



MARTIN-LUTHER-UNIVERSITÄT
HALLE-WITTENBERG

Projektbericht zum Modul
Information Retrieval und Visualisierung
Wintersemester 2021/22

Visualisierung von Nährwerten und Kalorien in Lebensmitteln

Eingereicht bei: Dr. Alexander Hinneburg
Eingereicht am: 20. Dezember 2021
Eingereicht von: Delia Storch
Matrikelnummer: 220233073
GitHub Repository: <https://github.com/95deli/ElmFoodProject>
Projektwebseite: <https://95deli.github.io/ElmFoodProject/>

Inhaltsverzeichnis

Abbildungsverzeichnis	III
1. Einleitung	1
1.1 Anwendungshintergrund	2
1.2 Zielgruppen.....	2
1.3 Überblick und Beiträge.....	3
2. Daten	4
2.1 Technische Bereitstellung der Daten	5
2.2 Datenvorverarbeitung	6
3. Visualisierungen.....	7
3.1 Analyse der Anwendungsaufgaben	7
3.2 Anforderungen an die Visualisierungen	9
3.3 Präsentation der Visualisierungen	10
3.3.1 Visualisierung Eins	10
3.3.2 Visualisierung Zwei	12
3.3.3 Visualisierung Drei	13
3.4 Interaktion.....	14
4. Implementierung	15
5. Anwendungsfälle.....	16
5.1 Anwendung Visualisierung Eins	16
5.2 Anwendung Visualisierung Zwei	17
5.3 Anwendung Visualisierung Drei	19
6. Verwandte Arbeiten	21
7. Zusammenfassung und Ausblick	22
Anhang: Git-Historie.....	IV
Literatur	VII
Eidesstattliche Erklärung.....	VIII

Abbildungsverzeichnis

Abbildung 1: Scatterplot	11
Abbildung 2: Parallele Koordinaten.....	12
Abbildung 3: Ausschnitt aus der Baumhierarchie.....	13
Abbildung 4: Vergleich von Proteinen und Kohlenhydraten im Scatterplot	16
Abbildung 5: Vergleich der Nährstoffgehalte in Parallelen Koordinaten.....	18
Abbildung 6: Lebensmittelkategorie Getreideprodukte in der Baumhierarchie	19
Abbildung 7: Lebensmittelkategorie Nüsse in der Baumhierarchie	20
Abbildung 8: Lebensmittelkategorie Früchte in der Baumhierarchie.....	20

1. Einleitung

Die exakte Zusammensetzung von Lebensmitteln wird für Verbraucher:innen immer relevanter. Von ihr hängen die Kaufentscheidungen ab und letztendlich auch – wie aus zahlreichen Studien hervorgeht – die körperliche und geistige Gesundheit des Menschen. Die D-A-CH-Referenzwerte¹ für die Nährstoffzufuhr sowie Referenzwerte anderer nationaler und internationaler Gesundheitsorganisationen bieten eine Grundlage für die Umsetzung einer vollwertigen Ernährung in der Praxis [1]. Sie definieren altersspezifische Mengen der Kalorien und Nährstoffe, die täglich benötigt werden, um lebenswichtige Funktionen sicherzustellen und Krankheiten vorzubeugen. Je nach Alter, Lebenssituation und äußeren Einflüssen können sich Nährstoffbedarf und Ernährungsgewohnheiten leicht ändern [2]. Beiträge zu einer ausgewogenen oder klimafreundlichen Ernährung, der täglichen Referenzmenge an Zucker, Informationen zu Nahrungsergänzungsmitteln und unzähligen weiteren verwandten Themen können meist schnell im Internet gefunden werden.

Auf der Suche nach der geeigneten Ernährungsform oder alternativen Produkten ist allerdings ein direkter Vergleich der Produkte nützlich. Dafür werden häufig die Nährstoffe und Kalorien der Produkte untersucht. Eine Nährwertinformation pro 100 Gramm befindet sich praktischerweise auf jeder Verpackungsrückseite. Zuerst erfolgt die Kalorienangabe, danach die Angabe von Fett, davon gesättigte Fettsäuren, Kohlenhydraten, Ballaststoffen und Proteinen. Je nach Produkt können auch zusätzliche Angaben enthalten sein. Jedoch sind Verbraucher:innen lediglich in der Lage, das Produkt zu bewerten, das sie gerade in der Hand halten. Um mehrere Produkte zu vergleichen, könnten sie nun unzählige verfügbare Artikel lesen oder die Lebensmittel einzeln vergleichen. Fraglich ist dabei aber, ob die benötigten Informationen auch übersichtlich und verständlich aufbereitet sind und vor allem schnell gefunden werden können.

Das Ziel dieses Projekts ist es, die Daten so aufzubereiten, dass Interessierte der Zielgruppe einen umfassenden Überblick über die ausgewählten Lebensmittel mit ihren entsprechenden Nährwerten und Kalorien erhalten. Im Fokus steht dabei die Vergleichbarkeit der Produkte anhand ihrer Nährwerte und Kalorien pro 100 Gramm. Dafür sollte auf den ersten Blick verständlich sein, welches Verhältnis zwischen zwei oder mehreren Nährstoffen in einem Produkt besteht. Darüber hinaus sollten die Produkte auch untereinander vergleichbar sein. Von besonderer Bedeutung ist außerdem die Darstellung verschiedener Lebensmittelgruppen mit den

¹ Fachgesellschaften Deutschlands, Österreichs und der Schweiz (D – Deutschland; A – Österreich; CH – Schweiz)

zugehörigen Produkten und Informationen der entsprechenden Kalorien. Die Umsetzung des Projektziels erfolgt anhand drei ausgewählter Visualisierungstechniken mit der Programmiersprache Elm.

1.1 Anwendungshintergrund

In dieser Forschungsarbeit werden die Lebensmitteldaten so aufbereitet, dass Anwender:innen übersichtliche Informationen zu Nährwerten und Kalorien erhalten können. Die Umsetzung erfolgt mithilfe der drei Visualisierungstechniken Scatterplot, Parallele Koordinaten und Baumhierarchie. Es folgt ein kurzer Überblick.

Die erste Visualisierungstechnik zur Darstellung von Nährwerten in Lebensmitteln ist der Scatterplot (Streudiagramm). In einem Scatterplot lassen sich zwei verschiedene numerische Attribute in einem zweidimensionalen Koordinatensystem darstellen und vergleichen [3]. Die Attribute sind dabei die verschiedenen Nährstoffe der Lebensmittel. Die Anwender:innen der Visualisierung können Kohlenhydrate, Proteine, Ballaststoffe, Fette und gesättigte Fettsäuren für die Koordinatenachsen auswählen. So können Informationen über die Verhältnisse von zwei ausgewählten Nährstoffgehalten gewonnen werden, welche darüber hinaus auch mit anderen Lebensmitteln in der Darstellung verglichen werden können.

Bei der zweiten Visualisierungstechnik handelt es sich um Parallele Koordinaten. Dadurch lassen sich mehrdimensionale Daten in einem zweidimensionalen Raum darstellen [4]. Jedes ausgewählte Attribut erhält eine eigene von vier Achsen. In dieser Darstellung können vier Nährwerte gleichzeitig miteinander verglichen werden. So lässt sich ein Produkt umfassend bewerten und darüber hinaus in seiner Gesamtbilanz mit anderen Produkten vergleichen.

Die dritte Visualisierungstechnik ist die Baumhierarchie [5]. Dadurch wird eine Darstellung der hierarchischen Beziehungen sowie eine Sortierung der Lebensmittel mit entsprechenden Kalorien realisiert. Möglicherweise möchten Anwender:innen die Kalorienwerte nicht nur für ein Produkt ermitteln, sondern für mehrere Produkte auf einmal. Bei dieser Visualisierung werden die Lebensmittel kategorisiert und erhalten Informationen zu Kalorien, sodass der Kaloriengehalt ähnlicher Produkte schnell verglichen werden kann.

1.2 Zielgruppen

Für die vorliegenden Daten sind im Rahmen dieser Arbeit zwei potenzielle Zielgruppen denkbar: ernährungsbewusste (interessierte) Konsument:innen und Produktentwickler:innen aus der Lebensmittelindustrie. Im Folgenden werden die Zielgruppen der Visualisierungen analysiert.

Ernährungsbewusste Konsument:innen weisen ein besonders großes Interesse an den Lebensmitteln auf, die sie regelmäßig konsumieren. Sie achten auf eine ausgewogene Ernährung und verfolgen dabei Ziele wie die Reduktion oder Ergänzung spezieller Nährwerte, die Einhaltung von Referenzwerten oder den Vergleich verschiedener Nährwerte und Kalorien von Produkten aus gleichen Produktkategorien. Es ist davon auszugehen, dass sich Verbraucher:innen dieser Zielgruppe inhaltlich mit dem Thema Ernährung auseinandersetzen und dass sie dadurch über ein umfangreiches Vorwissen verfügen. Sie sind in der Lage, Lebensmitteldaten intuitiv zu erfassen und einzuordnen. Diese Zielgruppe kann ihr Vorwissen mithilfe der Visualisierungen um Informationen zu Lebensmitteln erweitern, die Nährwerte ähnlicher Produkte vergleichen und darauf basierend Kaufentscheidungen für Lebensmittel zu ihrem gesundheitlichen Vorteil treffen. Unter der Annahme, dass zu ihrem Vorwissen auch die Kenntnis der Referenzwerte zählt, können sie relevante Informationen vor allem aus dem Scatterplot und aus den Parallelen Koordinaten gewinnen. Handelt es sich bei den Anwender:innen um Interessierte, könnten diese allerdings einen geringeren Wissensstand aufweisen. Auch sie können durch die Visualisierungen einen umfassenden Überblick über die Lebensmittel erhalten, wenn sie bereit sind, etwas Zeit zu investieren, um die Daten einordnen zu können. Alle drei Visualisierungen bieten für sie vor allem den Vorteil, dass die Informationen nicht für jedes Produkt einzeln herausgesucht werden müssen.

Eine weitere potenzielle Zielgruppe ist die der Produktentwickler:innen aus der Lebensmittelindustrie. Diese sind sich den Nährwerten, Kalorien und Referenzwerten von Lebensmitteln sicher bewusst. Unter Umständen verfügen sie aber nur über Nischenwissen zu den Lebensmitteln aus ihrer spezifischen Branche. Um innovativ zu agieren, müssen Entwickler:innen stets neue Möglichkeiten in Betracht ziehen. So könnte die Substitution von Komponenten in der Rezeptur eine Möglichkeit darstellen, um eigens festgelegte Grenzwerte nicht zu überschreiten (beispielsweise eine maximale Kalorienanzahl) oder um einen besonders hohen Anteil von einem bestimmten Nährstofftypen im Endprodukt zu verarbeiten (beispielsweise ein hoher Anteil an Proteinen). Diese Zielgruppe kann von den drei Visualisierungen profitieren, indem sie ausgehend von ihrem Vorwissen die benötigten Informationen aus den Darstellungen heraus sucht.

1.3 Überblick und Beiträge

Das Interesse an gesunder Ernährung, speziellen Ernährungsformen oder der Zusammensetzung von Lebensmitteln steigt. Im Internet und in zahlreichen Büchern und Zeitschriften sind Informationen zum Thema Ernährung mit unterschiedlichsten Ansätzen zu finden. Diese liegen allerdings meist in Text- oder Tabellenform vor. Um sich schnell und unkompliziert einen

Überblick über die Nährwerte und Kalorien in Lebensmitteln zu verschaffen, eignet sich eine Visualisierung entsprechender Daten. Diese soll möglichst so gestaltet sein, dass Anwender:innen keine Programmierkenntnisse benötigen, um mit den Darstellungen interagieren und Informationen erhalten zu können.

Die Visualisierungstechnik Scatterplot wird als verbreitete Technik eingeschätzt, aus der Informationen intuitiv entnommen werden können. Ein Scatterplot bietet die Möglichkeit, interaktiv zwei konkrete Nährstoffe auszuwählen und gegenüberzustellen. Dadurch können Anwender:innen genau die Informationen erhalten, die für sie relevant sind.

Die Visualisierungstechnik der Parallelen Koordinaten ist möglicherweise nicht vergleichbar intuitiv und verbreitet. Eine Interaktion mit der Darstellung ist zwar auch ohne Vorwissen möglich. Jedoch ist es hilfreich, wenn eine gezielte Suche nach Lebensmitteln anhand vorher bekannter charakteristische Nährstoffeigenschaften durchgeführt wird, um aus den Daten den maximal möglichen Mehrwert zu ziehen. Dennoch wird diese Visualisierungstechnik im Vergleich zu potenziellen Alternativen als am besten geeignete Möglichkeit entsprechend ihres Anwendungshintergrunds eingeschätzt.

Die Visualisierungstechnik Baumhierarchie ermöglicht eine strukturierte und übersichtliche Darstellung der Daten ohne Interaktionsmöglichkeit. Inbegriffen sind die Lebensmittel in Lebensmittelgruppen sowie die entsprechenden Kalorienwerte. Es ist davon auszugehen, dass Anwender:innen bei dieser Darstellung die Informationen leicht erfassen können, selbst wenn sie über kaum oder kein Hintergrundwissen verfügen.

In Kapitel 2 folgen zunächst Informationen zur Datengrundlage, zur technischen Bereitstellung und zur Datenvorverarbeitung. Eine genaue Beschreibung der Umsetzung der Visualisierungen erfolgt in Kapitel 3. Kapitel 4 beinhaltet die Implementierung in Elm. In Kapitel 5 werden die Anwendungsfälle zu den Visualisierungen aus Kapitel 3 ausgeführt. Verwandte Arbeiten in Kapitel 6 sowie Zusammenfassung und Ausblick in Kapitel 7 bilden den Abschluss dieser Arbeit.

2. Daten

Der für die Visualisierungen verwendete Datensatz wurde von einer Nutzerin auf der Online-Plattform *Kaggle* bereitgestellt. Die Daten dienten ursprünglich als Grundlage für ein Python-Projekt mit dem Titel „Nutritional facts for most common foods“ [6]. Die Datei beinhaltet im Original zehn Spalten mit insgesamt 329 Datensätzen. Der Datensatz beginnt mit den Lebensmittelnamen in der Spalte *Food*. Die folgenden Mengenangaben unter *Measure* stellen eine durchschnittliche Portion dar. Die Angabe des Gewichts unter *Grams* gibt das Gewicht der

entsprechenden Portionen an. Die Werte der Kalorien unter *Calories* sowie die Werte für Proteine unter *Proteins*, Fette unter *Fats*, gesättigte Fette unter *Sat.Fat*, Ballaststoffe unter *Fiber* und Kohlenhydrate unter *Carbs* beziehen sich auf das angegebene Gewicht pro Portion. Unter *Category* ist für jedes Produkt eine aus insgesamt zwölf Kategorien erfasst.

Grundsätzlich sind die Daten gut geeignet, um die Fragestellungen der Zielgruppen beantworten zu können. Sie beinhalten die wesentlichen Informationen über ausgewählte Lebensmittel, die für die Zielgruppe interessant sein könnten. Allerdings enthält der Datensatz einige fehlerhafte Daten in Form von doppelten, nicht zuordenbaren oder unvollständigen Werten, welche entfernt werden müssen. Es ist darüber hinaus eine Erweiterung um drei Spalten erforderlich, um die benötigten Daten insbesondere für die Baumhierarchie aufzubereiten.

Die Datei wurde zunächst um die Spalte der Portionsgrößen reduziert, anschließend um drei Kategoriespalten erweitert und beschädigte Datensätze wurden entfernt. Daraus resultiert eine Datei mit zwölf Spalten und 280 Datensätzen. Es wurden außerdem geringfügige Anpassungen bei den Bezeichnungen der Lebensmittel und Kategorien vorgenommen. Zudem wurden alle Werte der Kalorien, Proteine, Fette, gesättigten Fettsäuren, Ballaststoffe und Kohlenhydrate umgerechnet, sodass sich die Nährwerte auf ein einheitliches Gewicht von 100 Gramm für jedes Lebensmittel beziehen. Die detaillierten Schritte der Datenvorverarbeitung sind dem Kapitel 2.2 zu entnehmen.

2.1 Technische Bereitstellung der Daten

Wie bereits genannt, wird der ursprüngliche Datensatz als CSV-Datei über die Online-Plattform Kaggle zur Verfügung gestellt. Die technische Bereitstellung der verwendeten originalen und verarbeiteten Daten erfolgt weiterhin über ein öffentliches GitHub Repository. Die Dateien befinden sich im Ordner *Daten* in den Unterordnern *CSV*, *Excel* und *JSON*. Da die Originaldaten im CSV-Dateiformat vorliegen, werden sie zunächst in Excel eingefügt, bearbeitet und anschließend wieder in CSV überführt sowie in JSON umgewandelt.

Im Scatterplot und in den Parallelen Koordinaten wird die CSV-Datei mit den verarbeiteten Daten der Lebensmittel verwendet. Die modifizierte CSV-Datei mit der Bezeichnung *NutrientsFINAL.csv* befindet sich im Ordner *Daten/CSV*. Alle Werte innerhalb der CSV-Datei sind mit einem Komma getrennt und jedes Dezimaltrennzeichen ist ein Punkt. Die CSV-Datei beinhaltet alle Werte der IDs und Lebensmittelnamen, drei Kategorien, einheitliche Mengenangaben (100 Gramm), Kalorien, Proteine, Fette, gesättigte Fettsäuren, Ballaststoffe und Kohlenhydrate.

In der JSON-Datei werden die Beziehungen für die Baumhierarchie dargestellt. Dafür ist eine Umwandlung der Daten erforderlich. Die Baumhierarchie greift auf die JSON-Datei mit den verarbeiteten Daten der Lebensmittel zu. Die JSON-Datei hat die Bezeichnung *BaumhierarchieJSON.json* und befindet sich im Ordner *Daten/JSON*.

2.2 Datenvorverarbeitung

Der ursprüngliche Datensatz erfordert einige Anpassungen, um die geeignete Form für die Visualisierungen zu erhalten. Da nicht alle Attribute für die Visualisierungen relevant sind und wiederum andere für die Visualisierungen gebraucht werden, wurde der Originaldatensatz in sieben Schritten verarbeitet, um die benötigte Datengrundlage zu erhalten.

Zunächst wurde die CSV-Datei mit den Originaldaten in eine Excel-Datei konvertiert. Diese ist im Ordner *Daten/Excel* mit der Bezeichnung *NutrientsRAW.xlsx* gespeichert. Für alle Datensätze wurde eine eindeutige fünfstellige ID vergeben, um einen transparenten Datenvorverarbeitungsprozess zu ermöglichen und um in weiteren Verarbeitungsschritten eine eindeutige Zuordnung zu gewährleisten.

Anschließend wurden alle *t*-Werte² entfernt und die Formatierung der Zellen überprüft und in eine einheitliche Form gebracht. Die Werte der Kalorien und Nährstoffe wurden auf eine Nachkommastelle formatiert. Letztendlich wurden fehlerhafte Datensätze aufgrund von doppelten, nicht zuordenbaren oder unvollständigen Informationen ausgeblendet.

Im folgenden Schritt wurde eine zweite Tabelle im gleichen Dokument angelegt. Aus Gründen der besseren Lesbarkeit wurden alle Spaltennamen sowie die Namen der einzelnen Lebensmittel und die Kategorienamen angepasst und vereinheitlicht. Es wird davon ausgegangen, dass die Zielgruppen in der Lage sind, englischsprachige Informationen zu erfassen. Deswegen wurde auf eine Übersetzung der Lebensmittelnamen und Nährstoffe verzichtet.

Die angegebenen Werte für Kalorien und Nährwerte beziehen sich in der zweiten Tabelle auf eine einheitliche Lebensmittelmenge von 100 Gramm. Im Originaldatensatz hingegen sind die Lebensmittel in Gramm pro Portion angegeben. So richten sich beispielsweise die Kalorien und Nährwerte von Kuhmilch (*Cows' milk*) mit der ID *11000* nach der Portionsgröße von einem Quart (*1 qt.*), was umgerechnet einem knappen Liter Milch entspricht. Da sich die Angabe des Gewichts in Gramm (*Grams*) für die verschiedenen Lebensmittel nach den unterschiedlichen Portionsgrößen richtet, ist keine Vergleichbarkeit der einzelnen Lebensmittel möglich.

² Spurenelemente (*trace amounts*)

Aufgrund dessen wurde die Portionsgröße aus allen für die Visualisierung relevanten Datensätzen entfernt und alle Werte für Kalorien und Nährstoffe wurden mithilfe des entsprechenden Umrechnungsfaktors in der *solver*-Spalte auf jeweils 100 Gramm umgerechnet.

Da mittels der vorliegenden Daten keine sinnvolle Visualisierung der Baumhierarchie umsetzbar ist, musste der Datensatz um die Attribute *supercategory* und *caloriescategory* erweitert werden. In der Überkategorie *supercategory* ist für alle Lebensmittel eine Zuordnung zu Nahrung (*Food*) oder Getränke (*Drinks*) erfasst. Die Kategorie der Kalorien *caloriescategory* beinhaltet kategorische Informationen über die Kalorien der Lebensmittel, wobei es zehn verschiedene Kategorien gibt. Dies ermöglicht die strukturierte Integration der Daten in eine JSON-Datei für die Baumhierarchie.

Darüber hinaus ist in der Excel-Datei auch eine Spalte für die Erstellung der JSON-Bezeichnungen enthalten. Die Bezeichnungen wurden bei der Erstellung der JSON-Datei aus der Excel-Datei entnommen und an der entsprechenden Stelle eingefügt. Die JSON-Datei wurde mit Visual Studio Code erstellt und ist im Ordner *Daten/JSON* mit der Bezeichnung *BaumhierarchieJSON.json* gespeichert.

Die zweite Tabelle mit den vollständigen Werten aller Attribute wurde aufgrund der Übersichtlichkeit in einem neuen Excel-Dokument zusammengefasst. Dieses Dokument mit der Version aller finaler Daten stellt die Grundlage für die zu erstellende CSV-Datei dar, die für die Visualisierung von Scatterplot und parallelen Koordinaten benötigt wird. Dabei stellte das Konvertieren der Excel-Datei in eine CSV-Datei eine besondere Herausforderung dar, da dabei zu beachten ist, dass das Dezimaltrennzeichen innerhalb der Werte ein Punkt ist und dass die Daten kommasetrennt sind. Die CSV-Datei ist im Ordner *Daten/CSV* mit der Bezeichnung *Nutrients-FINAL.csv* gespeichert.

3. Visualisierungen

In diesem Kapitel erfolgt die Vorstellung der drei Visualisierungen. Im Vordergrund stehen dabei die Anforderungen der Zielgruppe, die Präsentation der drei Visualisierungen und die Interaktionsmöglichkeiten.

3.1 Analyse der Anwendungsaufgaben

Durch die Visualisierungen soll eine übersichtliche Darstellung der Informationen zu Kalorien und Nährstoffen ausgewählter Lebensmittel ermöglicht werden. Nutzer:innen soll es möglich sein, die Daten schnell zu erfassen und zusätzliche interessante Erkenntnisse zu gewinnen.

Ziel der ersten Visualisierung ist es, zwei Attribute (Nährwerte) unterschiedlicher Lebensmittel miteinander zu vergleichen. Dafür wurde die Visualisierungstechnik Scatterplot gewählt. Daraus lassen sich Erkenntnisse zu Zusammenhängen zwischen zwei ausgewählten Nährwerten in Lebensmitteln ableiten: *Wie ist das Verhältnis zwischen Kohlenhydraten und Proteinen in Lebensmitteln?* Es können aber auch allgemeine Aussagen zu Lebensmitteln getroffen werden: *Welche Lebensmittel haben die meisten Kohlenhydrate? Welche Lebensmittel haben am wenigsten Fett?* Die Punkte im Scatterplot stellen die Lebensmittel dar und zeigen bei der Auswahl mit der Maus den entsprechenden Lebensmittelnamen und die Namen und Werte der zwei ausgewählten Nährwerte an. Da der Vergleich von zwei Faktoren mittels eines zweidimensionalen Koordinatensystems zu den gängigen Techniken bei der Darstellung von Statistiken gehört, sollten Anwender:innen die Bedeutung der Daten schnell einordnen können. Über einen Hinweis an den Buttons zur Zuweisung des Attributs auf der x- und y-Achse soll Anwender:innen die Benutzung erleichtert werden.

Ziel der zweiten Visualisierung ist es, mehrere Attribute unterschiedlicher Lebensmittel miteinander zu vergleichen. Dafür wurde die Visualisierungstechnik Parallele Koordinaten verwendet. Daraus lassen sich grundsätzliche Zusammenhänge von mehreren Nährstoffen in Lebensmitteln ableiten: *Welche Lebensmittel haben umfassend eher schlechte Nährwerte, beispielsweise einen hohen Wert an Kohlenhydraten und Fett, aber wenige Proteine und Ballaststoffe?* Es lassen sich aber auch ähnliche Lebensmittel in ihrer Gesamtbilanz vergleichen, wenn Anwender:innen über entsprechendes Vorwissen verfügen: *Welches Produkt aus einer bestimmten Kategorie deckt den täglichen Bedarf an Ballaststoffen am besten?* Anders als beim Scatterplot werden die Daten der Lebensmittel jetzt auf vier Achsen mit durch Linien verbundenen Attributen dargestellt. Bei der Auswahl der Linien mit der Maus wird der Lebensmittelname sowie Namen und Werte der vier ausgewählten Nährwerte angezeigt. Auch hier gibt es einen Hinweis an den Buttons zur Zuweisung der Attribute.

Ziel der dritten Visualisierung ist es, die in Nahrung oder Getränken enthaltenen Kalorien abzubilden. Dafür werden die Lebensmittel den Lebensmittelgruppen zugeordnet und aufsteigend nach Kalorien sortiert. Dafür wurde die Visualisierungstechnik Baumhierarchie verwendet. Die Baumdarstellung gibt Auskunft über die folgenden Fragestellungen: *In welchen Lebensmittelgruppen haben die Lebensmittel besonders viele Kalorien? Welche Lebensmittel haben im Einzelnen besonders wenige Kalorien?* Die Anforderungen an die Visualisierung unterscheiden sich von den beiden vorher genannten Visualisierungen, da hierbei die Kalorien, nicht aber die Nährstoffe abgebildet werden. Zudem sollen in dieser Darstellung die Lebensmittelgruppen

stärker mit in die Visualisierung einbezogen werden. Die Baumdarstellung sollte für Anwender:innen mit unterschiedlichen Wissensständen leicht verständlich sein.

Betrachten Anwender:innen darüber hinaus die Gesamtheit der bereitgestellten Informationen in allen drei Visualisierungen, sollen sie zusätzliche Erkenntnisse gewinnen und Rückschlüsse auf weitere Zusammenhänge ziehen können. Sie könnten so beispielsweise herausfinden, wie Nährstoffe den Kaloriengehalt in Lebensmitteln beeinflussen.

3.2 Anforderungen an die Visualisierungen

Aus dem vorhergehenden Kapitel geht hervor, dass für jede der drei Visualisierungen eine spezifische Zielstellung besteht, aus der sich die zu erfüllenden Anforderungen ergeben. Zu den Anforderungen gehört vor allem der Vergleich von zwei oder mehreren Nährstoffen verschiedener Lebensmittel sowie eine Übersicht über die enthaltenen Kalorien. Anwender:innen sollen die Lebensmittel in einem Scatterplot und in den Parallelen Koordinaten interaktiv auswählen und selbst festlegen, welche Nährstoffe miteinander verglichen werden. Da der Wert der Kalorien eines Produktes nicht direkt zu den Nährstoffen zählt, der Vergleich für Interessierte aber trotzdem von Bedeutung ist, soll dieser durch die Umsetzung der Baumhierarchie dargestellt werden.

Da im Rahmen des Projekts angenommen wird, dass die Zielgruppen nicht aus dem Fachbereich der Informatik stammen, ist die verständliche und übersichtliche Darstellung der Daten die entscheidendste Anforderung an die Visualisierungen. Dazu gehört eine sprachlich korrekte sowie optisch gut lesbare Beschriftung der Elemente wie Buttons, Achsen, Punkte und Linien. Auch die farblichen Darstellungen der Punkte im Scatterplot sowie der Linien in Parallele Koordinaten sollen zur besseren Übersicht beitragen.

Die erste Visualisierung soll das Verhältnis zwischen zwei Attributen der Lebensmittel in einem Scatterplot abbilden. Dafür müssen die Koordinatenachsen mit den ausgewählten Attributen beschriftet sein und eine Einteilung erhalten, die sich nach den Zahlenwerten dieser Attribute richtet. Die Zahlenwerte der Attribute sind in Gramm angegeben und beziehen sich auf ein einheitliches Gewicht der Lebensmittel von 100 Gramm. Deswegen wurde auf Einheiten verzichtet, ein entsprechender Hinweis befindet sich aber auf der Projektwebseite. Zusätzlich sollen Anwender:innen die Möglichkeit erhalten, für jede Koordinatenachse eines aus den verfügbaren Attributen auszuwählen. Die Umsetzung erfolgt durch Buttons, wobei es für die x- und die y-Achse die Option gibt, ein Attribut individuell auszuwählen bzw. das Attribut zu wechseln, indem der zugehörige Button betätigt wird. Aus der Kombination der Zahlenwerte zweier Attribute ergibt sich der zugehörige Punkt im Scatterplot. Damit die Punkte unterschieden

werden können ist es notwendig, dass sie eine eindeutige Beschriftung erhalten, einschließlich Lebensmittelname, ausgewählte Nährstoffe und deren Wert pro 100 Gramm. Die Beschriftung soll allerdings nur angezeigt werden, wenn Anwender:innen die Punkte mit der Maus auswählen, damit sich die Beschriftungen nicht überschneiden, falls sich Punkte überlagern.

Die zweite Visualisierung soll den Zusammenhang zwischen mehreren Attributen der Lebensmittel mithilfe der Visualisierungstechnik Parallele Koordinaten abbilden. Die Einteilung der vier Spalten für die ausgewählten Attribute richtet sich wie beim Scatterplot nach den Zahlenwerten der Attribute in Gramm (für 100 Gramm von jedem Lebensmittel). Die dynamische Anordnung der Spalten soll über Buttons erfolgen, wobei es für jede Spalte die Option gibt, eines aus den verfügbaren Attributen auszuwählen. Die Spalte soll automatisch den Namen entsprechend dem ausgewählten Button erhalten. Lebensmittelnamen, ausgewählte Attribute und deren Zahlenwerte sollen auf zwei verschiedene Arten visualisiert werden, um eine Unterscheidung der Daten zu gewährleisten. Zum einen sollen zusammengehörende Attributwerte durch Linien verbunden werden, die bei der Auswahl mit der Maus farbig angezeigt werden. Zum anderen soll die zugehörige Beschriftung, ähnlich wie beim Scatterplot, angezeigt werden, in diesem Fall aber für vier Attribute.

Die dritte Visualisierung soll dazu dienen, die Lebensmittel aus dem Datensatz unter Berücksichtigung der Lebensmittelgruppe nach Kalorien zu sortieren und in Form eines Baumdiagramms abzubilden. Um herausfinden zu können, ob ein Produkt viele oder wenige Kalorien beinhaltet, soll die hierarchische Strukturierung so gewählt werden, dass zunächst aus ihr hervorgeht, zu welcher Lebensmittelgruppe die Produkte gehören, um die Suche zu erleichtern. Auf der letzten Stufe sollen die Lebensmittel unter der entsprechenden Kaloriengruppe aufgelistet werden. In dieser Visualisierung besteht keine Interaktionsmöglichkeit, durch vertikales oder horizontales Scrollen soll lediglich der Bildausschnitt verändert werden.

3.3 Präsentation der Visualisierungen

Nachdem die Anforderungen an die Visualisierungen analysiert wurden, werden im folgenden Kapitel die drei Visualisierungstechniken vorgestellt. Darüber hinaus werden Interaktionsmöglichkeiten beschrieben.

3.3.1 Visualisierung Eins

In der ersten Visualisierung ist der Vergleich von zwei Attributen der Lebensmittel in einem Scatterplot dargestellt. Die Koordinatenachsen sind mit den ausgewählten Attributen beschriftet und passen sich in ihrer Skalierung beim Wechsel der Attribute automatisch an. Die Auswahl

der Attribute wird durch die Anwender:innen per Button festgelegt und kann für die x- und die y-Achse erfolgen. Jedes Lebensmittel wird im Scatterplot durch einen Punkt visualisiert. Bewegen Anwender:innen die Maus über einen Punkt, wird dieser farbig und ein Text erscheint. Dieser Text setzt sich aus Lebensmittelname und den beiden ausgewählten Nährstoffen zusammen. Der Scatterplot ist in Abbildung 1 dargestellt.

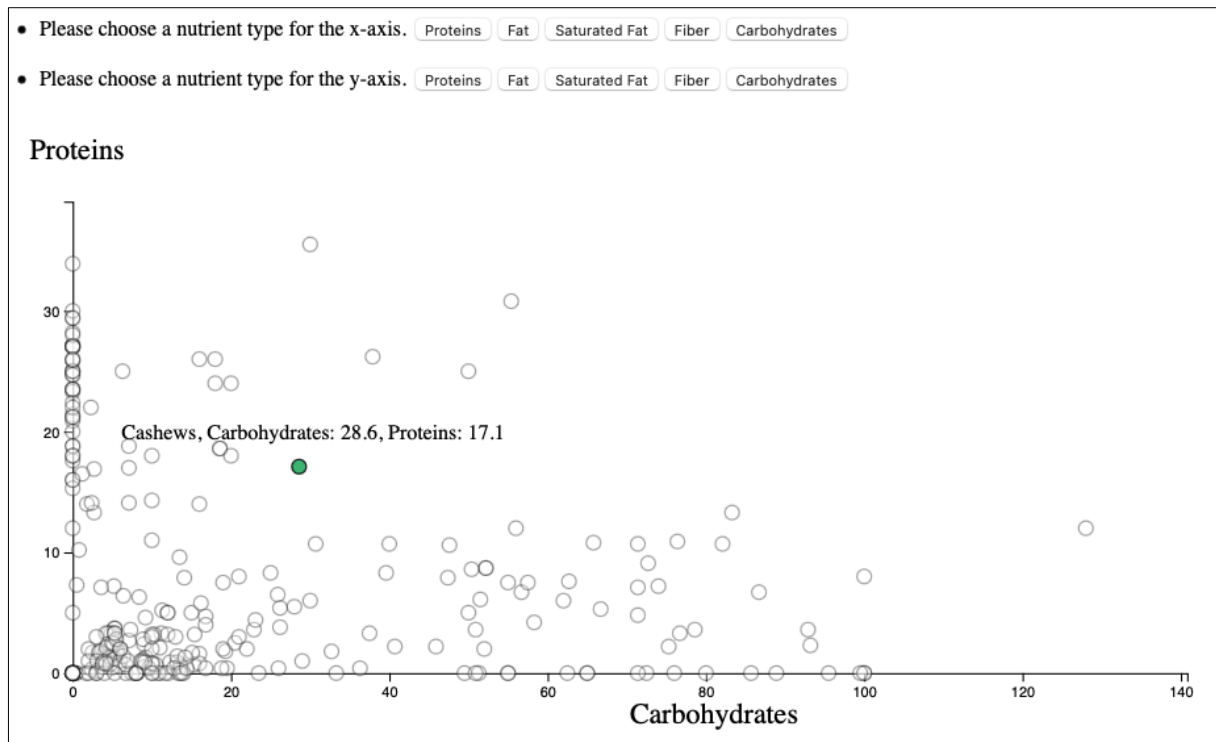


Abbildung 1: Scatterplot (Quelle: eigene Darstellung)

Die bestehenden Anforderungen an den Scatterplot konnten durch die beschriebenen Merkmale erfüllt werden, indem die Gegenüberstellung von zwei Nährstoffen für verschiedene Lebensmittel in übersichtlicher Weise visualisiert wurden. Darüber hinaus kann die Darstellung durch Anwender:innen dynamisch angepasst werden.

Der Scatterplot erweist sich im Vergleich zu alternativen Darstellungsmöglichkeiten zweidimensionaler Zusammenhänge als am besten geeignet. Um zwei Attribute ins Verhältnis zu setzen, ist eine Darstellung als Liniendiagramm denkbar, die im Fall der vorliegenden Daten allerdings nicht gewählt wurde. Dies begründet sich darin, dass die Punkte im Liniendiagramm durch Linien verbunden sind, was im Fall der dargestellten Daten als teilweise überlagerte Punkte unübersichtlich wäre. Durch den Scatterplot ist es möglich, die Datenpunkte voneinander zu differenzieren.

3.3.2 Visualisierung Zwei

Für den Vergleich von mehr als zwei Attributen eignet sich die Visualisierungstechnik Parallele Koordinaten. Dabei werden vier Attribute von jeweils einer Achse repräsentiert. Die Achsen verlaufen vertikal, liegen mit Abstand parallel nebeneinander und sind mit den ausgewählten Attributen beschriftet. Darüber hinaus erfolgt die Anpassung ihrer Skalierung beim Wechsel der Attribute automatisch. Anwender:innen können für jede der vier Achsen mithilfe von verschiedenen Buttons ein Attribut individuell auswählen. Die Lebensmittel werden durch Linien dargestellt, welche die zugehörigen Datenwerte auf den Achsen schneiden. Bewegen Anwender:innen ihre Maus über eine Linie, wird diese farbig und ein Text erscheint. Dieser Text zeigt den Lebensmittelnamen und die vier ausgewählten Nährwerte an. Die Parallelen Koordinaten sind in Abbildung 2 dargestellt.

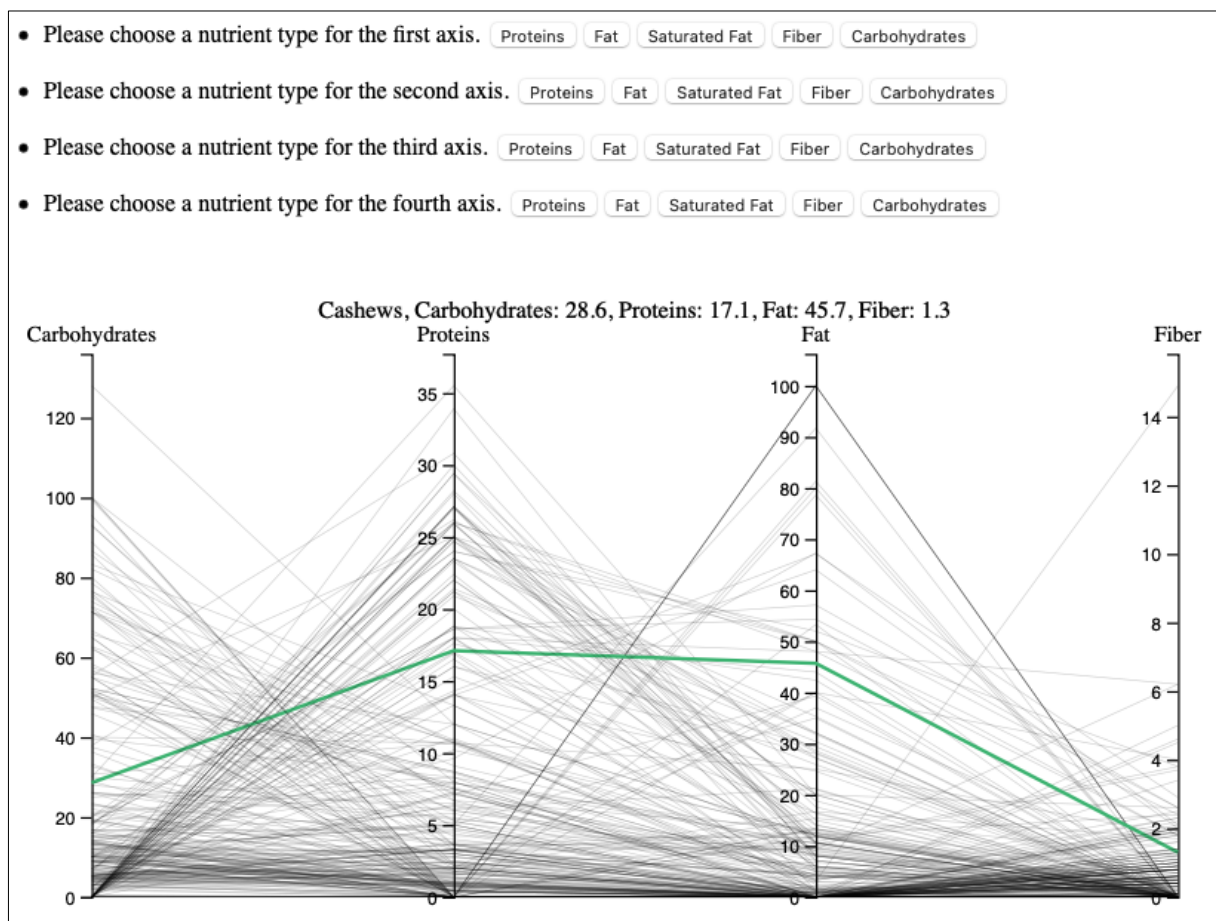


Abbildung 2: Parallele Koordinaten (Quelle: eigene Darstellung)

Da Parallele Koordinaten den Vergleich von mehreren Nährstoffen in Lebensmitteln ermöglichen, konnten die Anforderungen an die Visualisierung erfüllt werden. Dabei können die vier Attribute mithilfe von Buttons ausgewählt werden. Die Lebensmittel werden durch beschriftete und farbige Linien dargestellt, wenn sie mit der Maus ausgewählt werden.

Als alternative Darstellungsformen zu den Parallelen Koordinaten kommen Datentinte, Projektion oder Selektion in Frage, da sie sich zur Visualisierung mehrdimensionaler Zusammenhänge eignen. Allerdings werden diese Visualisierungstechniken im Vergleich als unübersichtlich und weniger intuitiv eingeschätzt, was zur Fehlinterpretation der Daten führen könnte.

3.3.3 Visualisierung Drei

Die dritte Visualisierung erfolgt in Form eines Baumdiagramms. Neben den Nährstoffdaten gehen aus den Datensätzen noch weitere Informationen hervor, die in den beiden vorhergehenden Visualisierungen nicht benötigt wurden. Diese sollen in der dritten Visualisierung Anwendung finden. Durch die Baumdarstellung wird eine hierarchische Struktur der Lebensmittelkategorien mit den zugehörigen Lebensmitteln abgebildet. Dafür sind beschriftete Kreise hierarchisch angeordnet und mit Linien verbunden. Die vorletzte und letzte Stufe liefert Erkenntnisse über die Anzahl der Kalorien in den einzelnen Lebensmitteln. Die Lebensmittel sind innerhalb ihrer Gruppe zudem aufsteigend sortiert. Ein Ausschnitt aus der Baumhierarchie ist in Abbildung 3 dargestellt.

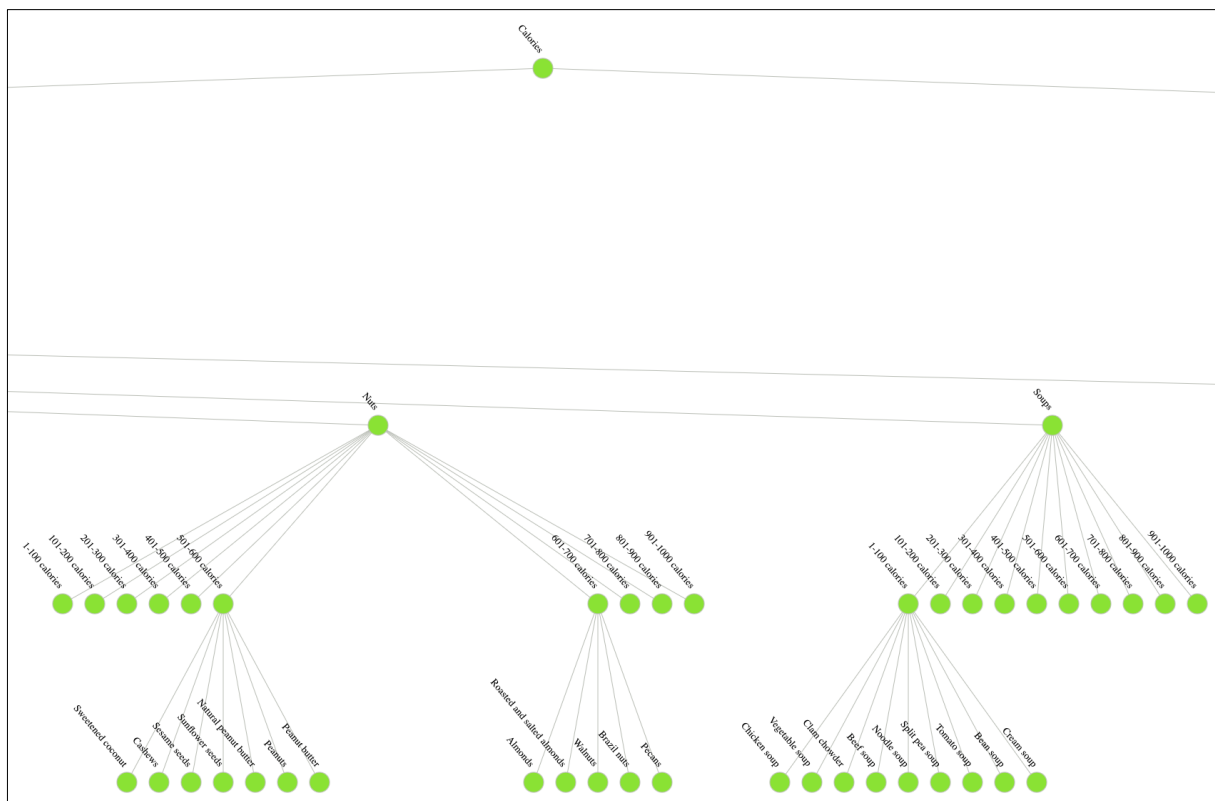


Abbildung 3: Ausschnitt aus der Baumhierarchie (Quelle: eigene Darstellung)

Die Anforderungen an das Baumdiagramm konnten nur teilweise erfüllt werden. Zwar lassen sich dadurch Lebensmittel in Kategorien einordnen und Anwender:innen erhalten einen groben Überblick darüber, welche Lebensmittel in welchen Lebensmittelkategorien eher kalorienarm

oder kalorienreich sind. Da die Einteilung der Kategorien für die Kalorien sehr grob gewählt wurde, können keine konkreten Aussagen über die tatsächlichen Kalorienwerte getroffen werden. Es ist lediglich eine Sortierung erkennbar, da die Lebensmittel sich innerhalb eines Rahmens von 100 Kalorien befinden und auf der letzten Stufe von links nach rechts aufsteigend sortiert sind. Auch die Gesamtübersicht geht in dieser Visualisierung etwas verloren, da ein aktives Scrollen zur Seite nötig ist, um bestimmte Informationen zu finden oder alle Daten anzuschauen.

Andere Visualisierungstechniken, wie beispielsweise das Sunburst-Diagramm, eignen sich ebenfalls zur Darstellung von Hierarchien. Dieses bietet zwar den Vorteil einer ansprechenden Optik, hat jedoch den Nachteil, dass die Kindelemente ab einer gewissen Stufe kaum noch lesbar sind. Dieses Problem besteht beim Baumdiagramm nicht. Es ist allerdings zu überlegen, ob die Informationen aus der Baumhierarchie zusätzlich in die anderen beiden Visualisierungen integriert werden könnten.

3.4 Interaktion

In den Visualisierungen sind verschiedene Interaktionsmöglichkeiten enthalten. Dadurch sollen Anwender:innen die Möglichkeit erhalten, die Darstellungen so zu verändern, dass sie relevante Informationen schneller finden.

Scatterplot und Parallele Koordinaten enthalten Buttons, durch die die Attribute der Achsen ausgewählt werden können. Dadurch ist es möglich, dass gezielt relevante Informationen angezeigt werden, während andere aus der Darstellung ausgeblendet werden. Außerdem besteht für die Anwender:innen die Möglichkeit, die Maus über die eingezeichneten Punkte im Scatterplot zu bewegen und so die Lebensmittelnamen und ausgewählte Nährwerte anzuzeigen. Der ausgewählte Punkt wird farbig hervorgehoben und der entsprechende Text erscheint. Die Interaktionsmöglichkeit für Parallele Koordinaten ist ähnlich, wobei in diesem Fall Linien die Lebensmittelnamen und vier ausgewählte Nährwerte repräsentieren, wenn Anwender:innen die Maus über die Darstellung bewegen. Auch hier wird die ausgewählte Linie farbig und der Text wird angezeigt. Die Baumhierarchie bietet keine Interaktionsmöglichkeit, da die Datengrundlage in Form einer statischen JSON-Datei vorliegt.

Die Startseite *index.html* im Ordner *docs* ermöglicht die Darstellung der Visualisierungen als Webseite. Es ist somit möglich, auf der Webseite zwischen den Visualisierungen hin- und herzuwechseln, um anwendungsübergreifende Zusammenhänge zu untersuchen. Die drei HTML-Seiten der Visualisierungen wurden mittels des Terminalbefehls *elm make* in Visual Studio Code erzeugt und befinden sich ebenfalls im Ordner *docs*.

4. Implementierung

Die Funktionalität der Visualisierungen basiert auf den Codes der Übungen des Moduls „Information Retrieval und Visualisierung“. Dabei zählen insbesondere Übung 1 und Übung 3 als Grundlage für den Code des Scatterplots. Für die Parallelen Koordinaten wurde Übung 7 modifiziert. Weiterhin wurden Teile aus Übung 8 verwendet, um den CSV-Decoder für Scatterplot und Parallele Koordinaten zu erstellen. Als Grundgerüst für die Baumhierarchie diente die Übung 10.

Die Aufteilung der Codes wurde weitestgehend einheitlich strukturiert, damit Anwender:innen schneller einen Überblick gewinnen und sich besser zurechtfinden können. Alle drei Codes beinhalten zunächst die benötigten importierten Elm-Module. Es folgt die Funktion *main* sowie die Definitionen von *type* und *type alias*. Weiterhin erfolgt die Initialisierung durch die Funktion *init* und die CSV-Daten werden geladen und mittels des CSV-Decoders *decodingNutrients* decodiert. Anschließend ermöglicht die Funktion *update* den dynamischen Wechsel der Attribute an den Achsen. Darüber hinaus schließen sich *subscriptions*, visualisierungsspezifische Designanpassungen und die Sektion *view* an.

Übung 1 diente als Grundlage zur Darstellung der Lebensmittel im Scatterplot. Zudem konnten die Funktionen zum Laden der CSV-Datei übernommen werden. Der CSV-Decoder erforderte allerdings einige Anpassungen, da möglichst viele Informationen aus dem Datensatz visualisiert werden sollten und wurde entsprechend auf mehrere Werte erweitert. Außerdem musste der *type Msg* angepasst werden. Die Funktion *update* wurde basierend auf Übung 3 und Übung 8 aufgebaut, wobei sie festlegt, welche Attribute zuerst ausgewählt sind und was bei der Bedienung eines Buttons passiert. Eine große Herausforderung der ersten Visualisierung bestand in der Verarbeitung der Daten, insbesondere beim Export von Excel zu CSV, da die exportierten Werte zunächst durch ein Semikolon getrennt sind. Für die erfolgreiche Weiterverarbeitung wird allerdings als Trennzeichen ein Komma benötigt. Auch die Einstellung des Dezimaltrennzeichens als Punkt anstelle eines Kommas ist zu beachten. Außerdem bestanden einige Schwierigkeiten bei der Verknüpfung der Daten mit den Buttons, welche jedoch durch den Einsatz genannter Übungen gelöst werden konnten. Der Code des Scatterplots wird über das GitHub Repository im Ordner *src/Develop* unter dem Namen *ELMScatterplot.elm* bereitgestellt.

Übung 7 wurde für die Visualisierung der Lebensmittel in den Parallelen Koordinaten verwendet. Zudem konnte der Code zum Laden der CSV-Datei aus dem Scatterplot übernommen werden. Der Code der Parallelen Koordinaten wird über das GitHub Repository im Ordner *src/Develop* unter dem Namen *ElmParalleleKoordinaten.elm* bereitgestellt.

Übung 10 konnte nahezu vollständig bei der Umsetzung der Baumhierarchie einfließen. Es ist lediglich eine Anpassung des Pfads der JSON-Datei notwendig. Die Datenbereitstellung für die Baumdarstellung ist allerdings deutlich aufwendiger. Die CSV-Datei musste dafür in ein JSON-Format umgewandelt werden, in dem zusätzlich eine Hierarchie zu berücksichtigen ist. Dies musste im Rahmen des Projekts manuell mit Visual Studio Code umgesetzt werden. Der Code der Baumhierarchie wird über das GitHub Repository im Ordner *src/Develop* unter dem Namen *ElmBaumhierarchie.elm* bereitgestellt.

Nach der Fertigstellung drei funktionsfähiger Codes erfolgten designtechnische Anpassungen in Bezug auf Diagramm-, Punkt-, Linien- und Schrifteigenschaften.

5. Anwendungsfälle

Die Visualisierungen ermöglichen Erkenntnisse über Lebensmittel und deren Nährstoffe und Kalorien. Im folgenden Kapitel wird für die drei Visualisierungen jeweils ein konkretes Anwendungsszenario vorgestellt.

5.1 Anwendung Visualisierung Eins

Die Visualisierung des Scatterplots in Abbildung 4 zeigt den Vergleich der beiden Attribute Proteine (*Proteins*) und Kohlenhydrate (*Carbohydrates*). Die Anwendung der Visualisierung wird anhand einer Lebensmittelkonsumentin dargestellt, die sich vegan ernährt.

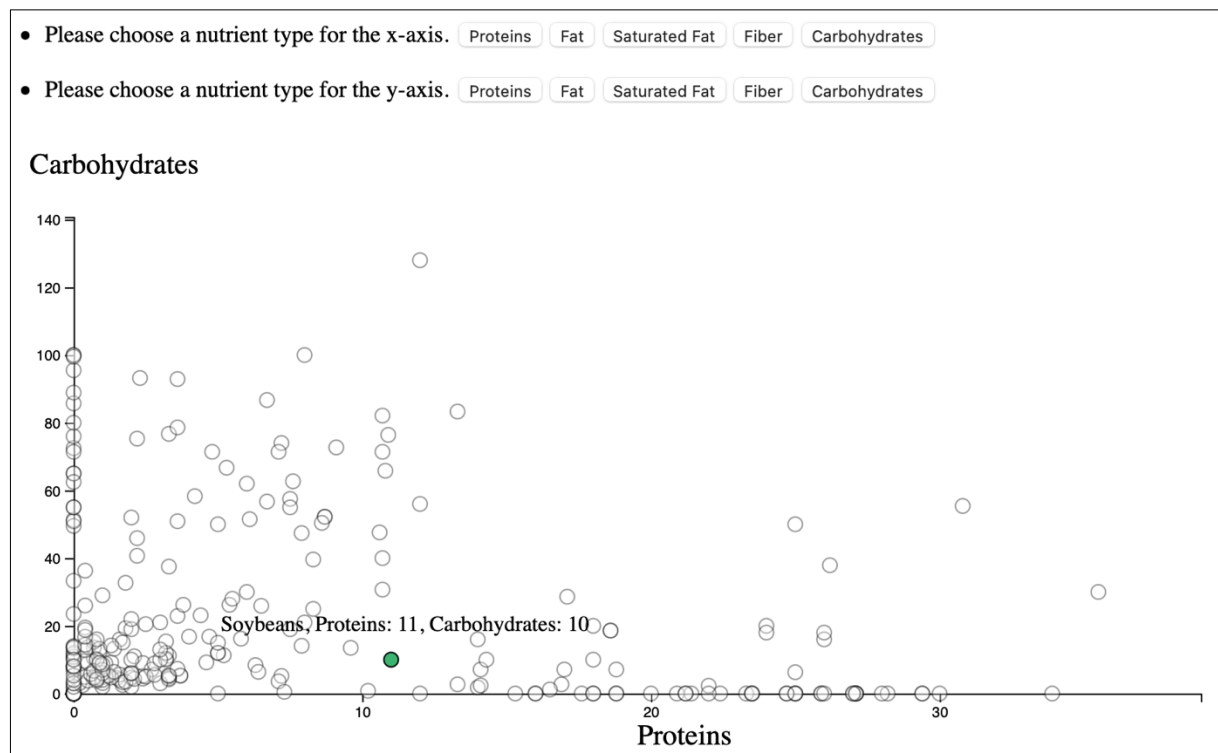


Abbildung 4: Vergleich von Proteinen und Kohlenhydraten im Scatterplot (Quelle: eigene Darstellung)

Die Konsumentin möchte fleischlose Burger zubereiten. Weil sie sich vegan ernährt, muss sie besonders darauf achten, dass sie ausreichend Proteine aufnimmt. Sie möchte aber auch, dass die Basis für ihren Burger nicht übermäßig viele Kohlenhydrate enthält. Sie wählt für die x-Achse das Attribut *Proteins* und für die y-Achse das Attribut *Carbohydrates*.

Bei der Betrachtung von Abbildung 4 fällt auf, dass einige Lebensmittel pro 100 Gramm sehr viele Proteine enthalten. Beim Prüfen dieser Punkte durch die Auswahl mit der Maus stellt die Lebensmittelkonsumentin fest, dass es sich bei Produkten mit sehr hohem Proteingehalt und niedrigem Kohlenhydratgehalt entweder um Fleisch oder Fisch handelt. Grundsätzlich ist daraus abzuleiten, dass tierische Produkte einen hohen Proteingehalt besitzen. Dem gegenüber steht ein sehr geringer Kohlenhydratgehalt – die meisten Produkte dieser Kategorie enthalten gar keine Kohlenhydrate. Bei ihrer Überprüfung weiterer Punkte mit vielen Proteinen lernt die Konsumentin außerdem, dass Nüsse im Allgemeinen sehr proteinreich sind. Diese eignen sich allerdings nicht für ihr Vorhaben. Schließlich entscheidet sie sich dafür, Sojabohnen (*Soybeans*) zu verwenden. Diese sind in Abbildung 4 an dem farbigen (grünen) Punkt mit Beschriftung erkennbar. Sojabohnen beinhalten pro 100 Gramm 11 Gramm Proteine (*Proteins: 11*) und 10 Gramm Kohlenhydrate (*Carbohydrates: 10*) und eignen sich entsprechend ihrer Anforderungen im Vergleich zu den Alternativen der umliegenden Punkte am besten.

5.2 Anwendung Visualisierung Zwei

Die Visualisierung der Parallelen Koordinaten in Abbildung 5 zeigt den Vergleich der vier Attribute Kohlenhydrate (*Carbohydrates*), Proteine (*Proteins*), Fette (*Fat*) und Ballaststoffe (*Fiber*). Die Anwendung der Visualisierung wird anhand eines interessierten Verbrauchers dargestellt, der sich über die Nährwerte von Nusssorten informieren und diese vergleichen möchte. Er hat bereits herausgefunden, dass Nüsse ein guter Ballaststofflieferant sind. Nun möchte er sich weitere Nährwerte ansehen und verschiedene Nusssorten anhand ihrer Nährwerte vergleichen. Er entscheidet sich dafür, die Attribute Kohlenhydrate (*Carbohydrates*), Proteine (*Proteins*), Fette (*Fat*) und Ballaststoffe (*Fiber*) auszuwählen. Auf den ersten Blick sind keine Muster in den Daten erkennbar.

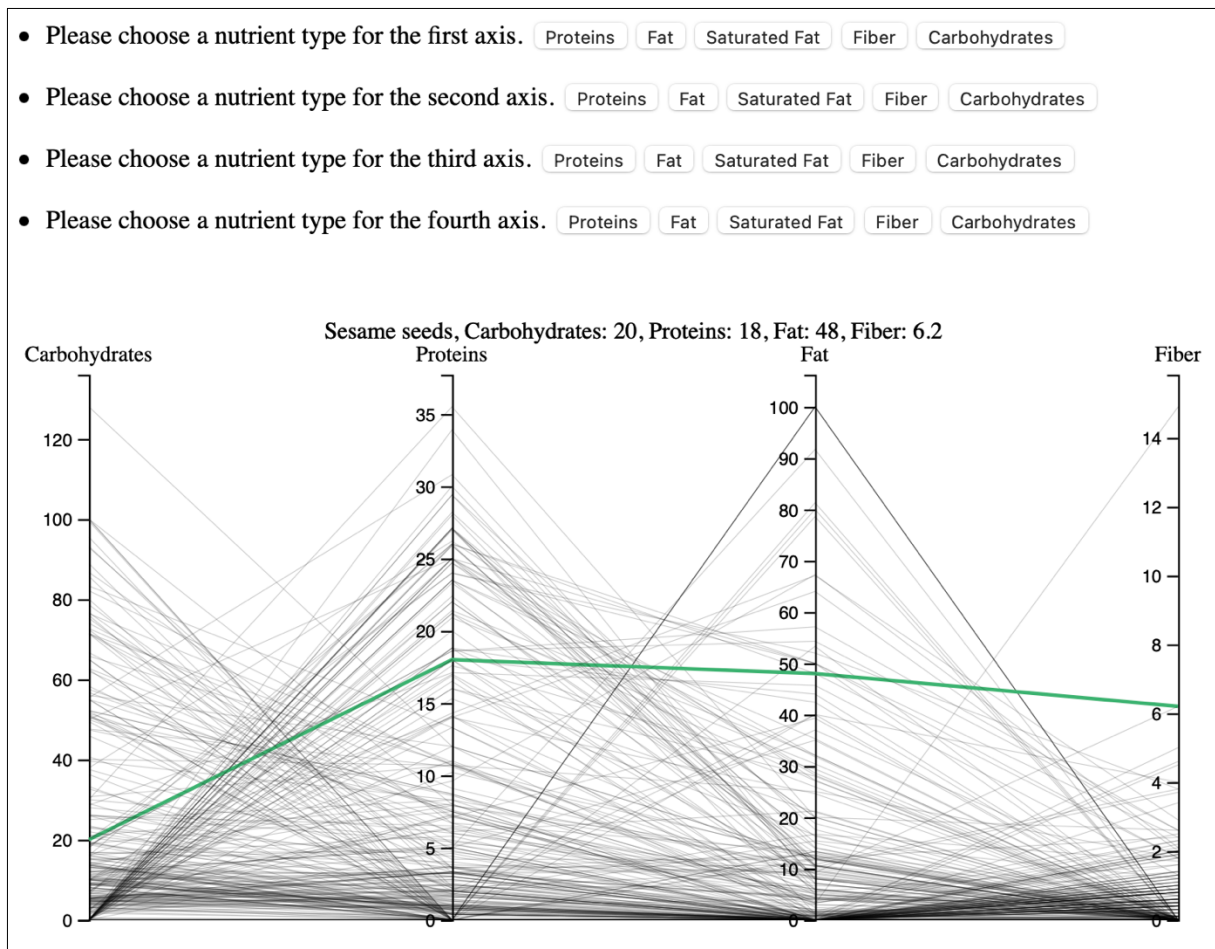


Abbildung 5: Vergleich der Nährstoffgehalte in Parallelen Koordinaten (Quelle: eigene Darstellung)

Da er bereits weiß, dass Nüsse einen hohen Ballaststoffgehalt haben, sucht er entlang der Achse Ballaststoffe (*Fiber*) von oben nach unten. Schnell findet er heraus, dass Sesam (*Sesame seeds*) mit 6,2 Gramm (*Fiber: 6.2*) den höchsten Ballaststoffgehalt pro 100 Gramm beinhaltet. Darüber hinaus kann er auf den ersten Blick erkennen, dass Sesam 20 Gramm Kohlenhydrate (*Carbohydrates: 20*), 18 Gramm Proteine (*Proteins: 18*) und 48 Gramm Fett (*Fat: 48*) pro 100 Gramm enthält. Er vergleicht die verschiedenen Produkte und sieht für Nüsse allgemein bestätigt, dass diese neben einem hohen Ballaststoffgehalt auch einen vergleichsweise hohen Fettgehalt haben. Zwar ließ sich auf den ersten Blick kein Muster in der Darstellung erkennen, aber aus den Daten geht bei genauer Betrachtung hervor, dass ähnliche Produkte ähnliche Nährwerte aufweisen. So haben beispielsweise Öle und Fette einen hohen Fettgehalt, aber keine Kohlenhydrate, Proteine oder Ballaststoffe. Viele Proteine sind in Fleischprodukten enthalten, diese haben aber kaum Fett und keine Kohlenhydrate oder Ballaststoffe.

5.3 Anwendung Visualisierung Drei

Die Visualisierung der Baumhierarchie in Abbildung 6 zeigt die Lebensmittel der Kategorie Getreideprodukte (*Cereal Products*), geordnet nach Kalorien pro 100 Gramm. Die Anwendung der Visualisierung wird anhand eines Produktentwicklers dargestellt, der sich für eine kalorienarme Zusammensetzung seines neuen Produktes interessiert.

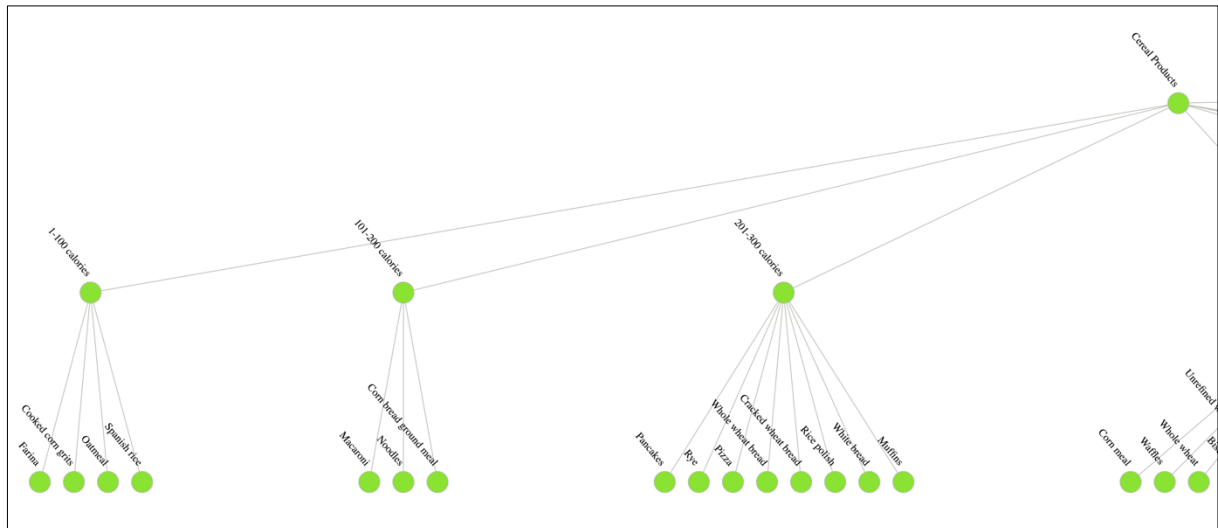


Abbildung 6: Lebensmittelkategorie Getreideprodukte in der Baumhierarchie (Quelle: eigene Darstellung)

Dabei sucht er zunächst eine Basis für einen Müsliriegel in der Kategorie Getreideprodukte (*Cereal Products*). Aus der Visualisierung entnimmt er, dass dafür Haferflocken (*Oatmeal*) in Frage kommen, da diese nur zwischen 1 und 100 Kalorien (*1-100 calories*) enthalten. Weiterhin möchte er eine geeignete Zutat aus der Kategorie Nüsse (*Nuts*) heraussuchen. In Abbildung 7 ist der Ausschnitt aus der Baumhierarchie mit dieser Kategorie dargestellt. Zunächst fällt auf, dass Nüsse allgemein verhältnismäßig viele Kalorien pro 100 Gram enthalten. Aufgrund seiner Anforderung, dass der Müsliriegel möglichst kalorienarm ist, entscheidet er sich für Cashewnüsse (*Cashews*). Außerdem möchte er gesüßte Kokosnuss (*Sweetened Coconut*) verarbeiten. Beide Zutaten beinhalten mit 501-600 Kalorien (*501-600 calories*) die geringste Kalorienzahl.

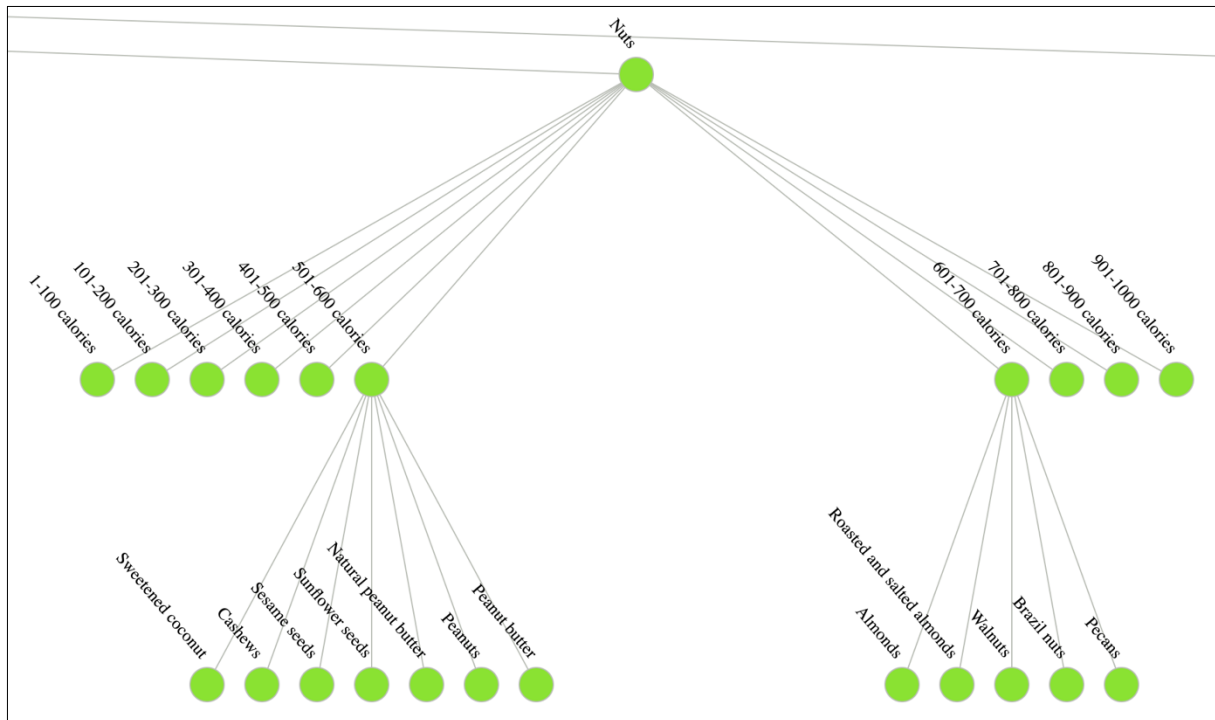


Abbildung 7: Lebensmittelkategorie Nüsse in der Baumhierarchie (Quelle: eigene Darstellung)

Anschließend schaut er sich die Kategorie Früchte (*Fruits*) an, um für seinen Müsliriegel die geeigneten Früchte auszuwählen. Diese Kategorie ist in Abbildung 8 dargestellt. Es könnten für ihn Bananen und Himbeeren in Frage kommen, da diese mit 1-100 Kalorien (*1-100 calories*) noch im unteren Bereich liegen.

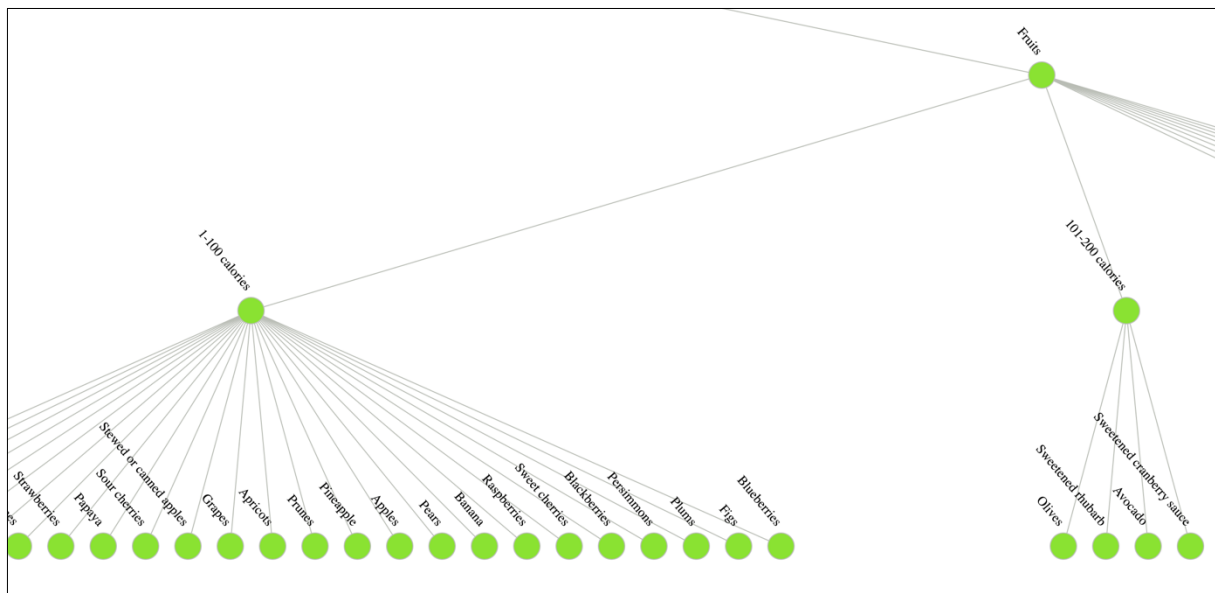


Abbildung 8: Lebensmittelkategorie Früchte in der Baumhierarchie (Quelle: eigene Darstellung)

Ein Produktentwickler würde sicherlich weitere Optionen in Betracht ziehen, um sich über die Kalorien der Lebensmittel zu informieren. Dennoch bietet die Baumhierarchie eine Möglichkeit, sich einen ersten Überblick zu verschaffen und die Lebensmittel direkt zu vergleichen.

Der Mehrwert der Visualisierungen entsteht vor allem dann, wenn alle drei Visualisierungen übergreifend betrachtet werden. So könnten interessierte oder ernährungsbewusste Lebensmittelkonsument:innen zusätzliche interessante Informationen erhalten, wenn sie sich auch die Baumhierarchie anschauen. Sie können beispielsweise feststellen, wie Nährstoffe den Kaloriengehalt beeinflussen. Im umgekehrten Fall könnten für den Produktentwickler nicht nur die Kalorien von Relevanz sein, sondern auch die Nährwerte, um ein wirklich qualitativ hochwertiges Produkt zu entwickeln.

6. Verwandte Arbeiten

Dieses Kapitel beinhaltet die Vorstellung ausgewählter Publikationen zum Thema Ernährung, in denen vergleichbare Visualisierungstechniken verwendet wurden. Es ist anzumerken, dass bislang wenige wissenschaftliche Visualisierungen mit ähnlichen Datensätzen und Anwendungsmöglichkeiten umgesetzt wurden.

Eine Forschungsarbeit befasst sich mit der Protokollierung der Nahrungsaufnahme von Patienten und mit der damit verbundenen Problemstellung, dass verschiedene Produkte unterschiedliche Inhaltsstoffe und Portionsgrößen aufweisen, die entweder gemessen oder geschätzt werden können. Da eine einzelne Abmessung aller Lebensmittel mühsam wäre, werden die Daten so modifiziert, dass sie einheitlich visuell geschätzt werden können. Dafür werden Lebensmittelgruppen auf der Grundlage der Energiedichte pro Volumen gebildet. Die Visualisierung basiert dabei auf einem Datensatz der USDA-NNDSR³ und die Kohlenhydrat- und Kalorienwerte wurden für jedes Lebensmittel berechnet. Die Darstellung erfolgt in einem Scatterplot mit den zwei Attributen *Kalorien pro Volumen* und *Kohlenhydrate pro Volumen*. Im Vergleich zur in Kapitel 3.1 beschriebenen Visualisierung bietet dieser Scatterplot den Vorteil, dass die Lebensmittelkategorien farbig differenziert dargestellt werden. Außerdem bildet die Darstellung mit 8618 Lebensmitteln einen wesentlich größeren Überblick ab. Allerdings ist es nicht möglich, zwischen verschiedenen Attributen der Lebensmittel zu wechseln – es können lediglich Erkenntnisse über Kalorien und Kohlenhydrate gewonnen werden. Als Gemeinsamkeit ist die Interaktionsmöglichkeit zu nennen, die eine ähnliche Funktionalität bei der Anzeige der Lebensmittelinformationen aufweist. [7]

Im Rahmen der Recherche wurde außerdem eine Visualisierung der Nährstoffe in Parallelen Koordinaten ermittelt. Dafür liegt kein Bericht vor, sondern lediglich eine Darstellung auf einer

³ USDA National Nutrient Database for Standard Reference

Webseite, die nach einem ähnlichen Prinzip aufgebaut ist, wie die in Kapitel 3.2 beschriebene Visualisierung. Die interaktive Darstellung repräsentiert die Daten der USDA Nutrient Database. Sie beinhaltet mehrere statische Achsen mit erweiterten Nährstoffen und Kalorien, eine farbige Differenzierung der Lebensmittelkategorien und eine alphabetisch sortierte Liste aller inbegriffenen Lebensmittel mit ihren Nährwerten, die bei der Auswahl mit der Maus in den Parallelen Koordinaten hervorgehoben werden. In der optischen Umsetzung könnte diese Darstellung als Inspiration zur Weiterentwicklung des vorliegenden Projekts dienen. [8]

Nicht zu vergessen ist an dieser Stelle auch das Visualisierungsprojekt, welches den zugrunde liegenden Datensatz von der Plattform *Kaggle* erstmals visuell umsetzte. Die eingesetzten Visualisierungstechniken und der Anwendungsumfang unterscheiden sich jedoch stark von dem vorliegenden Projekt. Es werden vor allem farbige Kreisdiagramme verwendet, um Informationen abzubilden, aber auch Säulen- und Portfolio-Diagramme sowie 3D-Scatterplots. Ursprünglich wurden vor allem die einzelnen Lebensmittel oder Lebensmittelgruppen betrachtet, nicht aber ihre Gesamtheit, sodass ein Vergleich aller Lebensmittel bzw. Nährwerte nicht möglich ist. Zwar ist es möglich, den Visualisierungen sehr spezifische Informationen zu entnehmen (zum Beispiel *Top 10 protein rich foods*) und mit ihnen zu interagieren, um beispielsweise im 3D-Scatterplot zu zoomen. Allerdings wird die Komplexität der meisten Darstellungen mit Interaktionsmöglichkeit als recht hoch eingeschätzt, sodass es ohne Vorwissen kaum möglich ist, sich intuitiv zurecht zu finden. [6]

7. Zusammenfassung und Ausblick

Die Visualisierungen ermöglichen eine veranschaulichte Darstellung der vorliegenden Lebensmitteldaten. Anwender:innen sind dadurch in der Lage, die Informationen zu Nährstoffen und Kalorien schneller zu erfassen und darüber hinaus intuitiv mit den Darstellungen zu interagieren.

Anhand des Scatterplots können Anwender:innen zwei verschiedene Nährwerte der Lebensmittel interaktiv vergleichen und sich die zugehörigen Daten anzeigen lassen. So lassen sich Zusammenhänge zwischen den Nährstoffen innerhalb eines Produkts sowie zwischen ähnlichen Produkten ermitteln. Die Visualisierung erleichtert es auch, Unterschiede zwischen Lebensmitteln zu untersuchen, wenn beispielsweise ein Produkt mit einem spezifischen (hohen oder niedrigen) Nährwert gesucht wird. Die Visualisierung der Parallelen Koordinaten bietet darüber hinaus die Möglichkeit, mehrere Nährwerte der Lebensmittel gegenüberzustellen, die ebenfalls interaktiv festgelegt werden können. Dadurch entsteht ein erweiterter Überblick sowie eine vergleichbare Gesamtbilanz verschiedener Lebensmittel. Die Baumhierarchie bildet hierarchisch

strukturierte Lebensmitteldaten ab, wobei insbesondere der Kaloriengehalt der einzelnen Lebensmittel aus der Darstellung hervorgeht. Durch die Visualisierung können Anwender:innen bestimmte Lebensmittel über Lebensmittelkategorien schnell finden und sich über die Kalorien informieren. Zusätzlich ist auf den ersten Blick erkennbar, welche Lebensmittelkategorien besonders kalorienreiche bzw. kalorienarme Produkte enthalten. Der größte Nutzen dieser Arbeit entsteht, wenn alle drei Visualisierungen in ihrem Zusammenhang untersucht werden.

Durch die Visualisierungen wird eine greifbare Darstellung der Daten erreicht. Dies bietet einen Vorteil gegenüber der tabellarischen Form des ursprünglichen Datensatzes. Dadurch können Informationen schneller erfasst werden und folglich können Anwender:innen schneller eine Konsum- oder Kaufentscheidung treffen.

Die Grenzen dieser Arbeit bestehen vor allem in den simplen, funktionalen Darstellungen, sodass eine Überarbeitung denkbar ist. So wäre es möglich, den Scatterplot und die Parallelen Koordinaten ausdrucksstärker zu gestalten, indem zum Beispiel die Punkte und Linien auch im nicht ausgewählten Zustand unterschiedliche Farben erhalten, um zwischen Nahrung und Getränken oder verschiedenen Lebensmittelkategorien differenzieren zu können. Diese Differenzierung könnte auch durch zusätzliche Filterfunktionen realisiert werden, um präzisere Suchergebnisse zu ermöglichen. Allgemein könnte der Datensatz durch weitere Lebensmittel ergänzt werden. Die Baumdarstellung könnte zudem um die exakte Kalorienanzahl erweitert werden. Zu überlegen ist auch, ob eine dynamische Darstellung von Ausschnitten aus der Baumhierarchie umsetzbar ist, um weniger relevante Inhalte herauszufiltern und nur den relevanten Ausschnitt anzuzeigen. Durch eine Modifizierung der Daten können zudem noch mehr Zielgruppen erreicht werden (beispielsweise durch die Einführung einer neuen Kategorie *vegan/vegetarisch* im Baumdiagramm).

Visualisierungen zu Lebensmitteldaten sind in dieser Form kaum zu finden. Abseits der Verbesserungsmöglichkeiten stellen die Visualisierungen einen Mehrwert gegenüber den ursprünglichen Daten dar und sie bieten Potenzial zur Weiterentwicklung.

Anhang: Git-Historie

- * 466c9e3 (HEAD -> main, origin/main, origin/HEAD) (Ordner aufräumen, Zwischenstände entfernen, 2021-12-20)
- * 26049ae (Upload des finalen Projektberichts,, 2021-12-20)
- * 32c9b43 (Korrektur der Kapitel, 2021-12-19)
- * b381596 (Fertigstellung Kapitel 6 und Überarbeitung der Einleitung, 2021-12-18)
- * 1533f32 (Upload Bericht Version 3; Einleitung, Kapitel 7, Formatierung, 2021-12-17)
- * e4db1f7 (Abbildungen speichern, 2021-12-17)
- * 060eea5 (README Aktualisierung, 2021-12-17)
- * f3c959d (README Korrektur, 2021-12-17)
- * c5b73a0 (Layout Update, 2021-12-17)
- * 6a7b193 (HTML Update, 2021-12-17)
- * fd4e2b8 (Geringfügige Anpassungen der Seiten, 2021-12-17)
- * c7d9348 (Update, 2021-12-16)
- * 7d61d77 (.html-Dateien in docs verschieben, 2021-12-16)
- * 63dbd51 (Update Index HTML, 2021-12-16)
- * fa0cca4 (Startseite anlegen, 2021-12-16)
- * 36fb181 (HTML aus ELM-Dateien erstellen, 2021-12-16)
- * 7924db3 (Upload Bericht Version 2, 2021-12-14)
- * 3c7d45b (Aktualisierung, 2021-12-04)
- * 7202d0b (Aktualisierung, 2021-12-04)
- * 024eee3 (elm-stuff/0.19.1 ausblenden, 2021-12-04)
- * 0db2390 (Pfade in README einsetzen, 2021-12-04)
- * 22bbef0 (Upload Bericht Version 1, 2021-12-04)
- * c13ba7a (Farbanpassungen, 2021-12-04)
- * ce45d82 (Farbanpassung, 2021-12-04)
- * b3221e0 (Anpassung der Code-Struktur, 2021-12-04)
- * 873d3f6 (Anpassung der Code-Struktur, 2021-12-04)
- * 40fd7ba (Anpassung der Code-Struktur, 2021-12-04)
- * 58355b1 (Umbenennen der Baumhierarchie, 2021-12-04)
- * ed7dc67 (Umbenennen der Baumhierarchie, 2021-12-04)
- * c1cdb5b (Pfad anpassen, 2021-12-04)
- * 8788ded (Umbenennung der Dateien und des Datenordners erfassen, 2021-12-04)

- * be1aa18 (Datenordner umbenennen, 2021-12-04)
- * d94397c (CSV-Dateien sortieren und umbenennen, 2021-12-04)
- * 523a989 (CSV-Dateien sortieren und umbenennen, 2021-12-04)
- * e1a7b45 (CSV-Dateien sortieren und umbenennen, 2021-12-04)
- * 7e38ce1 (Excel-Dateien sortieren und umbenennen, 2021-12-04)
- * 76fbab9 (CSV-Dateien sortieren, 2021-12-04)
- * 673bd6f (JSON Testversion entfernen, 2021-12-04)
- * 26f6515 (BaumTest entfernen, 2021-11-26)
- * 67e681c (Aktualisierung README.md, 2021-11-26)
- * 7a32bc2 (JSON Update, 2021-11-26)
- * 72ac3f3 (Layout Update, 2021-11-26)
- * 6fd892a (Anpassung der Baumhierarchie, 2021-11-26)
- * a9b33b9 (Parallele Koordinaten Update, 2021-11-26)
- * 4b22f3e (Baumhierarchie Update, 2021-11-26)
- * ebc844d (JSON Update, 2021-11-26)
- * 33a095d (TreeLayout, 2021-11-26)
- * b2140a6 (Update Baumhierarchie, 2021-11-26)
- * d4d01c5 (JSON Update, 2021-11-26)
- * 6a3939a (Anpassung der Achsen, 2021-11-26)
- * e9045c2 (Kalorien aus "Parallele Koordinaten" entfernen, 2021-11-26)
- * 83e5d12 (Anpassung Punkbezeichnung, 2021-11-26)
- * 6ea992f (Kalorien aus Nutrients entfernen, 2021-11-26)
- * c51b596 (Anpassung der Punkt und Achsenbezeichnung, 2021-11-26)
- * bb47d64 (Update CSV, 2021-11-26)
- * 579f7e1 (CSV einlesen, 2021-11-26)
- * 19abf31 (CSV einlesen, 2021-11-26)
- * 02f7e87 (CSV bearbeiten, 2021-11-26)
- * a50856f (Anpassung Baumhierarchie, 2021-11-25)
- * ffbaa36 (Anpassung JSON, 2021-11-25)
- * 72ce6d4 (Excel-Dateien entfernen, 2021-11-25)
- * 1f9fc22 (Aktualisierung, 2021-11-25)
- * a2f8fa0 (Anpassung der Codes, 2021-11-25)
- * 5e6fcff (Löschen der Testversion Scatterplot und Überarbeitung, 2021-11-25)
- * 3400834 (Grundgerüst für Parallele Koordinaten, 2021-11-25)

- * 44b3a15 (Bearbeitung Scatterplot, 2021-11-25)
- * a179fe9 (Bearbeitung Scatterplot, 2021-11-25)
- * 0e6b275 (Anpassung der Baumhierarchie, 2021-11-25)
- * 08f0b96 (Grundgerüst für Baumhierarchie erstellen, 2021-11-24)
- * 3b1500b (Finale Datenvorverarbeitung, 2021-11-24)
- * be19b5f (Erstellung JSON für Baumhierarchie, 2021-11-24)
- * 1f153eb (Test ELMScatterplot, 2021-11-12)
- * 933d812 (Packages Update, 2021-11-12)
- * b32f51 (Ordner für Teil 1 und 2 erstellen, 2021-11-08)
- * 4666a43 (Starten der Datenbearbeitung, 2021-11-06)
- * adcd02f (Strukturierung der Datenordner, 2021-11-06)
- * a473771 (Überschrift zur READ.md-Datei hinzufügen, 2021-11-06)
- * 0307470 (Erstellung von .gitignore und Ignorieren von elm-stuff, 2021-11-06)
- * f96bfa (CSV-Datei einlesen, 2021-08-18)
- * efe1d8d (Erstellung der Projektordnerstruktur, 2021-08-18)
- * bc0c2ef (Initial commit, 2021-08-18)

Literatur

- [1] Deutsche Gesellschaft für Ernährung e.V. (2021): <https://www.dge.de/wissenschaft/referenzwerte/?L=0>. Zugriffen: 08.12.2021.
- [2] Lebensmittelverband Deutschland (2021): <https://www.lebensmittelverband.de/de/lebensmittel/ernaehrung/energie-naehrstoff-bedarf>. Zugriffen: 08.12.2021.
- [3] Yi, M. (2019): A Complete Guide to Scatter Plots. <https://chartio.com/learn/charts/what-is-a-scatter-plot/>. Zugriffen: 15.12.2021.
- [4] Parallel Coordinates Plot. https://datavizcatalogue.com/methods/parallel_coordinates.html. Zugriffen: 15.12.2021.
- [5] What is a Tree Diagram? <https://t2informatik.de/en/smartpedia/tree-diagram/>. Zugriffen: 15.12.2021.
- [6] Pandit, N. (2020): Food Nutrition Analysis. <https://www.kaggle.com/niharika41298/food-nutrition-analysis-eda/comments>. Zugriffen: 08.12.2021.
- [7] Tiefengrabner, M. & Ginzinger, S. (2014): Nutrition Data Analysis and Visualization for Deterministic Food Categorization. <https://phaidra.fhstp.ac.at/open/o:127>. Zugriffen: 17.12.2021.
- [8] Nutrients Contents – Parallel Coordinates. <https://syntagmatic.github.io/parallel/>. Zugriffen: 17.12.2021.

Eidesstattliche Erklärung

Hiermit erkläre ich, Delia Storch, dass ich die vorliegende Arbeit eigenständig und ohne Nutzung unerlaubter Hilfsmittel bzw. ohne die Hilfe anderer Personen verfasst habe. Sämtliche Textstellen, die den Literaturquellen entnommen wurden, sind als solche gekennzeichnet. Die Arbeit liegt nicht in gleicher oder ähnlicher Form bei einer Prüfungsbehörde oder an anderer Stelle vor.

Halle (Saale), 20.12.2021

A handwritten signature in black ink, reading "Delia Storch". The script is cursive and elegant, with the first letters of the first and last names being capitalized and prominent.

Delia Storch