and tales of alternative medicine systems are spread widely. By analysing the public awareness of various health issues, it is possible to prepare better policy responses that could be of special value during different pandemics.

The results can also be valuable for those food computing researchers who deal with the development of personalised food and health apps (Rostami et al., 2020). An important contribution is the context knowledge on the society in which these apps are introduced and methodology for using big data in analytics. The relationship between our analogue and digital lives and how our experiences in the digital life (including social networks) affect various choices in the analogue world is studied more and more. These publications can be used as an important data source by those who want to study the links between digital and analogue life (Andersen et al., 2021), as well as other issues related to multimedia in computer science (Yamakata et al., 2022).

3. Theoretical Basis

Since food is an interdisciplinary topic, it is important to understand food consumption, on the interdisciplinary level. Food computing is a research direction dealing with food using the computer science methodology. It is a comparatively new sub-branch of computer science, which has rapidly developed since 2017 and consists of five basic tasks: perception, recognition, retrieval, recommendation, and monitoring of food (Min et al., 2019a). These research tasks arise from the availability of large food-related data sets, starting from images, image recognition, computer vision, analysis of natural language data to development of personalised apps. It is expected that food computing will rapidly develop towards Food Knowledge Graphs, personalised health apps (Kāle and Jain, 2023), which is indicated by the industry developments (see MediPiatto (Vasiloglou et al., 2020) and JoinZoe that, using human microbiome data from PREDICT measurements (The PREDICT program, 2020), build algorithms for individual food recommendation system (Kāle and Jain, 2023) and contextual food choices with weather data (Kāle and Rikters, 2023).

However, taking into account that food choice is often impulsive because it is affected by the surrounding environment and social norms, it can thus be concluded that choosing food is context-based and food-related decisions depend on its availability, the culture, and social and personal habits. Moreover, these factors are unpredictable and changing. Consumers' choice of food is not always rational, often this process is guided by emotions and different accidental context factors (Kāle and Agbozo, 2020). To better understand the rationale behind consumers' food choices, the author first reviews the cognitive and social science literature, followed by an in-depth look at the development of the theoretical framework of food computing.

3.1. Review of cognitive science literature

This paper is largely based on the work of Spence who stresses the need of understanding the role of 'everything else' in the process of food consumption by trying to identify what makes food and beverages enjoyable, stimulating and, most important, memorable (Spence, 2017). Spence (2021) explains the seasonal models of food consumption by stressing that season and climate in general impacts what consumers choose and how is flavour experience formed. Spence (2015b) also recognises the importance of sound by studying what role sounds connected with eating play in our perception and when enjoying multi-sensory flavour experience. Examples for this are specially designed audio sweetener and potato crisps crunching sound effects. Crunchiness indicates freshness of food adding also more pleasant cognitive experience. It must be remembered that flavour starts in our mind and we first eat with our eyes (Spence, 2015a). The background of flavour thus reaches further than a list of well-combined ingredients. Such multi-sensory qualities like shape, sound, smell, as well as context and environment also define the taste of a meal. But the name and description of a meal can have an impact on how we perceive and enjoy the meal (Spence, 2017). This makes us ask questions about the nature of multi-sensory experience and its role in eating.

The author defines multi-sensory experience as impressions that comprise sensory elements, which are built around particular events and which are derived from how we see, hear, feel, taste and smell the world (Velasco and Obrist, 2020). Research of the multi-sensory experience is useful for studying various aspects that characterise the affective and emotional attitudes towards food. In general, multi-sensory experience comes from how we perceive the world that surrounds specific events, in this result we gain impressions with sensory elements (Velasco and Obrist, 2020). Food research is especially favourable for studying the multi-sensory experience as it provides such consumption-related elements (going to restaurants, food presentation, etc.) that are important for tracing the multi-sensory experience and for the development of food computing. The notion 'multi-sensory experiences' includes both the sensory factors characteristic of food (for example, the colour, smell, texture, viscosity) and the external factors related to product packaging, utensils and contextual factors, like music, atmosphere, etc. which affects how we perceive meal (Wang et al., 2019).

Multi-sensory food experience has a crucial role in food perception and figurative, affective communication about it (Kāle et al., 2021). Analysing tweets about food collected for more than ten years, it is possible to not only trace which multi-sensory experiences are communicated by consumers most frequently but also specify the day of the week and the time of the day (morning, midday, evening) when it happens. According to the cognitive science theory, consumers can better express themselves when speaking about colour than, for example, smell, because the everyday lexicon for describing colours is more specific. This knowledge can help us to build successful public health communication campaigns on eating recommendations.

Furthermore, most of the researches until now have been based on corrective food aspects, like reducing weight, studying healthier eating habits, etc. (Obrist et al., 2019). However, it should also be noted that the amount of textual and visual information on food grows day by day and data on food, which includes blogs by famous people, TV cooking shows and other entertaining food-related events, play an increasingly important role on digital media (Spence et al., 2016). Another important aspect is the transfer of hedonistic experience to new media, and it becomes more important also in the communication of food suppliers and consumers which is intensifying both in the digital and analogue environment (Andersen et al., 2021). Thus, one of the notions used to identify the positive food-related experience is the idea of complexity, which is understood as the versatility and richness of positive experiences connected with eating and, despite the vague definition, it allows to distinguish the food consumption experience from the simplified and rational perspective of choosing food (Kāle and Agbozo, 2020a). This theoretical approach is innovative as it deals with large scale analysis of social network data related to food and allows to track food perception, which describes hedonistic experience. In this case, complexity is analysed using bigrams and by studying nouns and adjectives that characterise a particular food product. This approach is explained in more detail in the chapters devoted to methodology and result analysis. But one of the main conclusions is that healthy food descriptions lack hedonistic sentiments if compared to comfort foods, like chocolate and wine (Kāle and Agbozo; 2020a; Kāle and Agbozo 2020b). While wine is referred to as ‘fleshy’, ‘muscular’ and ‘sinewy’ but dessert as ‘sexy’ and ‘seductive’ (Jurafsky, 2015), healthy food is described by its pragmatic characteristics, for example, ‘easy to cook’ and ‘healthy’. Healthy food lacks the eloquent descriptions of comfort foods and, therefore, associations that healthy food can also be tasty do not evoke. And this means that when looking for a pleasure people will rather choose food that is already described figuratively and in a way that stirs our imagination not with pragmatic lines emphasising the rational aspects of it.

Division between healthy and tasty food in our minds (Mai and Hoffmann, 2015), therefore, is one of the central topics of this paper, because understanding of tastiness and healthiness is vital to comprehend the factors underlying the choice of food. This is also one of the central opinions in the field of food computing. A large portion of food-choice related apps offer different recipes according to individual’s flavour preferences. Studies of the interchangeability of recipe ingredients according to flavour preferences are based on the same food selection principles. So, the aspect of ‘tastiness’ is important in relation to individual’s choice to cook or eat some specific meal, and knowledge that ‘tastiness’ conflicts with ‘healthiness’ is also important for the analysis of big data in relation to these aspects and for taking into account their mutually exclusive nature. However, any individual making choices in the modern urban and digital world interacts with different social groups, and these groups have a significant impact on individual choices; therefore, the next chapter examines the theoretical underpinnings from a social science perspective.

3.2. Review of social science literature

In her paper on flavour and social class, Margo Finn studies how flavour preferences form within social classes by analysing how it is manifested in the language used for describing the food perception. By analysing the social network Twitter, Abbar et. al. (2015) identified various differences in eating experiences as they are described in urban and rural settings. Interesting, that "alcoholic beverages tend to be mentioned in urban environments, whereas pizza and chocolate, are popular in the rural ones" (Abbar et al., 2015), which, in a way, signalises life-style and status quo issues in different social groups. While inequality and the related eating differences is not one of the research questions of this paper, it is important to be aware of it when analysing large scale social network data in order to understand the research limits.

It should be noted that the author analyses food consumer that can be called a super-consumer, who can be characterised as a comparatively wealthy and digitally active consumer that embodies the change in people's habits and everyday life (Spence et al., 2016). The author is aware that a large portion of the global population cannot be regarded as super-consumers due to the limited access to food. Society is becoming more and more polarised, and next to super-consumers there is a spreading poverty and, paradoxically, those working in the food sector often receive the smallest wages and are financially and socially most vulnerable (Grivins et al., 2020). It is important for computer science researchers working with the analysis of social network posts to gain an in-depth understanding of those who are active in social network discussions, since it is critical to know that data that is generated from posting and uploading photographs on social networks mainly describes comparatively wealthy social groups.

Such wider account of social processes is necessary to understand other food-related critical aspects, like food supply and logistics, which has spread over the last decades thanks to globalisation and faces such problems like long supply chains and increase of the consumption of convenience foods having negative impact on the health, as indicated in the strategy From Farm to Fork (Council of European Union, n.d.). If we look at the history of food supply chains, we see that the history of food packaging follows the history of war. Similarly to military leaders who have always tried to synchronise army's movement with logistical roads, there has always been a search for methods for keeping the delivered food fresh. Canning was the first modern military-inspired packaging technology developed further as the industrial revolution at the end of the 18th century began (Metcalfe, 2019). It explains why food products with long expiry dates is a priority for food supply chains nowadays. Unfortunately, it undermines the health of food consumers, therefore more and more emphasis is laid on the development of short food supply chains, at least within the European Union.

Understanding the context of eating allows the author to discuss the limitations of her research work's objectivity and to establish the level of permissible interpretation and generalisation of the research results. Thus, important social science issues for studying the eating trends of an individual and society include the wellbeing, differences between urban and rural environments, aspects related to culture and food history, different social roles in family, and many other topics that have been studied in different social science researches in connection with food choice and eating (Finn, 2019; Metcalfe, 2019; Burnett and Ray, 2012).

3.3. Review of computer science literature

The food sector as such is characterised by many non-efficiencies be it choosing of healthy food or food waste management, for example food waste in the developed countries accounts for up to 40% and majority of this amount is due to household food purchasing habits (Lopez Barrera and Hertel, 2021). Many hopes of solving the inefficiencies of food systems are largely associated with the development of computer science: new technological solutions could be used to address various risks in both agriculture and food supply chains and also the choices made by consumers. At the same time, rapid development can be observed currently in precision agriculture (Carolan, 2015) and as various personalised food app solutions (Nag et al., 2021). And, while the expectations are high, the information and communication technology sector still face significant challenges in relation to developing sustainable solutions for food systems (Svenfelt and Zapico, 2016). The author thinks that it is the interdisciplinary nature what poses challenges and this paper can thus be regarded as an example of how sustainability and health issues can be addressed and how the methodology used in computer science can be used to acquire new valuable information.

Over the last years, the interest about food data analysis is increasing in computer science, as several computer science fields are developing at the same time. One of them is human-oriented computing, which is central for the International Conference on Human–Computer Interaction of the Association for Computer Machinery (ACM Digital Library, n.d.), which devotes an increasing number of workshops (Ferran et al., 2019) for studying the food and human interaction and generates new approaches (UC Santa Cruz, n.d.) and a vision how food could be analysed by the computer science and what the future of food computing might be (Obrist et al., 2018). Similar developments have been observed in the ACM Multimedia: in 2022, the international workshop on eating, cooking and food apps was organised during the conference (CEA++, 2022). Taking the direction of food analysis by the important ACM conferences is natural, given the challenges faced by people and the environment in securing the operation of global food systems.

In essence, food computing is an interdisciplinary field that studies food-related issues using computer science. Food computing survey, which is a comprehensive publication by the Association for Computing Machinery, gives an overview of the available data sets, methods of data analysis and illustrates the possibilities created by the increasing amount of big data in food and computing (Min et al., 2019a). In general, data sets are based either on image or natural language data sets. Since this work focuses on the data sets based in natural language, the author will further present the possibilities offered by these data sets for a better understanding of food consumption.

Initially, the recipe data set was the most available and biggest data set. One of the most significant works in this field was performed by Ahn and co-authors in 2011 when they studied the principles of flavour networks and food pairing revealing that "the flavor compound (chemical) profile of the culinary ingredients is a natural starting point for a systematic search for principles that might underlie our choice of acceptable ingredient combinations" (Ahn et al., 2011). The work of Kāle and Agbozo (2021) is largely based on the approach by Ahn and co-authors stating that North American and Western European recipes tend to combine ingredients with similar flavour compounds, while East Asian recipes include ingredients with distinct flavour compounds (Jurafsky, 2015). To analyse the ingredient combinations in different national cuisines, this work applied the topic modelling methodology (Kāle and Agbozo, 2021). All the researches mentioned here analyse food ingredients from the information available in recipes. While this approach provides important information on the ingredients characteristic of particular countries and regions, this data does not allow to identify the result of enjoying the meal—factors related to flavour and other aspects related to eating (Ahn et al., 2011).

Although recipes are one of the simplest large-scale data sources available for the analysis of food-related texts and have thus served as one of the key elements in food computing, there are several limitations to the analysis of this data set. Food computing studies mainly analyse recipes to compare different cuisines, national dishes and ingredients, to identify characteristics of flavour. The key assumption is that flavour matters and that taste depends on the combination of ingredients (Ahn et al., 2011). Most of the recipe data analysts base their work on the assumption that recipes contain important information about the meal and that this information concerns flavour, nutritional and health aspects (Kusmierczyk and Nørvågg, 2016; Ahn et al., 2011; Min et al., 2019B; Herruzo et al., 2016). This assumption seems to be an obvious truth, however, an increasing number of studies in cognitive sciences reveals that 'everything else' pertaining to the food matters as well (Spence, 2017). In fact, the flavour or taste of a product might be subordinate to other more determining factors at play – such as context, social class and even 'food fashion'. Multisensory food research is on the rise (Obrist et al., 2019), which shows that the explanation of palatability based on a sole factor – such as flavour – is insufficient in terms of the human relationship with food.

Taking this into account, we may conclude that while recipes can be utilized, up to a certain degree, as the indicators of a food's flavour, doing so without a holistic approach that includes cognitive and social sciences as well as computing might turn out of a marginal benefit for the research community. Working with large-scale flavour-related data has proven to be complex, and most of the analytics has been done in a comparative framework looking at various cuisines in different geographical areas, since the data on flavour analysis per se does not provide any substantial clues for further understanding of the human and food relationship. A comparative geographical analysis proves, however, that there are differences between various cuisines when it comes to flavour and that there is a particular divide between the East and the West (Jurafsky, 2015). How to proceed with this knowledge in tow remains an open question though.

3.4. Analysis of natural language data: computer linguistics

Recipes as one of the most readily available big data sources for early analysis of food-related texts are considered to be one of the basic elements of food computing, which performs analysis of food consumption using big data sets. Researchers have mainly focused on comparing different cuisines, national dishes and ingredients to establish flavour, for example by identifying authentic, frequently used ingredients (Ahn et al., 2011; Min et al., 2019b; Kim and Chung, 2016). In most of the cases, studying of recipes has centred around combination of ingredients and their different representation in various national cuisines. Large share of such researches has been inspired by the availability of large-scale digital data and the possibilities to use them (Kāle and Agbozo, 2021). The author has analysed the usefulness of recipe data by developing a new conceptual framework in order to offer new research questions and to raise awareness of the big data related to food.

Meanwhile, analysis of social network data has become more frequent in different consumer researches: in case of Twitter, the analysis includes both tweets and hashtags (Puerta et al., 2020). As the COVID-19 pandemic began, the amount of social network text data increased. Changes to food consumption and habits have been analysed in many countries by identifying changes in microblogging posts or tweets (Laguna et al., 2020). Analysis of social network data provides information on consumer experience of choosing or consuming food and allows tracing the multi-sensory food experience. It should be noted that analysis of social networks gives information on how users group their taste qualities and on interlinked structural bonds. The author is of the opinion that better understanding of language use—in relation to both healthy and unhealthy food—can bring us to new methods of how to reduce contradictions suggesting that healthy food is less tasty than the unhealthy.

It is worth recalling that one of the central themes of this paper is the intuition of "unhealthy = tasty" (Mai et al., 2011), which creates cognitive dissonance in food consumers, who immediately associate healthy food with something bland and describe unhealthy food rather eloquently and sensually (Jurafsky, 2015). Results such as these - where healthy food is associated with comparatively pragmatic and rational benefits, rather than palatability and eating pleasure - can be observed in the analysis of social networks, particularly Twitter (Kāle and Agbozo, 2020a; Kāle and Agbozo, 2020b) more described in the next chapters. Although consumer studies are more and more trying to understand and interpret tweet texts (Puerta et al., 2020), until now no one has ever tried to analyse multi-sensory experiences in the way this paper does. More detailed description of the innovative approach is provided in the next chapter, which collects the publications in cognitive science.

4. Research Design and Methods

The author has analysed Twitter posts in which consumers have shared their experiences related to different foods and beverages. Twitter is mainly a platform for text (not images), and it is widely used to analyse sharing food-related experience. Twitter was the most popular social network from 2011 to 2015, but afterwards many users migrated to other social networks (Kāle et al., 2021). In comparison with other data sources, social networks allow tracing spontaneous reactions when people tweet immediately, thus avoiding possible deviations that might emerge if other methods for gathering opinions were used, like surveys (Puerta et al., 2020). For the author, “microblog posts on Twitter” are microblog posts in the length of one tweet. For the purposes of research, several Twitter data corpora, including the recently published data set the Latvian Twitter Eater Corpus or the LTEC (Sproģis and Rikters, 2020), were used. Data had been collected for ten years by using 363 eating-related key words. The total size of the data set exceeds 2.3 million tweets generated by 168,703 users.

The methods described below were used for the analysis of social networks and other large-scale data bases. Methods are grouped by their purpose: 1) analysis of healthy-food related hedonistic connotations on Twitter; 2) detecting of frequency rates of food-related multi-sensory aspects on Twitter; 3) identification of the multi-sensory experiences related sentiments on Twitter; 4) comparison of different national cuisines using the recipe data analysis, and 5) correlating LTEC food tweet data with weather data.

4.1. Analysis of healthy-food related hedonistic connotations on Twitter

One of the assumptions tested with data analysis methods is that there are few hedonistic connotations in the discourse on healthy food. Namely, healthy food is discussed as rational and pragmatic choice, but it is not associated with tastiness and enjoyment. To reveal hedonistic connotations that are related to healthy food on the social media network Twitter, data mining methods were applied as “a process of analyzing data to capture key concepts and themes and uncover hidden relationships and trends without prior knowledge of the precise words or terms that authors have used to express those concepts” (Khatai et al., 2019). Using Twitter API (application programming interface) in the R programming language, information in English was searched on the following topics: #healthyfood and #recipeoftheday. To obtain meaningful contextual information from the data set, N symbols were generated (Gargouri et al., 2003). The final method applied to language data analysis was establishing word associations to identify such associations like 1) #healthyfood and the corresponding supportive terms #sustainable, #nutritive and #tasty; 2) #vegan un #vegetarian, supported by the terms #lowfat, #healthy and #tasty (Kāle and Agbozo, 2020). Detailed results are presented in the chapter devoted to research results.

4.2. Detection of the frequency rates of food-related multi-sensory aspects on Twitter

To detect the frequency rates of food-related multi-sensory aspects on the social network Twitter, a large, domain-specific tweet data set (with semi-automatic annotations of the mentioned food products and beverages for each tweet) was gathered, normalised and filtered to obtain only those units that are directly related to the research question. Workflow was created based on the following tasks: 1) to identify the most common adjectives and verbs used to describe characteristics of different meals (and drinks), and to automatically include all the possible language modifications; 2) to select those tweets that contain at least one of the specified adjectives or verbs; 3) to select only those tweets that mention also at least one type of food, dish or beverage; 4) to illustrate the number, timescale and other aspects of the most popular combinations visually. Following these tasks, the results were analysed and interpreted to identify the key trends that are based on the frequency of multi-sensory aspects (Kāle et al., 2021).

4.3. Identification of the multi-sensory experiences related sentiments on Twitter

To identify the sentiments related to aspects of multi-sensory food experiences on the social media network Twitter, the author employed two different methodological approaches. The first was applied to the analysis of tweets in English, but the other to the food-related tweets in Latvian. In the study by Kāle and Agbozo (2020b), text mining was applied as the technique for discovering unknown data models by using the latent semantic analysis (LSA), to obtain and depict the meaning of words in context using statistical calculations that are usually applied to big language data sets. Tweets with the keywords ‘kale’ and ‘food’ were marked additionally, using the UDPipe model, to identify the adjectives necessary for analysis (Straka et al., 2016). Then, bigrams containing adjectives were extracted from the data set, and sentiment analysis was performed of adjective bigrams from a clean, marked data set. To evaluate the general emotions expressed by a tweet, the Loughran-McDonald Sentiment Word Lists were used, which was originally built for sentiment analysis in financial analytics (Loughran and McDonald, 2011) but then was adapted for the analysis of food data. The second semantic lexicon was AFFIN (now ANEW) (Nielsen, 2011), which attributes a number of polarity points to each word and is thus created by words that are classified as reflecting positive or negative sentiments.

A different approach was used by Kāle and Rikters (2021), where the sentiment analysis tool was adapted and used for general tweets in Latvian (Thakkar and Pinnis, 2020). At the same time, also the model characteristic of the Latvian language (Vīksna and Skadiņa, 2020) was used by applying it to the whole Latvian Twitter Eater Corpus (LTEC). Namely, in order to identify the unique tweets containing at least one of the pre-defined adjectives or verbs using a morphology tool for Latvian, the LTEC was used to identify the aspects related to smell, taste, and temperature (Deksne, 2013). It allowed to teach the sentiment classifier to distinguish between negative, positive and neutral tweets (Kāle and Rikters, 2021).

4.4. Comparison of different national cuisines by using the recipe data analysis

To compare different national cuisines using the analysis of recipe data, the topic modelling method was applied. It was considered to be the most appropriate of methods for recipe data analysis as it allows researchers to determine the prevalence of names of different meal ingredients within the selected topics. And it can bring to an in-depth analysis of more specific topics (Nikolenko et al., 2017). In the work, the topic modelling principle called the latent Dirichlet allocation (LDA) was applied, which is a widely used general probability model, despite its inability to model topic correlation. As explained in an article by Kāle and Agbozo (2021), the initial data set obtained from the data set Yummly-66 k, which consists of 66,615 recipes from ten cuisines on the site Yummly (a website that generates personalised recipes and serves as a search engine for recipes), is a repository of cross-regional recipes. This data set was considered to be suited for the research that focused on North American and Mexican cuisines. The topic model function was generated based on the LDA model to acquire ten (10) topic models from the corpus. After cleaning and filtering of the data set using the part-of-speech (POS) tagging model, (udpipe developed by Straka et al. (2016)), the udpipe model was used to tag POS in relation to all tokens. This allowed to establish the adjectives used most frequently. Finally, the test data set and training data sets were analysed to obtain complete overview of the results (Kāle and Agbozo, 2021).

4.5. Correlating food and weather data

Using a combination of two data sources - the LTEC for tweets and weather data exported from Meteostat, Kāle and Rikters (2023) mainly focused on tweets and weather relating to Riga, the capital of Latvia, since most tweets with location data originated there, and it was difficult to obtain detailed historical weather data for the smaller regions. From the Meteostat website, it was possible to reliably obtain only data for temperature and precipitation, while data for snowfall was only available up to the end of 2017, and data for wind speed and air pressure was only available from July 2018 and onward. Figure 1 shows a visual depiction of the data gathered. There was no available data to trace daily sunshine directly, but it can be inferred from looking at precipitation, snowfall and air pressure (Kāle and Rikters, 2023).

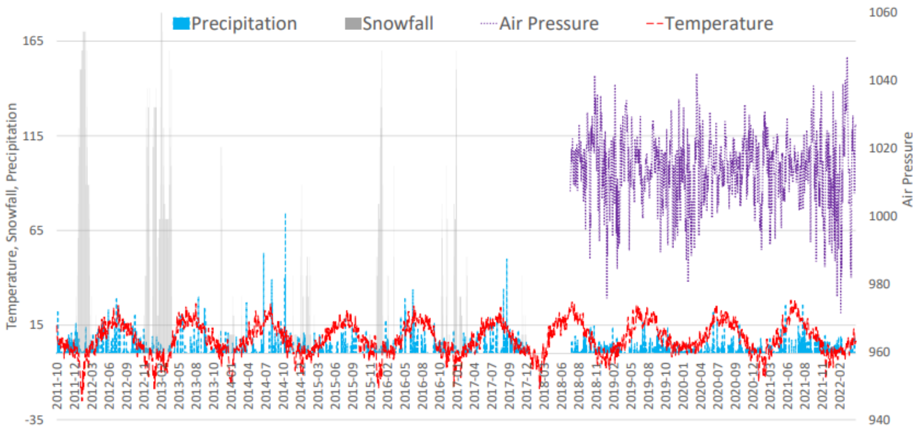

Figure 1: Visualisation of available weather data from Meteostat (Kāle and Rikters, 2023).

5. Conceptual Framework

In the following sub-chapters, the author explains and substantiates the new conceptual framework (see

Table **1**) by illustrating each of its elements with an example of the results acquired in publications by the author and co-authors. The current results are presented together with a review of topic analysis possible in the future, thus contributing for further research development. The conceptual framework is studied step by step.

5.1. Gaining context knowledge in the chosen topic

The initial goal by comparing various cuisines using recipe data was to classify dishes into healthy and unhealthy food products. This goal was not achieved because healthiness of food is contextual and depends on the overall diet and lifestyle (Kāle and Agbozo, 2021). However, it was possible to build topic models for each of the analysed national cuisines thus gaining an overview of their meals, their ingredients, and specifics. While large-scale recipe data can give food computing researchers certain added value, the assumptions used to analyse food consumption in different cultures should be developed constantly, otherwise analysis of the obtained data will not result in sufficient added value (Kāle and Agbozo, 2021).

In-depth context knowledge draws significant limits that have to be taken into consideration when generalising the obtained data. By focusing on the wealthy and digitally active part of the society, it was important to analyse the trends that build discussion topics by including, for example the discourse on vegan and vegetarian food in the context of climate change, the issues related to human health and wellbeing, the peculiarities of national cuisines, and others. Without the contextual understanding on food systems (Grivins et al., 2020), there would not be sufficient substantiation for the chosen focus and an in-depth analysis of food-consumption related aspects.

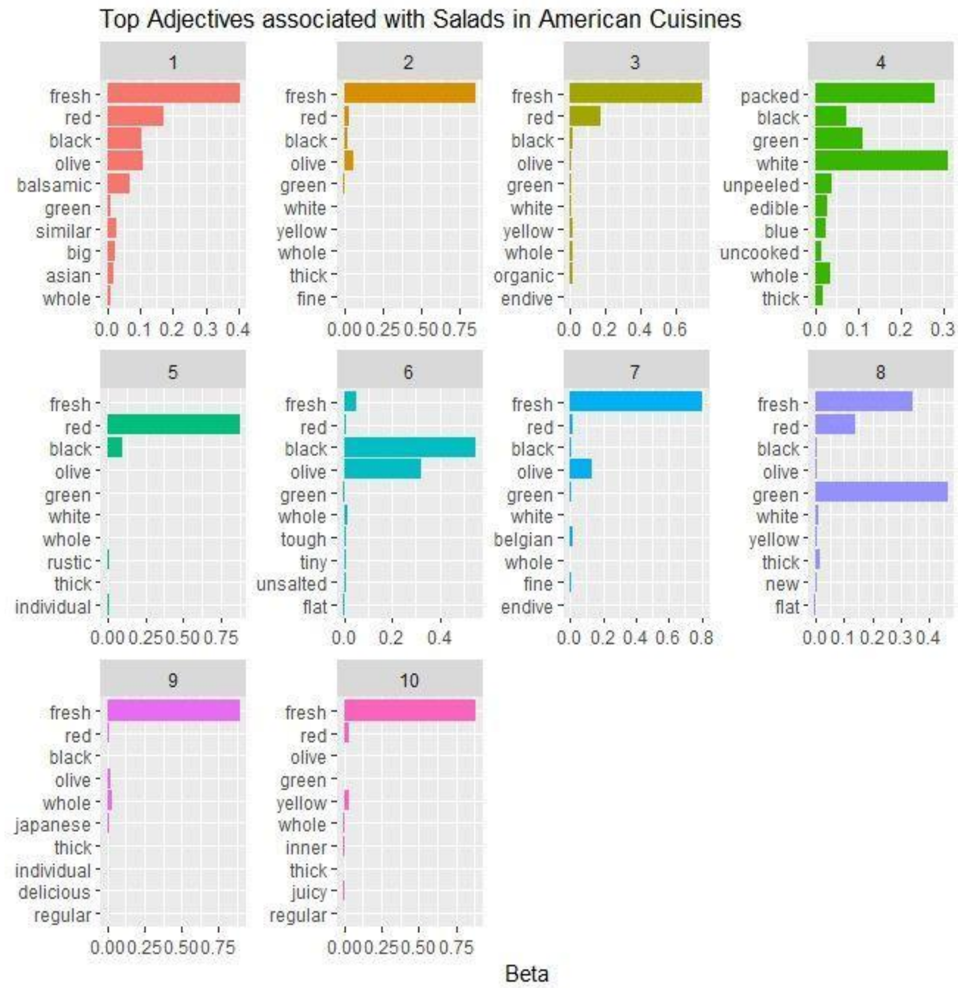

Figure 2: Topic models for adjectives (Kāle and Agbozo, 2021)

This example from the author’s publication demonstrates how context knowledge is acquired and what are the most often used adjectives and nouns in different cuisines. To illustrate, here is an example of adjectives describing salads and adjectives describing desserts in the North American cuisines (see Figure 2 and Figure 3). More detailed knowledge about what are the adjectives and nouns used most frequently in a particular cuisine can help researchers to clarify search queries in social media researches, in which selecting the right keywords is the first important step for the research to be successful.

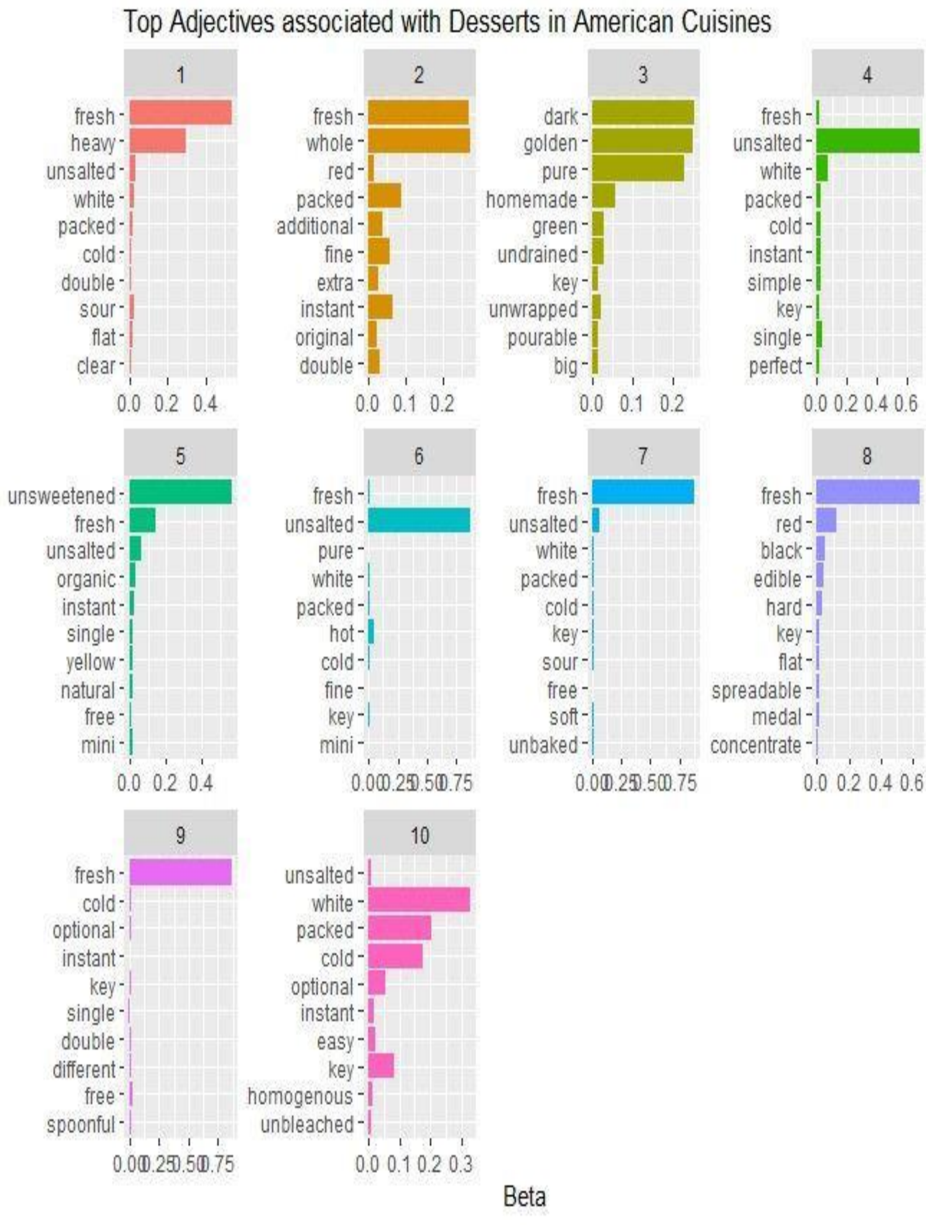

Figure 3: Topic models for adjectives (Kāle. and Agbozo, 2021)

This example from the author's publication demonstrates how context knowledge is acquired and what are the most often used adjectives and nouns in different cuisines. To illustrate, here is an example of adjectives describing salads and adjectives describing desserts in the North American cuisines (see Figure 2 and Figure 3). More detailed knowledge about what are the adjectives and nouns used most frequently in a particular cuisine can help researchers to clarify search queries in social media researches, in which selecting the right keywords is the first important step for the research to be successful.

This specific information concerning a particular national cuisine can also be of value in comparative studies to identify differences in the key ingredients in recipes that build the flavour basis of a certain national cuisine. The topic modelling method, when an overview of the ingredient combinations used most frequently is obtained by analysing different foods (for example, salads, desserts, etc.), allows researchers to obtain the ‘meta-recipe’ of that particular cuisine and to build a lexicon specific of the food products and ingredients of that cuisine, which can serve for a further in-depth analysis. Thus, the topic modelling method applied to big data of recipes and their ingredients provides important raw data that can be then used to analyse, for example, the data of social networks. Recipes and ingredients present the food products that are actively used in a certain region, and they can offer valuable context information for the analysis of the attitudes expressed in social networks towards, for example, alternative proteins, meat consumption or other food-related issues topical in that particular society.

5.2. Detection of the frequency rates

In this paper, special attention is paid to the distinction between ‘healthy’ and ‘tasty’ food, since these, as already stated in the literature review, are mutually exclusive categories because of the ‘unhealthy = tasty intuition’ (Mai et al., 2011), which works in both directions—healthy food is automatically perceived as less tasty while unhealthy food is described as tastier. Therefore, assessment of the frequency of the bigrams associated with this topic in the affective reactions about healthy food on Twitter is one of the key research tasks chosen by the author for this chapter. The more frequently certain bigrams repeat, the more important is the discourse upheld by these bigrams. When analysing big data, detecting the frequency rate is one of the first analytical approaches to take for further classification and processing.

The results show that the expression ‘healthy food’ is usually accompanied by words that denote pragmatic decisions to use healthy food, but words that suggest of affective reactions related to food are used rarely (Kāle and Agbozo, 2020a; Kāle and Agbozo, 2020b). We can therefore conclude that healthy food is not described as tasty and enjoyable, but the reason behind choosing healthy food is rational and pragmatic care of one’s health. In this way, the existing tweets and posts on Twitter that are related to healthy food do not concentrate on the aspects of flavour and have less references to hedonistic expressions, they focus on ‘simple and easy’ cooking instead of ‘complex and enjoyable’ eating experience (Kāle and Agbozo, 2020a).

The connection between the Twitter language discussed here and healthy food is also related to the discussion on the extent quantitative analysis methods of language data / texts are compatible with research questions in cognitive science. Instead of directing studies towards the analysis of food ingredients alone, as it has been done in computer science until now (Ahn et al., 2011), the focus should be widened to include, for example food associations, affective reactions to food, links between choosing food and the social status, and other important context factors. By expanding the range of research questions, we will be able to better understand the phenomenon of food choice as such (Kāle and Agbozo, 2020b).

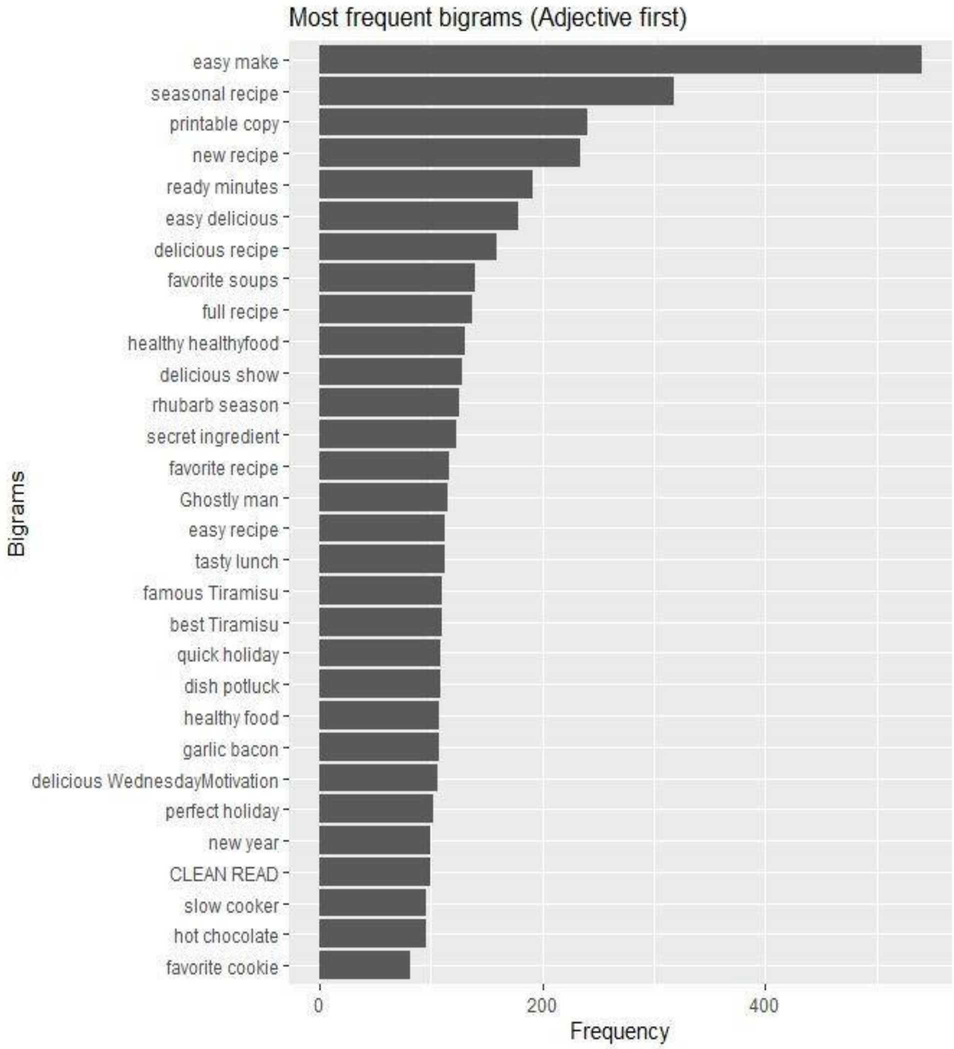

Figure 4: Frequency of bigrams for healthy food on Twitter (Kāle and Agbozo, 2020)

As presented in Figure 4, healthy food is mainly characterised by a set of bigrams describing food as simple, seasonal and easy to make and underlining the rational and pragmatic choice and pointing out benefits for health. No aspects of flavour and enjoying eating are emphasised. This allows to draw conclusions on how topics on healthy food and tasty food form on social media and provides the necessary elements for understanding how the manner we refer to food can affect our food choices. Results that emphasise specific health-related characteristics but do not contain descriptions of flavour are also presented by the analysis of particular healthy food product bigrams (see

Figure 5). Interestingly, that a recent study of superfood representations in news media (Ghandi et al., 2023) captures meaningful patterns that are consistent with discourse findings about known superfoods using the kale dataset (Kāle and Agbozo, 2020b).

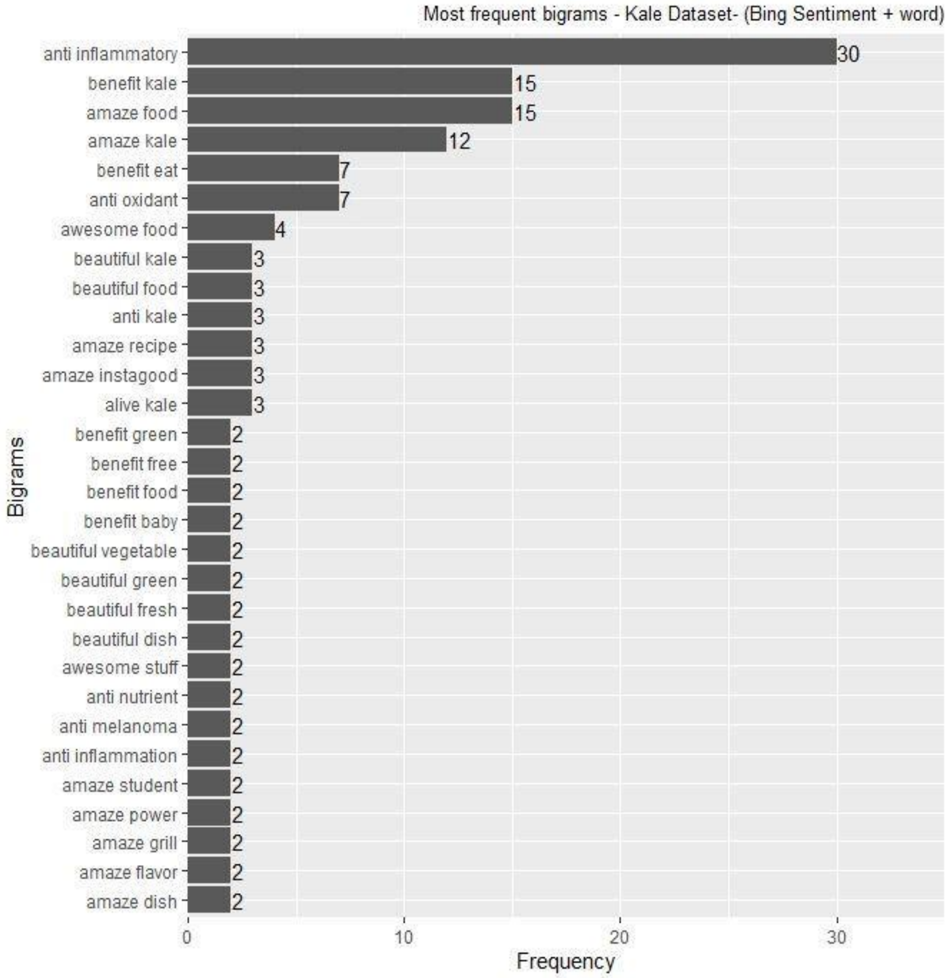

Figure 5: Frequency of bigrams for kale (Kāle and Agbozo, 2020b)

5.3. Identification of emotionally affective reactions and the timescale

Thorough identification of emotionally affective reactions and of the timescale of these emotions expressed by the Twitter users in Latvian was performed by using the Latvian Twitter Eater Corpus: the emotionally affective reactions at different times of day and week were assessed by looking at the depiction of the multi-sensory experiences. To establish affective reactions, sentiment analysis was performed (see Kāle et al., 2021).

According to results, some multi-sensory experiences are mentioned frequently (for example, those related to colour), while others—rarely (for example, the ones connected with texture and smell). Placing the affective reactions on the timescale allowed us to make conclusions on time-related dynamics, for example, to recognise that coffee and chocolate is more often perceived as ‘tasty’ closer to weekend, but Monday mornings usually start with healthy breakfast, like salad. From this perspective, Kāle et al. (2021) have added a unique, time-sensitive point of view to the research of multi-sensory eating experience, and this new approach provides important information on how and when social network users talk about multi-sensory experiences of eating on the social network Twitter. This data is significant for the analysis of the increasing impact daily rhythm has on food consumption (Boege et al., 2021). By taking into account the individual food preferences according to the circadian rhythm, it is possible to develop even more precise personalised food apps by taking note of also the group dynamics on social networks regarding the cycle of food consumption and the impact of this cycle on individual food choices.

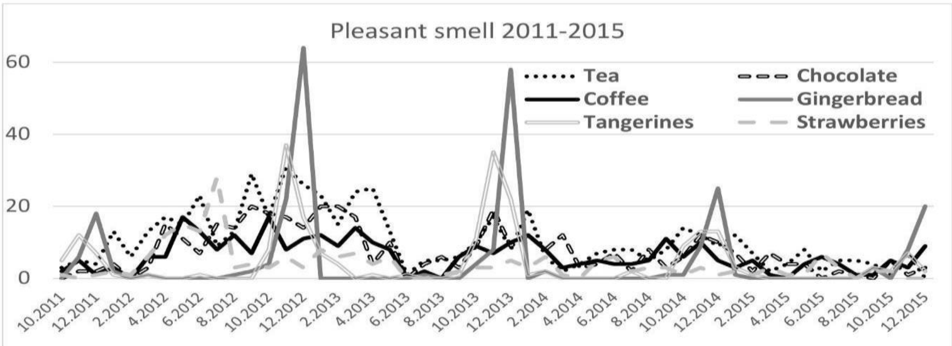

Figure 6: Sentiment graph for food products and drinks associated with pleasant smell, 2011–2015 (Kāle et al., 2021)

This example illustrates the previously mentioned research by Kāle et al. (2021), namely, how specific foods and drinks are described in the context of their multi-sensory characteristics, in this example—with affective reactions to pleasant smell (see Figure 6). The peaks signal increase in the frequency of food consumption depending on the seasonality. The decreasing height of peaks can be explained with Twitter users becoming less active and with their numbers dropping due to migration to Instagram and other social networks in 2015. By analysing the frequency rate of the representation of food and multi-sensory experiences, it was possible to spot seasonality and some cultural aspects that are related to traditions in a particular national cuisine, in this case—the Latvian. The analysis has also revealed a relation between food-related tweets and particular days of the week and times of the day, thus presenting links between tweets and particular food products during particular periods of the day (Kāle et al., 2021) and even verbs denoting sounds associated with food consumption (see Figure 7).

Knowledge and higher level of precision with regard to specific food products, which is reflected by the seasonal peaks, can help public health policy makers build such communication strategies that resonate with the society on social media. But the seasonality and cyclic nature of the consumption of different food products plays an increasing role in computer science when developing personalised food recommendation apps and building recommendation systems (Rosa et al., 2022).

	Munch	Crunch	Crackle	Chomp	Sizzle
Crisps	5	15	0	0	0
Chocolate	0	11	4	0	0
Meat	2	7	3	0	3
Potatoes	0	11	0	0	2
Apple	4	8	0	0	0
Gingerbread	1	8	3	0	0
Cream soup	0	11	0	0	0
Bacon	0	11	0	0	0
Tea	3	4	3	0	0
Salad	1	7	0	1	0
Sauce	0	9	0	0	0
Pork	3	5	0	0	1
Ice cream	0	4	2	0	2
Soup	0	8	0	0	0
Cheese	0	8	0	0	0
Chicken	2	2	1	0	2
Water	1	4	1	0	1
Pumpkin	0	6	0	1	0
Pancake	3	2	0	0	0
Strawberries	0	2	2	0	1

Figure 7: Top 20 foods and drinks most often described using top 5 verbs denoting sounds (Kāle et al., 2021)

5.4. Adding of additional context variables

The last step in the new conceptual framework for understanding food consumers is adding of new context variables identified in the previous analysis. The author adds two new context variables that emerged from the last analysis: weather conditions and perception of healthiness. These variables are further described below.

5.5. Weather condition data

Digital social networks are valuable because they are perceptive of trends and offer new ways how to trace seasonal differences in flavour and food products preferred by the society, which is especially noticeable in countries with more explicit differences between seasons in terms of temperature (Spence, 2021). Latvia with four seasons is a good example for analysing tweet information during the cold winter season and the increasingly warmer summer days (Kāle et al., 2021).

By linking the food-related data from Twitter with data on weather conditions, it could be possible to assess the connection between food choice and the language describing food, on one hand, and the weather conditions and the public assessment / perception, on the other hand (Bakhshi et al., 2014; Bujisic et al., 2019).

This knowledge on the frequency of tweets and the sentiment can be useful for makers of the public health policy, and they can be used to urge consumers to choose healthier food alternatives according to different weather conditions and seasons. Recognition of the impact weather conditions have on consumers and taking note and understanding of their affective reactions help explaining also several non-efficiencies related to food consumption: food wasting, choice of healthy or unhealthy food, and other issues.

The contextual knowledge created here can be of value for computer science researchers who work in the development of personalised food and health apps, because people are social beings and the behaviour of others affects their choice. Linking the data of weather conditions with food choices could be useful in creating a global food atlas—an open source platform containing information on the availability of food that takes into account different factors to illustrate how food and lifestyle navigation systems can use the global food atlas to provide personalised and context-oriented solutions (Rostami et al., 2022).

Moreover, this research can help to show the level of link between our digital and analogue lives: to trace tweet sentiments and frequency rates by using actual purchasing habits and food sale data.

In order to study the impact of weather conditions on food choices, detailed weather data is necessary for a statistically significant analysis of the links between weather and sentiments in food tweets.

While the results of tweet sentiment in terms of the percentage of negative, neutral and positive tweets are largely the same for all weather conditions, it was still possible to observe significantly fewer positive tweets during windy and high pressure weather conditions, as can be seen in Table 2.

Table 2: Weather relation to tweet sentiment ('-' negative sentiment; '0' neutral sentmient; '+' positive sentiment).

	Cold	Warm	Windy	Snowy	Rainy	High Pres	Low Pres	Overall
-	12.59%	13.20%	23.15%	11.88%	13.63%	23.10%	12.63%	13.07%
0	37.25%	38.68%	**48.40%**	36.06%	38.64%	**48.26%**	38.72%	38.38%
+	**50.17%**	**48.12%**	28.45%	**52.06%**	**47.73%**	28.63%	**48.65%**	**48.55%**

5.6. Identification of the perception of healthiness using computer linguistics methods

One of the further important aspects in the analysis of food-related big data is the perception of healthiness. Categorising of different types of food into ‘healthy’ and ‘unhealthy’ is challenging. Healthiness as a notion depends on various aspects, that are much wider than only food-related factors. The general observation is that “the objectives of health and taste often conflict—and taste usually prevails in food decision making” (Mai et al., 2011). Words that characterise food have decisive role in positioning food on the mind map, and, when describing food as healthy, the impact of the idea of taste according to the intuition ‘unhealthy = tasty’ reduces (Mai and Hoffmann, 2015). An intriguing direction for further research could therefore be connected with healthy food or what is referred to as healthy by people in the context of their health or environmental sustainability (Kāle et al., 2021).

Food choice and consumption greatly affect the public health and also the sustainability of the environment. Although food impact on individual’s health is a topic addressed by food policy makers and nutritionists all over the world, a new discourse has emerged in relation to food consumption: its impact on the planet’s health and the level of biodiversity, pollution and CO_2 emissions that affect and cause climate change and the planet’s ecosystem as a whole (Grivins et al., 2020). A third of global carbon dioxide emissions is generated by food systems, in which the largest contribution comes from the land use for agriculture (about 71% of the total emissions) but the food supply chain (transport, consumption, retail, and other processes) accounts for the remaining 29% (Crippa et al., 2021).

The new food policies not only include the impact on individual’s health but also look at the influence of food consumption on the environment and global sustainability. In 2019, the EAT-Lancet Commission, which was led by 37 world’s leading scientists, developed the first healthy and sustainable food guidelines known as the Planetary Health Diet (PHD), which recommends to raise the consumption of fruit, vegetable, legumes and nuts by more than twice while cutting the amount of added sugar and red meat by more than 50% (Willet et al., 2019). Meat production, which is responsible for almost 60% of all the greenhouse gas emissions generated by the food sector, affects the global environment most. Beef alone accounts for a quarter of the total emissions, and breeding animals for meat creates twice as big pollution as producing plant-based food, in general (Xu et al., 2021).

Analysis of social media data can be instrumental in establishing the population’s sentiment towards sustainable food consumption habits, for example, by analysing the tweets concerning meat consumption and by performing additional analysis of tweets about alternative proteins, vegetarian and / or vegan lifestyle and taking care of climate crisis, environmental sustainability and biodiversity. In addition, also the climate conditions of the particular society, as well as the peculiarities of the culture and national cuisine should be taken into consideration. Establishing the perception of healthiness in relation to an individual and environment is important also for the development of global food atlas, as context knowledge is key to a comprehensive big data system for food (Rostami et al., 2022).

6. Limitations

One of the limiting factors for an in-depth understanding of multi-sensory food aspects depicted in social media is the lack of food lexicon that could be used by computer scientists when defining the selection criteria of social network text data. Such lexicon should be developed according to the language of social networks and it should be context-sensitive (for example, by taking into account the urban environment and regional specifics in food culture). Lexicons that were used where derived from English lexicons, which means that some important aspects might be lost in the translation.

Development of food lexicons goes hand in hand with large-scale data digitalisation in local languages, which could expand the analytic possibilities of big data, without any doubt. If there had been access to digitalised recipe data in Latvian, it could have been used for the topic modelling analysis; but, since such data was not available, the topic modelling analysis was performed using North American and Mexican cuisine data.

Since food consumption is seasonal and closely related to the national food culture, in the context of food computing there is a need for close cooperation with food anthropology experts who could contribute to improving social media search queries by providing information on traditions and believes and a more nuanced picture of the food consumption by the society. While recipe data as such implies limited ability to analyse the actual cooking practices since recipes are strictly grouped according to the national cuisines, under the impact of globalisation these trends tend to mix nowadays and better indicators of eating habits could be provided by the analysis of food consumption habits in urban and rural settings (Grivins et al., 2020).

Another significantly restricting factor is studying a wealthy and digitally literate super-consumer who is active on social media networks, instead the society in general. While such fragmentation of information is characteristic of big data, big-scale social network data can still provide important knowledge on the behaviours of digitally active society groups. By choosing not to analyse big-scale social network data due to fragmentation, researchers cannot develop new methodologies and approaches for eliminating the drawbacks and still use the valuable but fragmented data. Although the author did not analyse how digital behaviour affects the analogue life, such study would be a logical continuation of the previous research.

Finally, there is a lack of important data sets (for example, consumer purchasing data, exact weather condition data, health-related data) necessary for carrying out a comprehensive analysis of food consumption.

7. Conclusion

As a result of this paper, **a new conceptual framework** for the analysis of food perception and consumption was developed by using the possibilities provided by big data. Results of the work significantly contributes to the development of food computing by offering different methodological approaches to analysing the big data of natural language about food, its perception and consumption trends. The new conceptual framework consists of four main elements:

1. collection of context knowledge on the topic;
2. detecting of frequency rates;
3. identification of affective reactions and analysis of the timescale;

4. adding of additional context variables.

In order to collect context knowledge, it is necessary to establish a global network of researchers representing different food-related research fields

It is necessary to focus on acquiring context knowledge on what affects the food systems and the society, to develop important research directions that have not been studied by researchers before. Food culture differs from country to country and from region to region, and it must be taken into account when working on big data about food. Food data that has been obtained in one field is rarely useful in another (Yamakata et al., 2022). Therefore, it is important to include analysis on national or regional food systems and their challenges, when doing research. Also social and economic data and the digital literacy of the society must be taken into account because social media data is only created by comparatively wealthy and digitally literate portion of the society, so these are restrictions that must be remembered when analysing big data on food choices and consumption. It is critical for the further development of food computing to include these context-knowledge related factors, and one way of doing it is creating cross-disciplinary research teams where computer scientists would work together with social science researchers—sociologists, food system historians and anthropologists. The current cooperation of computer science researchers with nutritionists is valuable but not sufficient for identifying those context factors that are important when analysing big data on food in different societies. Therefore, it is necessary to build a global network of food computing researchers that would bring together scientists from various disciplines.

Social media networks will be the most important source for detecting the frequency rates also in the future food computing, despite the fragmentation of such social media networks and partial possibilities to generalise such data

Food is one of the main topics discussed on the social network Twitter. As a platform for mainly text, not images and thanks to its availability, it is possible to trace the most accidental details of the everyday of Twitter users, including what, how and where they eat (Kāle et al., 2021). Compared to other sources of analysis, the social media data allow us to track spontaneous reactions when people tweet immediately, thus avoiding any possible prejudices that might arise using other methods for collecting opinions, like surveys and food diaries (Laguna et al., 2020). In addition, social media data is big data that allows to establish the food-related topics and issues discussed most frequently.

Detecting topic frequency rates allows to understand what dominates discussions on different foods. One of the most valuable divisions for studying the results of natural language data about food is the division into healthy and unhealthy food. Knowing how people describe healthy and unhealthy food is useful from the perspective of computer science for the development of various personalised apps for food recommendations and for the elaboration of the global food atlas (Rostami et al., 2022). By taking into consideration cognitive science studies that indicate the extent to which words affect our perception of food, specialists of computer linguistics have many opportunities for developing new research tasks, based on the information about the food-related natural language data.

Quality and availability of big data will play a critical role in identifying affective reactions and analysing the timescale

Findings of the cognitive science, for example understanding of affective reactions to food and multi-sensory experience related to food consumption, could be combined with methods applied by the computer science to big data analysis A large share of the new food and health apps are based on the circadian rhythm and the role of the food in this cycle, therefore better understanding of the cycle and depiction of food-related emotions is important. However, a factor limiting researches is the availability of data. While the data sets are not of sufficient size and detail, precise analysis is not possible. So, for example to correlate the sentiment analysis of food big data with the data on weather conditions, more detailed weather condition data is necessary—both daily weather data and information on geographic regions. The situation with individual health data is similar. While food computing researchers have partial access to important databases related to human health and lifestyle, the results of researches will be approximate. Thus, one of the primary tasks of the food computing research network would be to create and maintain open data sets that could be used in performing food computing analysis.

Adding of additional context variables will be an iterative process, taking into account the social, political and climate factors of specific regions

Cross-disciplinary approach raises understanding on food consumption and provides additional research impulse for food computing scientists. Although until now food computing has mainly been directed towards analysis of food ingredients, their compatibility and nutritional value (Min et al., 2019a, Ahn et al., 2011; Kim and Chung, 2016), an important turning-point was made by Jurafsky (2015) and Fenko et al. (2010) who promoted development of language analysis for language perception. In relation to food computing and language data analysis there is still a cross-disciplinary gap as important cognitive science issues have not been studied within the methodologies of these science branches. Thus, it is important to take into consideration the geographic and climate issues of the analysed big data and to include also regionally important policy documents that may impact the development of regional food systems.

The new conceptual framework adds to the research methodology of food computing and gives an example in the development of new research questions

The main task of this paper was to design a new conceptual research framework that is based on the articles by the author and co-authors and that solves important research challenges of cognitive and computer sciences in order to better understand big data on food consumption. The research work by the author illustrates what can be achieved if food consumption is analysed from an interdisciplinary point of view. In asking the following research question: **what methods should be applied to big data in order to better understand food consumers**, the author was motivated by the wide availability of social network data and the need to combine interdisciplinary science branches that can explain food consumption and promote healthier food and thus also healthier society. The author has solved some research gaps and provided a clear methodological system how to continue work with interdisciplinary researches while using the added value offered by big data for such analysis.

Acknowledgments

Part of this research (until August 2023) was supported by the project "Strengthening of the capacity of doctoral studies at the University of Latvia within the framework of the new doctoral model", identification No. 8.2.2.0/20/I/006.

References

Abbar, S., Mejova, Y., Weber, I. (2015). You tweet what you eat: Studying food consumption through twitter. *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, CHI '15*, pp. 3197–3206, New York, NY, USA. Association for Computing Machinery.

Ahn, Y.-Y. et al. (2011). Flavor network and the principles of food pairing. *Scientific Reports*, 1(1), p. 196. doi:10.1038/srep00196.

Allmendinger, J. (2015). *Quests for interdisciplinarity: A challenge for the ERA and HORIZON 2020.* Retrieved December 6, 2023, from https://commission.europa.eu/system/files/2019-04/allmendinger-interdisciplinarity.pdf

Andersen, T., Byrne, D. V., Wang, Q. J. (2021). How Digital Food Affects Our Analog Lives: The Impact of Food Photography on Healthy Eating Behavior. *Frontiers in psychology*, *12*, 634261.

Bakhshi, S., Kanuparthy, P., Gilbert, E. (2014). Demographics, weather and online reviews. *Proceedings of the 23rd International Conference on World Wide Web.* https://doi.org/10.1145/2566486.2568021

Boege, H. L., Bhatti, M. Z., St-Onge, M.-P. (2021). Circadian rhythms and meal timing: impact on energy balance and body weight. Current Opinion in Biotechnology, 70:1–6. Food Biotechnology Plant Biotechnology.

Bujisic, M., Bogicevic, V., Parsa, H. G., Jovanovic, V., Sukhu, A. (2019). It's raining complaints! How weather factors drive consumer comments and word-of-mouth. *Journal of Hospitality & Tourism Research*, 43(5), pp. 656–681.

Burnett, S. C., Ray, K. (2012). Sociology of Food. *The Oxford Handbook of Food History.* Oxford University Press.

Carolan, M. (2015). Publicising food: Big data, precision agriculture, and co-experimental techniques of addition. Sociologia Ruralis, 57:135–154.

Crippa, M., Solazzo, E., Guizzardi, D., Monforti-Ferrario, F., Tubiello, F.N., Leip, A. (2021). Food systems are responsible for a third of global anthropogenic GHG emissions. *Nature Food*.

Deksne, D. (2013). Finite State Morphology Tool for Latvian. *Proceedings of the 11th International Conference on Finite State Methods and Natural Language Processing.* St Andrews, Scotland: Association for Computational Linguistics, pp. 49–53. Available at: https://aclanthology.org/W13-1809 (Accessed: 13 September 2021). doi:10.3389/fcomp.2019.00003.

Fenko, A., Otten, J., Schifferstein, H. (2010). Describing product experience in different languages: The role of sensory modalities. *Journal of Pragmatics*, 42(12), pp. 3314–3327.

Ferran B., Jhaveri S., Lutz R., Isbister K., Wilde D. (2019). CHI '19: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, pp. 1–13. doi: https://doi.org/10.1145/3290605.3300908

Finn, S.M. (2019). Can "Taste" Be Separated from Social Class? *Food Fights*. University of North Carolina Press. doi: 10.5149/northcarolina/9781469652894.003.0005.

Gandhi, N., Meyer, C., Bogdanski, P., Walasek, L. (2023). Computational Analysis of Superfood Representations in News Media. *Journal of Food Products Marketing*, *29*(8-9), 270-290.

Gargouri, Y., Lefebvre, B., Meunier, J. (2003). Ontology Maintenance using Textual Analysis. *Unpublished Paper.* Available from: *Montreal: LCIA and LKMDA, UQAM*.

Grivins, M., Halloran, A., Kale, M. (2020). Eight megatrends in Nordic-Baltic food systems. Nordisk Ministerråd.

Ihler, A., Newman, D. (2011). Understanding errors in approximate distributed latent Dirichlet allocation. *IEEE Transactions on Knowledge and Data Engineering*, 24(5), pp. 952–960.

Jin, C., Bouzembrak, Y., Zhou, J., Liang, Q., van den Bulk, L. M., Gavai, A., Liu, N., van den Heuvel, L. J., Hoenderdaal, W., Marvin, H. J. (2020). Big data in food safety – a review. Current Opinion in Food Science, 36:24–32. Food Safety.

Jurafsky, D. (2015). *The Language of Food: A Linguist Reads the Menu*. Reprint edition. New York: W. W. Norton & Company.

Kāle, M., Agbozo, E. (2020a). Tracing complexity in food blogging entries. *CEUR Workshop Proceedings*

Kāle, M., Agbozo, E. (2020b). Healthy Food Depiction on Social Media: The Case of Kale on *Twitter. DHN Post-Proceedings*

Kāle, M., Agbozo, E. (2021). Utility of large-scale recipe data in food computing. *Baltic Journal of Modern Computing*, **9**(2), pp. 155-165.

Kāle, M., Rikters, M. (2021). Fragmented and Valuable: Following Sentiment Changes in Food Tweets. Workshop proceedings *Smell, Taste, and Temperature Interfaces (STT21), CHI 2021 Workshop.*

Kāle, M., Jain, R. (2023). The Future of Food Computing: Deepening the Scope by Broadening the Network. *Digital Humanities in the Nordic and Baltic Countries Publications.* 5, 1 (Oct. 2023), 70–78. DOI:https://doi.org/10.5617/dhnbpub.10652.

Kāle, M., Rikters, M. (2023). Exploring the Sentiment of Latvian Twitter Food Posts in Various Weather Conditions. *Digital Humanities in the Nordic and Baltic Countries Publications.* 5, 1 (Oct. 2023), 313–327. DOI:https://doi.org/10.5617/dhnbpub.10666.

Kāle, M., Šķilters, J., Rikters, M. (2021). Tracing multisensory food experiences on *Twitter*. *International Journal of Food Design*, 6:2, pp. 181–212.

Khatai, S. *et al.* (2019). An Iterative Model for Text Mining Using Big Data Technology. SSRN Scholarly Paper ID 3356552. Rochester, NY: Social Science Research Network.

Kim, K.-J., Chung, C.-H. (2016). Tell Me What You Eat, and I Will Tell You Where You Come From: A Data Science Approach for Global Recipe Data on the Web. *IEEE Access*, 4, pp. 1–1.

Laguna, L., Fiszman, S., Puerta, P., Chaya, C., Tárrega, A. (2020). The impact of COVID-19 lockdown on food priorities. Results from a preliminary study using social media and an online survey with Spanish consumers. *Food quality and preference*, 86, 104028. https://doi.org/10.1016/j.foodqual.2020.104028

Lopez, B., E., Hertel, T. (2020). Global food waste across the income spectrum: Implications for food prices, production and resource use. *Food Policy*. 98. 101874. 10.1016/j.foodpol.2020.101874.

Loughran, T., Mcdonald, B. (2011). When Is a Liability Not a Liability? Textual Analysis, Dictionaries, and 10-Ks. *The Journal of Finance*, 66(1), pp. 35–65.

Lucini, F.R. et al. (2017). Text mining approach to predict hospital admissions using early medical records from the emergency department. *International Journal of Medical Informatics*, 100, pp. 1–8.

Mai, R., Hoffmann, S. (2015). How to Combat the Unhealthy = Tasty Intuition: The Influencing Role of Health Consciousness. *Journal of Public Policy & Marketing*, 34(1), pp. 63–83.

Mai, R., Hoffmann, S., Helmert, J.R. et al. (2011). Implicit food associations as obstacles to healthy nutrition: the need for further research. *The British Journal of Diabetes & Vascular Disease*, 11(4), pp. 182–186.

Marvin, H.J.P. et al. (2017). Big data in food safety: An overview. *Critical Reviews in Food Science and Nutrition*, 57(11), pp. 2286–2295.

Metcalfe, R. (2019). Food Routes: Growing Bananas in Iceland and Other Tales from the Logistics of Eating. Cambridge, MA, USA: MIT Press.

Min, W. et al. (2018). You Are What You Eat: Exploring Rich Recipe Information for Cross Region Food Analysis. *IEEE Transactions on Multimedia*, 20(4), pp. 950–964.

Min, W., Jiang, S., Jain, R. (2019b). Food Recommendation: Framework, Existing Solutions and Challenges. *arXiv:1905.06269 [cs]* [Preprint].

Min, W., Jiang, S., Liu, L. (2019a). A Survey on Food Computing. *ACM Computing Surveys*, 52(5), p. 36.

Nag, N., Oh, H., Tang, M., Shi, M., Jain, R. (2021). Towards integrative multi-modal personal health navigation systems: Framework and application.

Nag, N., Pandey, V., Jain, R. (2017). Live Personalized Nutrition Recommendation Engine. *MMHealth'17: proceedings of the 2nd International Workshop on Multimedia for Personal Health and Health Care: October 23, 2017, Mountain View, CA, USA. ACM Workshop on Multimedia for Personal Health and Health Care.*

Nielsen, F. Å., (2011). A new anew: Evaluation of a word list for sentiment analysis in microblogs. *arXiv preprint arXiv*:1103.2903.

Nikolenko, S. I., Koltcov, S., Koltsova, O. (2017). Topic modelling for qualitative studies. *Journal of Information Science*, 43(1), pp. 88–102.

Obrist, M. et al. (2019). Space Food Experiences: Designing Passenger's Eating Experiences for Future Space Travel Scenarios. *Frontiers in Computer Science*, 1, p. 3.

Obrist, M., Marti, P., Velasco, C., Tu, Y. T., Narumi, T., Møller, N. L. H. (2018). The future of computing and food: Extended abstract. *Proceedings of the 2018 International Conference on Advanced Visual Interfaces*, AVI '18, New York, NY, USA. Association for Computing Machinery.

Puerta, P., Laguna, L., Vidal, L., Ares, G., Fiszman, S., Tárrega, A. (2020). Co-occurrence networks of Twitter content after manual or automatic processing. A case-study on "gluten free". *Food Quality and Preference*, 86, 103993.

Rosa, I., Lages, M., Grilo, C., Barros, R., Guarino, M. P. (2022). mHealth Applications to Monitor Lifestyle Behaviors and Circadian Rhythm in Clinical Settings: Current Perspective and Future Directions. *Frontiers in public health*, *10*, 862065.

Rostami, A., Nagesh, N., Rahmani, A., Jain, R. (2022). World Food Atlas for Food Navigation. *Proceedings of the 7th International Workshop on Multimedia Assisted Dietary Management*, pp. 39–47.

Rostami, A., Pandey, V., Nag, N., Wang, V., Jain, R. (2020). Personal Food Model. *Proceedings of the 28th ACM International Conference on Multimedia (MM '20)*. Association for Computing Machinery, New York, NY, USA, pp. 4416–4424.

Spence C. (2017). Gastrophysics: The new science of eating. Penguin Random House UK.

Spence, C. (2015a). Visual contributions to taste and flavour perception. *Colour Additives for Foods and Beverages,* Woodhead Publishing, pp. 189–210.

Spence, C. (2015b). Eating with our ears: Assessing the importance of the sounds of consumption to our perception and enjoyment of multisensory flavour experiences. *Flavour*, 4:3.

Spence, C. (2021). Explaining seasonal patterns of food consumption. *International Journal of Gastronomy & Food Science*, 24:100332.

Spence, C. et al. (2016). Eating with our eyes: From visual hunger to digital satiation. *Brain and Cognition*, 110, pp. 53–63.

Sproģis, U., Rikters, M. (2020). What can we learn from almost a decade of food tweets. Human Language Technologies – The Baltic Perspective. *Proceedings of the Ninth International Conference Baltic HLT 2020,* IOS Press, 328, pp. 191–198.

Straka, M., Straková, J. (2017). Tokenizing, POS Tagging, Lemmatizing and Parsing UD 2.0 with UDPipe. *Proceedings of the CoNLL 2017 Shared Task: Multilingual Parsing from Raw Text to Universal Dependencies*. *CoNLL 2017,* Vancouver, Canada: Association for Computational Linguistics, pp. 88–99.

Straka, M., Hajič, J., Straková, J. (2016). UDPipe: Trainable Pipeline for Processing CoNLL-U Files Performing Tokenization, Morphological Analysis, POS Tagging and Parsing. *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*. *LREC 2016*, Portorož, Slovenia: European Language Resources Association (ELRA), pp. 4290–4297.

Svenfelt, A., Zapico, J. L. (2016). Sustainable food systems with ICT. In Proceedings of ICT for Sustainability 2016, pp 194–201. Atlantis Press.

Thakkar, G., Pinnis, M. (2020). Pretraining and Fine-Tuning Strategies for Sentiment Analysis of Latvian Tweets. *Human Language Technologies—The Baltic Perspective-Proceedings of the Ninth International Conference Baltic HLT 2020*; IOS Press: Kaunas, Lithuania, 2020; pp. 55–61

Tran, B. X., Nghiem, S., Sahin, O., Vu, T. M., Ha, G.H., Vu, G. T., Ho, C. S. (2019). Modeling research topics for artificial intelligence applications in medicine: latent Dirichlet allocation application study. *Journal of medical Internet research*, 21(11), e15511.s

Vasiloglou, M. F., Lu, Y., Stathopoulou, T., Papathanail, I., Faeh, D., Ghosh, A., Baumann, M., Mougiakakou, S. (2020). Assessing mediterranean diet adherence with the smartphone: The medipiatto project. Nutrients, 12(12).

Velasco, C., Obrist, M. (2020). Multisensory Experiences, Oxford University Press.

Velasco, C., Michel, C., Spence, C. (2021). Gastrophysics: Current approaches and future directions. *International Journal of Food Design* 2021;6:137-52.

Vīksna, R., Skadiņa, I. (2020). Large Language Models for Latvian Named Entity Recognition. In Utka, A. et al. (Eds) *Frontiers in Artificial Intelligence and Applications*. IOS.

Wang, Q. J. et al. (2019). The Role of Intrinsic and Extrinsic Sensory Factors in Sweetness Perception of Food and Beverages: A Review. *Foods (Basel, Switzerland)*, 8(6), p. E211.

WEB (a). ACM Digital Library. (n.d.). *CHI: Conference on Human Factors in Computing Systems* Retrieved December 6, 2023, from https://dl.acm.org/conference/chi

WEB (b). CEA++'22 (2022). *CEA++'22: Proceedings of the 1st International Workshop on Multimedia for Cooking, Eating, and related APPlications.* New York: Association for Computing Machinery.

WEB (c). Council of European Union. (n.d.). *From farm to fork:* Retrieved December 6, 2023, from https://www.consilium.europa.eu/lv/policies/from-farm-to-fork/

WEB (d). European Parliament. (2021). *"Big Data: Definition, Advantages, Challenges (Infographic)"* Retrieved December 6, 2023, from https://www.europarl.europa.eu/news/lv/headlines/priorities/maksligais-intelekts-es/20210211STO97614/lielie-dati-definicija-prieksrocibas-sarezgijumi

WEB (e). *Personalized nutrition startup Zoe closes out Series B at $53M total raise* (2021). Retrieved December 6, 2023, from TechCrunch: shorturl.at/AXY06

WEB (f). The Lancet Gastroenterology & Hepatology. Obesity: another ongoing pandemic. Lancet Gastroenterol Hepatol 2021; 6: 411.

WEB (g). The PREDICT program. (2020, October 1). JoinZoe: Retrieved December 6, 2023, from https://joinzoe.com/whitepapers/the-predict-program

WEB (h). Twitter Developer Platform (n.d.a). Consuming streaming data. https://developer.twitter.com/en/docs/tutorials/consuming-streaming-data. Accessed 15 August 2021.

WEB (i). UC Santa Cruz. (n.d.). *Making Sense of Human Food Interaction:* Retrieved December 6, 2023, from https://www2.ucsc.edu/hfi/conceptualmodel.html

WEB (j). World Economic Forum. (2020, September 20). *How digital technology can accelerate food sustainability:* Retrieved December 6, 2023, from https://www.weforum.org/agenda/2022/09/how-digital-technology-can-accelerate-food-sustainability/

WEB (k). World Health Organization. (2021, June 9). *Obesity and overweight:* Retrieved December 6, 2023, from https://www.who.int/news-room/fact-sheets/detail/obesity-and-overweight

WEB (l). World Health Organization. (2021, October 20). *Making diets environmentally friendly: Nordic countries lead the way:* Retrieved December 6, 2023, from https://www.who.int/europe/news/item/20-10-2021-making-diets-environmentally-friendly-nordic-countries-lead-the-way

Willett, W., Rockström, J., Loken, B., Springmann, M., Lang, T., Vermeulen, S., Garnett, T., Tilman, D., DeClerck, F., Wood, A., Jonell, M., Clark, M., Gordon, L. J., Fanzo, J., Hawkes, C., Zurayk, R., Rivera, J. A., De Vries, W., Majele Sibanda, L., Afshin, A., Murray, C. (2019). Food in the Anthropocene: the EAT-Lancet Commission on healthy diets from sustainable food systems. *Lancet (London, England)*, *393*(10170), pp. 447–492.

Xu, X., Sharma, P., Shu, S., Lin, T.S., Ciais, P., Tubiello, F. N., Smith, P., Campbell, N., Jain, A. (2021). Global greenhouse gas emissions from animal-based foods are twice those of plant-based foods. Nature Food. 2. 1-9. 10.1038/s43016-021-00358-x.

Yamakata, Y., Mougiakakou, S., Jain, R. (2022). Cea++2022 panel - toward building a global food network. In Proceedings of the 1st International Workshop on Multimedia for Cooking, Eating, and Related APPlications, CEA++ '22, pp. 59–60, New York, NY, USA. Association for Computing Machinery.

Received January 15, 2024, accepted February 6, 2024