JOHANNES KEPLER
UNIVERSITÄT LINZ

Eingereicht von
David Zangerle, BSc

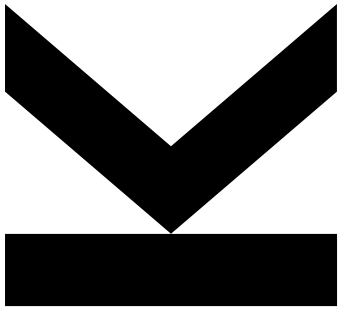
Angefertigt am
Institut für Wirtschaftsinformatik – Data & Knowledge Engineering

Beurteiler
Assoz.-Prof. Mag. Dr. Christoph Schütz

Mitbetreuung
Simon Staudinger, MSc
Dr. Emir Ganić

Monat Jahr
August 2024

Erhebung der aktuellen und geplanten Maßnahmen zur Reduzierung der Umwelteinflüsse des Flugverkehrs durch Flugsicherungsorganisationen



Masterarbeit

zur Erlangung des akademischen Grades

Master of Science

im Masterstudium

Wirtschaftsinformatik

JOHANNES KEPLER
UNIVERSITÄT LINZ
Altenberger Straße 69
4040 Linz, Österreich
jku.at

VORWORT

Die vorliegende Masterarbeit beschäftigt sich mit der Identifizierung der derzeit angewandten Maßnahmen europäischer Flugsicherungsorganisationen zur Reduktion negativer Umweltauswirkungen durch den Flugbetrieb, insbesondere in Bezug auf Luftqualität, Lärmentwicklung und Treibstoffverbrauch. Angesichts der zunehmenden globalen Herausforderungen im Umweltschutz und der Dringlichkeit, die Luftfahrt nachhaltiger zu gestalten, schien es mir wichtig, den aktuellen Stand der Maßnahmen im Flugbetrieb aus Sicht der europäischen Flugsicherungsorganisationen darzustellen.

Die Ergebnisse einer ersten Analyse dieses Sachverhalts, durchgeführt durch Internet- und Literaturrecherchen, wurden auf der EASN (European Aeronautics Science Network) Konferenz 2023 in Salerno, Italien, präsentiert. Diese Gelegenheit bot nicht nur die Möglichkeit, wertvolles Feedback aus der Fachwelt zu erhalten, sondern gab auch den Anstoß, tiefer in die Thematik einzutauchen und die Forschung weiterzuentwickeln.

Mein besonderer Dank gilt Herrn Dr. Emir Ganić, wissenschaftlicher Mitarbeiter am Institut für Verkehrs- und Transportwesen der Universität Belgrad, der mich während des gesamten Forschungsprozesses mit seinem umfassenden Fachwissen und seiner Unterstützung begleitet hat. Seine Expertise war entscheidend für die Ausarbeitung und Erstellung der Methodik sowie des verwendeten Fragebogens im Rahmen dieser Masterarbeit.

Ebenso möchte ich mich bei Dr. Christoph Schütz und Simon Staudinger, MSc, vom Institut für Data & Knowledge Engineering der Johannes Kepler Universität für ihre wertvolle Unterstützung und konstruktiven Rückmeldungen bedanken. Ihre Beiträge haben wesentlich zum Gelingen dieser Arbeit beigetragen.

KURZFASSUNG

In den letzten Jahren wurden erhebliche Anstrengungen unternommen, um die Lärmbelastung sowie den Kraftstoffverbrauch und die Schadstoffemissionen aus dem Flugzeugbetrieb zu reduzieren. Verschiedene Interessengruppen wie Flughäfen, Fluggesellschaften, Flugsicherungsorganisationen und Regulierungsbehörden gehen dieses Problem aus verschiedenen Perspektiven in ihrem Einflussbereich an. Die meisten Initiativen zur Reduzierung negativer Umweltauswirkungen erfordern erhebliche Ressourcen und konzentrieren sich hauptsächlich auf strategischer Ebene. Auf operativer Ebene können Flugsicherungsorganisationen (ANSPs) durch die Änderung von Betriebsabläufen an Flughäfen und die Verbesserung der Effizienz des Flugverkehrs sowohl auf dem Boden, als auch in der Luft kurzfristige Verbesserungen erzielen, die im Vergleich zu Lösungen anderer Interessengruppen weniger kostspielig umzusetzen sind. In diesem Zusammenhang wurde die Optimierung von Flughafenbetriebsverfahren mit dem Ziel, Lärmbelastung, Luftverschmutzung und Kraftstoffverbrauch zu reduzieren, in den letzten Jahrzehnten untersucht, und verschiedene Ansätze wurden ANSPs vorgeschlagen.

Daher untersucht diese Arbeit den aktuellen Stand der Systeme, Initiativen und Praktiken, denen ANSPs folgen, um die negativen Umweltauswirkungen des Luftverkehrs zu reduzieren. Um den aktuellen Stand zu ermitteln, wurden Textinformationen von den Webseiten der europäischen ANSPs gesammelt. Anschließend wurde der extrahierte Text systematisch vorbereitet und mithilfe verschiedener Text-Mining-Methoden analysiert. Das endgültige Ergebnis dieser Analyse ergab mehrere spezifische Listen für jeden der analysierten europäischen ANSP, welche die derzeit angewendeten Methoden und Techniken zur Reduzierung der negativen Umwelteinflüsse des Flugverkehrs zeigen. Alle derzeit angewendeten Methoden und Techniken wurden schließlich in einen gesamtheitlichen aktuellen Stand zusammengeführt und je nach Art in vier verschiedene Kategorien unterteilt.

Der nächste Schritt konzentriert sich auf die Überprüfung des abgeleiteten aktuellen Standes. Dazu wurde basierend auf den Text-Mining-Ergebnissen ein Fragebogen erstellt. Das angestrebte Ergebnis des Fragebogens besteht darin, von ANSP-Vertretern über die tatsächlich angewendeten Praktiken, Hindernisse bei der Implementierung, Motivationsfaktoren sowie zukünftige Absichten zur Einbeziehung umweltfreundlicher Praktiken in ihre Aktivitäten zu erfahren. Unter Berücksichtigung all dieser genannten Faktoren und Interessensgebiete wurde ein ganzheitlicher Überblick über den aktuellen Stand erarbeitet. Verschiedene ANSP-Vertreter in Europa wurden kontaktiert und gebeten, bei dieser Studie teilzunehmen und ihre Erkenntnisse im Fragebogen zu teilen. Als Gesamtergebnis dieses Schritts wurde der bestätigte aktuelle Stand der derzeit angewendeten Methoden und Techniken zur Reduzierung negativer Umweltauswirkungen des Luftverkehrs abgeleitet.

Eine Bewertung der Eignung des verwendeten Web-Scrapings- und Text-Mining-Prozesses zur Darstellung des tatsächlichen aktuellen Standes der angewandten Techniken schloss diese Studie ab. Dies wurde durch einen Vergleich der erwarteten Ergebnisse, die durch das Web-Scraping und Text Mining abgeleitet wurden, mit den tatsächlichen Ergebnissen von den ANSP-Vertretern durch die Umfrage, durchgeführt. Im Falle von Abweichungen zwischen den erwarteten und den tatsächlichen Ergebnissen wurde der Web-Scraping- und Text-Mining-Prozess sorgfältig überprüft, um etwaige Probleme zu identifizieren.

ABSTRACT

Significant efforts have been made over the past years to reduce noise impact as well as fuel consumption and pollutant emissions from aircraft operations. Different stakeholders such as airports, airlines, air navigation service providers, and regulatory bodies are tackling this problem from various perspectives in their sphere of influence. Most initiatives for the reduction of negative environmental impacts require substantial resources and are concentrated mainly on a strategic level. On a practical level, for example by changing operational procedures at airports and enhancing the efficiency of air traffic, both on ground and air level, Air Navigation Service Providers (ANSPs) can achieve short-term improvements, which are less expensive to implement compared to solutions at the disposal of other stakeholders. In this regard, the optimisation of airport operational procedures with the aim to reduce noise impact and fuel burn has been well studied over the past decades, and various approaches have been proposed to ANSPs. However, the implementation of possible theoretical solutions into ANSP practice proved not to be as widely spread as expected.

Therefore, in this paper, the state-of-the-art regarding current systems, initiatives and practices that ANSPs are following to reduce the environmental impact of air traffic, including the factors of local air quality, fuel consumption and noise emissions is investigated. In order to derive the current state-of-the-art, websites of the European ANSPs have been scraped to extract relevant information material. Subsequently, the extracted text information was systematically prepared and analysed by the usage of various text-mining-methods. The final outcome of this analysis resulted in several specific lists, for each of the scraped European ANSP, showing the currently applied methods and techniques for each of the analysed organisations. All of these currently applied methods and techniques were finally merged into the concluding current state, and allocated into four different categories depending on the type and expected result of usage.

The next step focuses on validation of the derived current state. In order to do that, the text mining results were integrated into a questionnaire. The pursued outcome of the questionnaire was to learn from ANSP representatives about the actual currently applied practices, barriers when implementing new or additional techniques, motivational factors, as well as future intentions to include environmental friendly practices in their activities. Considering all these mentioned factors and areas of interest, a comprehensive but holistic overview of the current state should be derived. Different ANSP representatives in Europe were contacted and asked to help with this important study and to share their insights in the questionnaire. As the overall result of this step, the confirmed and validated current state of currently applied methods and techniques in order to reduce negative environmental impacts of air traffic, was derived.

Commenting on the suitability of the used Web-Scraping and text mining process in order to reflect the true current state of applied techniques concluded this study. This was done by comparing the expected results which were derived by the usage of Web-Scraping and text mining, with the validated results obtained from the ANSP representatives. In case of discrepancies between the expected and the validated results, Web-Scraping and text mining process was carefully reviewed, starting from the end and working backwards, to identify any encountered issues.

INHALTSVERZEICHNIS

ABBILDUNGSVERZEICHNIS

TABELLENVERZEICHNIS

LISTINGS

1. Einleitung

Der Luftverkehr ist einer jener Sektoren, der in den Medien oft mit seinen negativen Auswirkungen in Verbindung für Umwelt und Klima gebracht wird (Europäisches Parlament, 2019). Es ist daher von entscheidender Bedeutung zu untersuchen, ob diese Darstellungen tatsächlich in diesem Ausmaß der Realität entsprechen, und falls ja, welche Maßnahmen dagegen ergriffen werden können und von wem. Mit den zahlreichen Vorteilen für den Personen- und Güterverkehr ist die Luftfahrt heutzutage aus unserem Alltag nicht mehr wegzudenken (Europäisches Parlament, 2019). Dennoch herrscht in der breiten Öffentlichkeit oft eine geringe Sensibilität für die Faktoren Umwelt und Klima in diesem Bereich (Europäisches Parlament, 2019).

Die verschiedenen Stakeholder und Akteure der Luftfahrtindustrie haben die Möglichkeit, die negativen Folgen des Luftverkehrs zu vermeiden oder zu verringern (CANSO, o. J.). In dieser Masterarbeit liegt der Fokus auf dem Stakeholder der Air Navigation Service Provider (ANSPs), oder zu Deutsch den Flugsicherungsorganisationen. Diese sind unter zahlreichen anderen Aufgaben für die Koordination, Planung und sichere Durchführung des Luftverkehrs sowohl am Boden als auch in der Luft verantwortlich (Austro Control, 2019). Aus dieser Rolle heraus lässt sich ableiten, dass sie über operative Methoden verfügen könnten, um den Luftverkehr umweltfreundlicher zu gestalten (Austro Control, 2019a).

Das Ziel dieser Arbeit ist es daher, den aktuellen Stand der europäischen ANSPs zu untersuchen und aufzuzeigen, welche Maßnahmen sie derzeit oder in Zukunft ergreifen, um die negativen Auswirkungen der Luftfahrtindustrie, insbesondere in den Bereichen Lärmbelastung, Luftqualität und Treibstoffverbrauch, zu vermeiden oder zu reduzieren.

1.1. Motivation

Die Luftfahrtindustrie spielt eine entscheidende Rolle in der globalen Wirtschaft, indem sie den weltweiten Handel, den Tourismus und die Mobilität von Menschen und Gütern auf schnelle und effiziente Weise, vor allem über lange Strecken hinweg, ermöglicht (EASA, 2022). Jedoch gehen mit dem Flugverkehr auch erhebliche negative Umweltauswirkungen einher, darunter die Emissionen von Treibhausgasen, Luftverschmutzung, Lärmbelastung und die Beeinträchtigung der natürlichen Lebensräume (EASA, 2022).

In Anbetracht der spürbar wachsenden Besorgnis über den Klimawandel und die Umweltbelastung, sowie auch der stetig wachsenden Luftfahrtindustrie, ist es von entscheidender Bedeutung, Wege zu finden, um die negativen Auswirkungen des Flugverkehrs zu verringern (Europäisches Parlament, 2019). Air Navigation Service Provider haben hierbei als einer der zahlreichen Stakeholder der Luftfahrtindustrie die Möglichkeit, fokussiert auf operationaler, aber auch auf strategischer Ebene, Einfluss zu nehmen (Jarošová & Pajdlhauser, 2022). Die Entwicklung, Implementierung und Anwendung umweltfreundlicher Praktiken und Technologien von Seiten der ANSPs in Europa ist daher ein wichtiger Schwerpunkt für Forschung und Innovation.

Diese Masterarbeit wird durch die Motivation angetrieben, einen Beitrag zur Erforschung und Bewertung der aktuellen Methoden und Systeme zur Reduzierung der Umweltauswirkungen des Flugver-

kehrs zu leisten. Durch die Analyse vorhandener und zukünftiger Ansätze sowie der Identifizierung von Herausforderungen und Motivationsfaktoren strebt diese Arbeit danach, Einblicke zu gewinnen, die dazu beitragen können, den Luftverkehr nachhaltiger zu gestalten sowie im Generellen die aktuelle Situation strukturiert darzustellen.

Darüber hinaus bietet die Integration von Web-Scraping als Datenerhebungsmethode die Möglichkeit, umfangreiche Daten aus verschiedenen Quellen zu sammeln und zu analysieren, um ein umfassendes Verständnis über den aktuellen Stand der Umweltschutzmaßnahmen im Luftverkehrssektor zu erlangen. Durch die Kombination von traditionellen Forschungsmethoden mit innovativen Ansätzen strebt diese Arbeit danach, neue Erkenntnisse zum Einsatz von Umweltschutzmaßnahmen im Bereich der europäischen ANSPs zu generieren.

1.2. Zielsetzung der Masterarbeit

Die Masterarbeit verfolgt jenes Ziel, den aktuellen Stand der angewandten Methoden, Systeme, Initiativen und Techniken abzuleiten, welche die europäischen ANSPs derzeit verwenden, um die negativen Einflüsse des Luftverkehrs auf die Umwelt zu vermeiden oder zu reduzieren. Dabei werden besonders die Einflüsse auf folgende drei Faktoren berücksichtigt beziehungsweise fokussiert. Bei diesen drei Faktoren handelt es sich um die Lärmentwicklung, den Kraftstoffverbrauch sowie auch die Luftqualität. Insgesamt ist das Herleiten und Aufzeigen des aktuellen Standes entscheidend, um eine solide Grundlage für die Entwicklung und Umsetzung effektiver Strategien zur Reduzierung der negativen Umweltauswirkungen des Luftverkehrs zu schaffen, und eine nachhaltigere Luftfahrtindustrie zu fördern. Darüber hinaus wird durch diese Arbeit das öffentliche Bewusstsein in dieser Thematik erhöht und Lücken und Herausforderungen sowie bewährte Methoden identifiziert. Dabei handelt es sich allesamt um Faktoren, welche eine informierte Entscheidungsfindung unterstützen können.

1.3. Forschungsfragen

Die Erforschung und Untersuchung der von Seiten europäischer ANSPs derzeit angewandten Methoden zur Reduzierung der umweltschädlichen Auswirkungen des Flugverkehrs in den Bereichen der Lärmentwicklung, Luftqualität sowie des Treibstoffverbrauchs, erfordert eine klare Definition der zu untersuchenden Fragen. Die folgenden Forschungsfragen dienen dazu, die verschiedenen Aspekte dieser Thematik zu beleuchten und den Rahmen für die folgenden Analysen zu setzen.

- *Welche konkreten Maßnahmen werden von ANSPs ergriffen bzw. geplant, um negative Umweltauswirkungen des Luftverkehrs zu vermeiden oder zu reduzieren?*
  Diese Frage zielt darauf ab, spezifische Handlungen, Praktiken oder Initiativen zum Vorschein zu bringen, die von den Luftsicherungsorganisationen ergriffen werden, um die Umweltbelastungen des Luftverkehrs zu reduzieren.
- *Wie relevant schätzen ANSP-Vertreter die vorhandenen Methoden zur Reduzierung der negativen Umweltauswirkungen des Luftverkehrs in Bezug auf Luftqualität, Lärmentwicklung sowie Treibstoffverbrauch ein?*
  Eine Bewertung der Effektivität der aktuellen Strategien und Methoden soll Aufschluss darüber geben, inwieweit die bestehenden Methoden den erwarteten Zielen gerecht werden.

- *Welche Herausforderungen und Barrieren bestehen nach Ansicht der ANSP-Vertreter bei der Implementierung umweltfreundlicher Praktiken im Luftverkehr?*
  Diese Frage zielt darauf ab, die Barrieren und Schwierigkeiten zu identifizieren, welche die Umsetzung und Implementierung von umweltfreundlichen Maßnahmen für die ANSPs behindern.

Diese Forschungsfragen stellen den Leitfaden für die Untersuchung und Analyse der Methoden zur Reduzierung der Umweltauswirkungen des Flugverkehrs dar, und bilden den Rahmen für die Erarbeitung von Erkenntnissen im Rahmen dieser Masterarbeit.

1.4. Struktur der Masterarbeit

Diese Masterarbeit beginnt damit, einen Überblick über die Grundlagen und Hintergründe zu schaffen, welche für das Verständnis des darauffolgenden praktischen Teils erforderlich sind. In diesem ersten Abschnitt werden daher alle relevanten Methoden, Techniken und Konzepte erklärt. Aus diesem Grund wird mit der Erklärung der Hintergründe begonnen. Dies beinhaltet eine Einführung in die Luftfahrtindustrie und deren Auswirkungen auf die Umwelt, eine Erklärung der Organisationen der Air Navigation Service Provider sowie abschließend die Darstellung der Möglichkeiten seitens der ANSPs, in die Thematik der negativen Umweltbelastungen des Luftverkehrs einzugreifen. Darauffolgend werden im Kapitel der Grundlagen die relevanten Techniken des Text Minings und Web-Scrapings vorgestellt. Nachdem die notwendigen Hintergründe und Grundlagen erläutert wurden, wird die Methodik der Masterarbeit näher beschrieben. In diesem Kapitel wird übersichtlich erklärt und dargestellt, wie die Zielsetzung der Masterarbeit erreicht wird. Darauffolgend geht es über in die Beschreibung und die Durchführung des praktischen Teils der Arbeit. Dies beginnt mit der Beschreibung des gesamten Prozesses des Web-Scrapings und des Text Minings. Übergreifend ausgedrückt also alle Schritte von der Datenerhebung bis hin zu der Datenanalyse, welche im Text Mining durchgeführt wurden. Daran anschließendes Kapitel fokussiert sich auf die Erstellung und Auswertung eines Fragebogens, welcher zum Informationsvergleich mit den Ergebnissen aus dem vorherigen Schritt dient. Zuletzt werden die erwarteten Ergebnisse aus dem Text Mining mit jenen aus dem Fragebogen verglichen und etwaige Diskrepanzen diskutiert und erklärt.

2. Hintergrund

Bevor im weiteren Verlauf auf die Methodik und die Analyse näher eingegangen wird, ist es zunächst wichtig den dafür benötigten Hintergrund vorzustellen. Ziel dieses Kapitels ist es daher den Hintergrund darzulegen, welcher für das Verständnis der Arbeit essentiell ist, sowie auch einen Kontext zu schaffen, um die Grundsituation zu verstehen. Dazu wird in diesem Abschnitt der Arbeit ein Überblick über die Umweltauswirkungen des Luftverkehrs gegeben. In diesem ersten Teil wird die Signifikanz des Luftverkehrs mit dessen Anteil am Klimawandel und Auswirkungen auf die Umwelt dargestellt. Damit wird in erster Linie geklärt, ob die Signifikanz des Luftverkehrs im Kontext Umweltbelastungen überhaupt gegeben ist, und es sich in Folge auch von Seiten der ANSPs überhaupt lohnt, spezifische Maßnahmen zum Umwelt- und Klimaschutz aufzunehmen und zu implementieren. Die Rolle der ANSPs mit einer generellen Definition, deren Aufgabenbereichen und Verantwortung wird im darauffolgenden Kapitel erläutert. Abschließend werden die möglichen Maßnahmen und Tätigkeiten aufgezeigt, welche von den ANSPs ergriffen werden können, um die negativen Auswirkungen des Luftverkehrs auf die Umwelt zu reduzieren oder zu vermeiden. Dieses abschließende Kapitel gibt einen generellen Überblick über zur Verfügung stehende Praktiken, und stellt gleichzeitig die gängigsten Maßnahmen, Systeme und Methoden dar, welche von den europäischen ANSPs angewendet werden.

2.1. Umweltbelastungen durch den Luftverkehr

Wie bereits vorher erwähnt, soll in diesem Kapital ein prägnanter aber holistischer Gesamtüberblick über den Flugverkehr mit Fokus auf die dadurch entstehenden Emissionen dargelegt werden. Neben einer generellen Einführung rund um das Thema der schädlichen Emissionen durch den Luftverkehr, wird ein weiteres Augenmerk auf aktuelle Zahlen und Fakten sowie zukünftige Entwicklungen gelegt.

Der Bundesverband der Deutschen Luftverkehrswirtschaft (BDL) (2023) fasst die negativen Auswirkungen des Luftverkehrs folgendermaßen zusammen. Der Luftverkehr verbrennt bekanntermaßen Kerosin und emittiert dadurch CO2 und andere umweltschädliche Emissionen, welche sich auf das Klima sowie auf andere Lebensbereiche auswirken (BDL, 2023). Für wie viel CO2-Emissionen der Luftverkehr verantwortlich ist, und was neben CO2 noch emittiert wird, wird im Verlauf noch näher erläutert. Laut dem Europäischen Parlament (2019) ist die Luftfahrt für etwa vier bis fünf Prozent an den globalen Treibhausgasemissionen verantwortlich. Obwohl diese Zahlen auf den ersten Eindruck als sehr niedrig wahrgenommen werden könnten, zählt die Flugindustrie zu jenen Bereichen mit den am stärksten zunehmenden Emissionsaufkommen, das zum Klimawandel beiträgt. Grund dafür ist hauptsächlich der überproportionale Anstieg der Passagierzahlen und des Handelsvolumen in den letzten Jahren innerhalb der EU und weltweit. Des Weiteren werden erst seit überschaubarer Zeit Anstrengungen unternommen und Technologien entwickelt, um den CO2-Ausstoß in dieser Branche zu verringern.

Abbildung 1 zeigt die Treibhausgasemissionen verursacht durch den Verkehr innerhalb der EU von 1990 bis hin zu prognostizierten Werten für das Jahr 2030. Deutlich zu sehen ist, dass der internationale Flugverkehr der Spitzenreiter bei der Emission von Treibhausgasen im Vergleich zu den restlichen dargestellten Verkehrsarten ist.

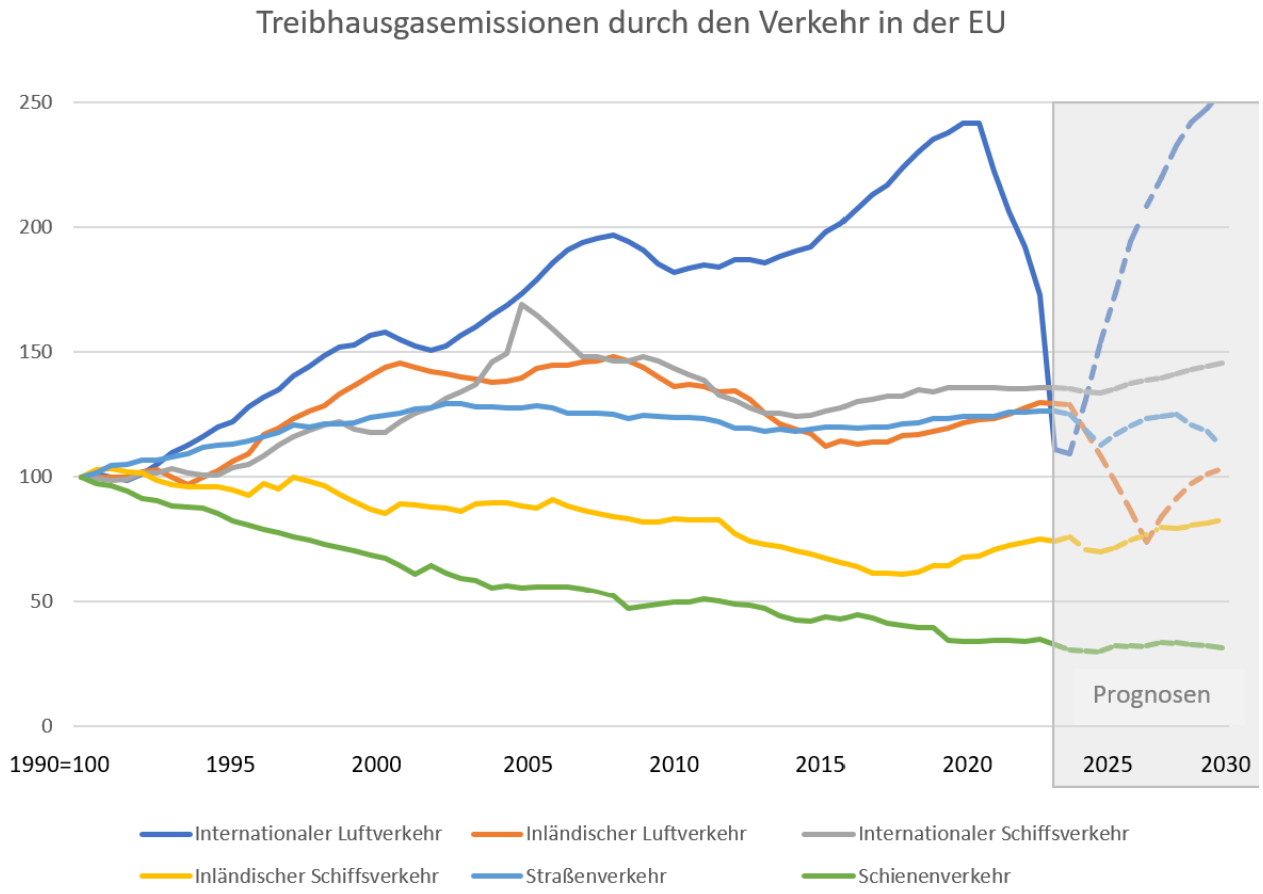


Abbildung 1: Treibhausgasemissionen durch Verkehr in der EU

Quelle: Eigene Darstellung in Anlehnung an Europäisches Parlament (2019)

Wie in Abbildung 1 sichtbar sind die Treibhausgasemissionen seit 1990 mit Ausnahme den Jahren 2020 und 2021 aufgrund der COVID-19 Pandemie, stetig angestiegen. Auch nach den COVID-19 Jahren wird der Ausstoß der Treibhausgasemissionen durch die Luftfahrt wieder an Fahrt aufnehmen und sehr rasch das Niveau vor der Pandemie erreichen. Prognostiziert wird im Jahr 2030 ein Ausstoß an Treibhausgasemissionen, verursacht durch den internationalen Luftverkehr, der über das 2,5 fache angestiegen sein wird im Vergleich zum Ausgangsjahr 1990.

Das Europäische Parlament (2019) bekräftigt hierbei, dass dieser Trend nicht nur beim Ausstoß der Treibhausgase zu sehen ist, sondern auch bei den Passagierzahlen im Flugverkehr innerhalb der EU. Genau wie die Treibhausgasemissionen, stieg auch diese Zahl seit 2010, mit Ausnahme der COVID-19 Jahre, stetig und stark an. Waren es im Jahr 2010 noch knapp 650 Millionen Passagiere, so hat sich diese Zahl 2019 bereits auf über 1 Milliarde erhöht mit der Tendenz weiter zu steigen. Der rapide Anstieg an Passagierinnen und Passagieren hat selbstredend eine Erhöhung der Flugfrequenzen, Steigerung der Anzahl an Flugzeugen in der Luft und damit auch den vermehrten Ausstoß von Treibhausgasemissionen zur Folge.

Aufgrund des soeben Gezeigten ist also in jedem Fall damit zu rechnen, dass die Passagierzahlen des Flugverkehrs und damit auch die CO2-Emissionen in den kommenden Jahren weiter in die Höhe klettern werden. Anzumerken ist, dass bis zum jetzigen Punkt dieser Arbeit nur die direkten CO2-Emissionen der Luftverkehrsindustrie behandelt wurden. Diese sind allerdings nicht die Einzigen

umweltschädlichen Stoffe und Gase, welche durch den nationalen und internationalen Luftverkehr freigesetzt werden (Europäisches Parlament, 2019). Neben den direkten CO2-Emissionen erzeugt der Flugverkehr auch einen wesentlichen Anteil an Nicht-CO2-Emissionen (Atmosfair, o. J.; BUND für Naturschutz und Umwelt Deutschland, o. J.). Die EASA (2022) sowie der BUND für Naturschutz und Umwelt Deutschland (o. J.) nennen beispielsweise folgende Nicht-CO2-Emissionen des Flugverkehrs, welche in die vorher dargestellten Zahlen noch nicht eingerechnet sind:

- Lärmentwicklung
- Minderung der Luftqualität
- Stickoxide oder Wasserdampf in hohen Luftschichten
- Ruß und Aerosole

Nicht-CO2-Emissionen haben teilweise genauso wie die direkten CO2-Emissionen einen negativen Einfluss auf das Klima und die Umwelt, andere Faktoren wie beispielsweise die Lärmentwicklung erzeugen negative Folgen für Anwohnerinnen und Anwohner (BUND für Naturschutz und Umwelt Deutschland, o. J.; EASA, 2022).

Die Auswirkungen dieser Nicht-CO2-Emissionen resultieren unter anderen in der Abnahme von Methan in der Luft, der Entwicklung von Kondensstreifen bis hin zur Sonnenabschirmung (Atmosfair, o. J.; BUND für Naturschutz und Umwelt Deutschland, o. J.). Eine Aufstellung der relevantesten Schadstoffe, erzeugt durch den Flugverkehr mit Ausnahme von CO2, sind in Tabelle 1 ersichtlich. Wie in dieser Tabelle erkenntlich ist, haben diese Nicht-CO2-Effekte eine dreimal höhere Klimawirkung als der direkte CO2-Ausstoß. Andere Studien, wie beispielsweise jene des BUND für Naturschutz und Umwelt Deutschland (o. J.), gehen von einer zwei- bis fünfmal stärkeren Klimawirkung durch die Nicht-CO2-Emissionen aus.

Tabelle 1: Klimawirkung von Flugreisen

Klimawirkung von Flugreisen Mehr als nur CO2							
Erzeugter Schadstoff	CO2	Stickoxide	Wasser-dampf	Ruß	Sulfat	Partikel	
Klimawirksamer Prozess	Direkt	Ozonbildung (+O3) Abnahme von Methan (-CH4)	Direkt	Direkt	Sonnen-abschirmung	Kondensstreifen Schleierwolken aus Eis	
Beitrag zur globalen Erderwärmung (im Verhältnis zu CO2)	1,0	0,33	0,04	0,02	-0,15	1,77	**Total 3,01**

Quelle: Atmosfair (o.J.)

Zusammenfassen lassen sich die schädlichen Auswirkungen der Luftfahrtindustrie also vor allem durch deren negativen Folgen für das Klima, die Umwelt aber auch auf Bürgerinnen und Bürger. Laut dem BUND für Naturschutz und Umwelt Deutschland (o. J.) verbrennt keine andere Art der Fortbe-

wegung pro Kopf so viel Energie wie das Fliegen. Die Klimawirkung beim Fliegen setzt sich einerseits aus den hohen direkten CO2-Emissionen, als auch aus anderen Faktoren zusammen, wie insbesondere Stickoxide und Wasserdampf in hohen Luftschichten. Derartige Faktoren werden unter dem Begriff der Nicht-CO2-Emissionen zusammengefasst. Diese Nicht-CO2-Emissionsen haben teilweise genauso wie direkte CO2-Emissionen einen Einfluss auf das Klima, teilweise sogar um ein vielfaches stärker. In den Zahlen und Fakten rund um die Auswirkungen des Flugverkehrs werden diese aber oft vernachlässigt.

Glücklicherweise gibt es laut der European Union Aviation Safety Agency (EASA) (2022) und auch weiteren Quellen aus der Literatur derzeit neben all diesen negativen Auswirkungen auf das Klima und die Umwelt des Flugverkehrs, auch zahlreiche Initiativen und neue Technologien, um die schädlichen Auswirkungen der Flugindustrie möglichst bald und auch rasch zu reduzieren. Bei diesen Initiativen handelt es sich meistens um Initiativen auf der strategischen Ebene, weshalb die Veränderung häufig von Seiten der Flugzeughersteller, Airlines oder Flughäfen erwartet wird. Durch die fortschreitende Produktion von effizienteren Triebwerkstypen ist es beispielsweise möglich die Lärmbelästigung durch den Flugverkehr nachhaltig zu senken. Darüber hinaus hat die Entwicklung effizienterer und nachhaltigerer Triebwerke natürlich auch positive Effekte auf die Energieeffizienz, Reichweite und Spritverbrauch. Eine weitere Initiative welche derzeit stark erforscht wird ist jene der Nachhaltigen Flugzeugkraftsoffen. Derzeit ist das Angebot an nachhaltigen Flugzeugkraftstoffen mit weniger als 0,05% des gesamten Flugkraftstoffverbrauchs in der EU sehr überschaubar. Darüber hinaus ist nachhaltiger Kraftstoff für Flugzeuge derzeit teurer als altbekanntes fossiles Kerosin. Kosteneinsparungen durch zukünftige Skaleneffekte werden hierbei allerdings erwartet. Nachhaltiger Flugzeugkraftstoff, welcher preiswert für die breite Masse in der Flugindustrie erworben werden kann, wird im Kontext des Umweltschutzes und des Klimawandels laut Literatur für enorme Einsparungen sorgen. Neben diesen beiden genannten Sektoren, welche derzeit verstärkt erforscht und entwickelt werden, gibt es natürlich noch zahlreiche weitere. Einige Beispiele hierfür wären marktbasierte Maßnahmen, Forschung und Entwicklung, ein grüner Flughafenbetrieb oder ähnliche.

Allerdings sind diese Entwicklungen und neuen Technologien von Seiten diverser Stakeholder zur Bekämpfung der negativen Umweltauswirkungen der Flugindustrie nicht der Fokus der Arbeit, weshalb diese auch an dieser Stelle nicht weiter vertieft werden. Weitere Maßnahmen, welche aber dennoch eine Rolle spielen sind unter dem Begriff des Flugverkehrsmanagement und -betrieb gesammelt. In diese Kategorie fallen auch die ANSPs. Die Rolle der ANSPs, deren Aufgaben und Verantwortungen werden im Laufe dieser Arbeit noch genauer beschrieben.

Nach diesem prägnanten Einblick in den Flugverkehr mit Fokus auf die Umweltauswirkungen und Emissionsausstoß ist nun geklärt, dass der Flugverkehr ein sehr bedeutsamer Faktor für den Klimawandel darstellt, dies obwohl das Fliegen auch heutzutage noch dem Großteil der Menschheit nicht zugänglich ist (EASA, 2022; Europäisches Parlament, 2019). In diesem Kapitel zusammengefasste Kernpunkte lassen sich folgendermaßen rekapitulieren. Zukünftiges Wachstum der Passagierzahlen im Flugverkehr wird den Anteil der Flugindustrie am globalen Klimawandel weiter erhöhen. Aufgrund dessen ist es umso wichtiger, dass in dieser Branche Initiativen zum Klimaschutz zeitnah geplant und implementiert werden. Initiativen und Maßnahmen, welche die negativen Umweltauswirkungen des Flugverkehrs eindämmen, sind hierbei gefragt. Neue Technologien und zukünftige Entwicklungen wie beispielsweise Nachhaltiger Flugkraftstoff, die Entwicklung effizienter Antriebs- und Turbi-

nentechniken oder andere Maßnahmen durch diverse Stakeholder des Flugsektors, helfen dieses Ziel zu erreichen. Derzeit sind diese innovativen Maßnahmen aber noch in deren Verfügbarkeit für die gesamte Flugindustrie stark beschränkt.

ANSPs haben allerdings die Möglichkeit durch operative Maßnahmen am Flugverkehr, welche größtenteils in deren Verfügbarkeit oder Finanzierbarkeit nicht von Drittparteien abhängig sind, und damit zeitnah umgesetzt werden könnten, einen Einfluss zu nehmen (Skyguide, 2023). ANSPs haben daher also die Möglichkeit den Flugsektor fokussiert auf operativer Ebene umweltfreundlicher zu gestalten (Skyguide, 2023).

2.2. Air Navigation Service Provider

Skyguide (2023) erklärt die Organisationen der Air Navigation Service Provider (ANSPs) oder zu Deutsch Flugsicherungsorganisationen, wie folgt. Sie sind grundsätzlich dafür zuständig, zu jedem Zeitpunkt einen sicheren und funktionierenden Flugverkehr zu gewährleisten. Dies gilt für alle Flugzeuge vom Start, über die gesamte Dauer des Fluges bis hin zur Landung. ANSPs sind dabei sowohl für die zivile als auch für die militärische Flugsicherung verantwortlich. Jede Flugsicherungsorganisation ist dabei für den Luftraum über einem gewissen Land oder definierten Bereich zuständig, sie können staatlich oder privat aufgebaut sein.

Laut zahlreichen Flugsicherungsorganisationen oder ANSPs wie beispielsweise der Deutschen Flugsicherung DFS (2023), dem Österreichischen ANSP Austro Control (2019) und dem Schweizer Flugsicherungsdienst Skyguide (2023) liegt der Fokus auf folgenden Bereichen und Hauptaufgaben:

- Tower
- Area Control Center/Kontrollzentrale
- Anflugkontrolle

Bezeichnungen von Aufgabenbereichen oder Funktionen können je nach Organisation variieren, die grundliegende Tätigkeit ist aber größtenteils ident (Austro Control, 2019; DFS, 2023; Skyguide, 2023). Die Abgrenzung der drei genannten Bereiche kann folgendermaßen durchgeführt werden.

Die ANSPs DFS (2023) und Skyguide (2023) erläutern den Aufgabenbereich des **Towers** folgendermaßen. Fluglotsinnen und Fluglotsen im Tower kontrollieren das gesamte Geschehen auf einem Flughafen. Dies beinhaltet die Überwachung und Kontrolle über das Rollfeld, die Start- und Landebahnen sowie auch des nahe umliegenden Luftraumes. Durch ihre Tätigkeit wird ein reibungsloser Ablauf des Flugverkehrs am Flughafen sowie im nahe angrenzenden Luftraum sichergestellt. Durch den Einsatz von Sprechfunk führen sie die Flugzeuge während des Rollens, Starts und Landens an und informieren die Pilotinnen und Piloten über die entsprechenden Verfahren für An- und Abflüge. Die Fluglotsinnen und Fluglotsen im Tower erteilen darüber hinaus die Freigabe zum Start.

Nach dem Start eines Flugzeuges und der vorhergehenden Kontrolle der Fluglotsinnen und Fluglotsen im Tower, geht der Verantwortungsbereich an das **Area Control Center(ACC)** oder zu Deutsch die Kontrollzentrale über (Austro Control, 2019). Die Mitarbeiter und Mitarbeiterinnen im ACC sind gemäß Austro Control (2019) und DFS (2023) für den gesamten Flugverkehr im entsprechenden Verantwortungsbereich, beispielsweise über eines bestimmten Bereiches oder Landes, verantwort-

lich. Dies beinhaltet die Verfolgung der Flugzeuge in der Luft auf Radarbildschirmen sowie der Sicherstellung, dass Kreuzungen in Flugrouten von verschiedenen Fluglinien sicher beflogen werden können. Dies wird erreicht mithilfe der Festlegung von Geschwindigkeiten und Flugrouten sowie Anweisungen zur Veränderung bestimmter Gegebenheiten wie beispielsweise der Flughöhe. Gesamt betrachtet überwachen die Fluglotsinnen und Fluglotsen im ACC den gesamten Luftverkehr in einem bestimmten Zuständigkeitsbereich. Die Kontrollzentrale spielt damit eine entscheidende Rolle bei der Gewährleistung der Sicherheit und Effizienz des Luftverkehrs durch ein ANSP.

Die dritte Hauptaufgabe stellt nach Austro Control (2019) und DFS (2023) die **Anflugkontrolle** dar. Nachdem ein Flugzeug die Reiseflughöhe verlassen hat, befindet es sich im Landeanflug. Die Fluglotsinnen und Flutlotsen in der Anflugkontrolle führen es am sichersten und im Optimalfall am schnellsten Weg zur Landebahn. In der Nähe eines Flughafens ist die Frequenz von startenden und landenden Flugzeugen in der Regel sehr hoch. Hinzu kommt, dass dies auch noch auf engem Raum passiert. Aufgrund dessen kreuzen sich Flugwege unterschiedlicher Flugzeuge häufig. Die Anflugkontrollstelle sorgt hierbei für mehr Sicherheit der an- und abfliegenden Flugzeuge im Nahbereich des Flughafens und ist damit verantwortlich für die Koordination und Sicherheit der Flugzeuge im entsprechenden Flugstatus.

Natürlich gibt es neben den soeben vorgestellten Aufgabenbereichen der ANSPs noch zahlreiche weitere. Grundsätzlich sind aber die zuvor beschriebenen Bereiche die Hauptaufgaben der ANSPs in Europa. Im Grunde beinhalten diese für einen effizienten und sicheren Flugraum zu sorgen. Dies gilt vom Start bis zur Landung eines Flugzeuges. Im Rahmen dieser Masterarbeit gilt aber jenen Maßnahmen, Techniken oder Systemen besondere Aufmerksamkeit, mit welchen die ANSPs in der Ausführung ihrer Tätigkeit Einfluss auf die negativen Auswirkungen für Umwelt und Klima des Flugverkehrs nehmen können. Der Umweltaspekt hat auch bei den ANSPs in den letzten Jahren stark an Bedeutung zugenommen (DFS, 2023). Teilweise wurden eigene Abteilungen installiert oder Mitarbeiterinnen und Mitarbeiter eingestellt, welche sich mit potentiellen Systemen und Praktiken auseinanderzusetzen, um den Luftverkehr sowohl am Boden als auch in der Luft nachhaltiger gestalten zu können (Austro Control, 2019a; Skyguide, 2023). Deswegen wird es besonders aufschlussreich sein im Laufe dieser Arbeit zum Vorschein zu bringen, welche Maßnahmen, Praktiken und Technologien heutzutage bei den ANSPs Anwendung finden, um die negativen Auswirkungen des Flugverkehrs für die Umwelt zu reduzieren. Darüber hinaus soll auch die Wirksamkeit der identifizierten Praktiken zum Vorschein gebracht werden.

2.3. Umweltfreundliche Möglichkeiten im Air Navigation Service

Wie bisher geklärt, liegt die Hauptaufgabe der ANSPs im operationalen Geschäft des Flugverkehrs. Dies beinhaltet die Steuerung, Überwachung und Koordination des Flugvorkommens in einem bestimmten Verantwortungsbereich. Hauptziel ist es zu jeder Tages- und Nachtzeit einen effizienten und vor allem sicheren Luftraum zu gewährleisten. Obwohl ANSPs daher hauptsächlich auf der operationalen Ebene tätig sind, so haben auch diese natürlich die Möglichkeit gewisse Techniken oder Systeme zu nutzen, um die Luftfahrt klima- und umweltfreundlicher zu gestalten (DFS, 2023a). Im Folgenden werden einige Möglichkeiten und Praktiken näher beschrieben. Zur Identifikation der folgend beschriebenen Methoden wurden vor allem Webseiteninformationen verschiedener europäischer ANSPs analysiert.

Eine zentrale Möglichkeit durch die ANSPs zur Reduzierung der negativen Umweltauswirkungen der Luftfahrt, welche sehr häufig auf diversen Webseiten von europäischen ANSPs genannt wird, beinhaltet alle Methoden und Prozesse unter dem Begriff der Flugroutenoptimierung (Austro Control, 2019a; DFS, 2023a). Die Österreichische Flugsicherungsorganisation Austro Control (2019a) betont die Hauptaufgaben bei der soeben genannten Maßnahme als die Findung und Festlegung des kürzesten Luftwegs zwischen zwei Punkten. Ein Flug über die kürzest mögliche Strecke spart natürlich Flugdauer und Kraftstoff. Dadurch wird im Endeffekt eine Verringerung des CO2-Ausstoßes und anderen umweltschädlichen Emissionen angestrebt.

In diesem Zusammenhang wird sehr häufig der Begriff der Free Route Airspace (FRA) genannt. Laut Austro Control (2019a) stellt die Free Route Airspace einen wichtigen Meilenstein dar, um das Ziel der CO2-Reduzierung zu erreichen, und wird folgendermaßen definiert. Die FRA ermöglicht es einem Flugzeug den Luftraum zwischen frei gewählten Ein- und Ausflugspunkt auf der direkten, kürzest möglichen Strecke zu durchfliegen. Dies geschieht aber natürlich nach wie vor unter der Kontrolle der für den jeweiligen Luftraum zuständigen Flugsicherung. Dadurch werden Flugwege verkürzt und der Treibstoffverbrauch verringert. In der Vergangenheit wurden Flugzeuge von Fluglotsinnen und Fluglosten auf vordefinierten Luftstraßen geführt, wobei es sich meist nicht um die kürzeste Strecke handelte. Es handelte sich wie erwähnt um vordefinierte Luftstraßen, auf welchen ein Land oder bestimmte Bereiche eines Landes überflogen werden mussten. Die FRA ist aber nur ein Beispiel von sehr vielen Praktiken und Möglichkeiten im Sinne der operationalen Effizienzsteigerung von Flugrouten. Häufig werden auch bestimmte Verfahren während Sink- und Steigflügen angewendet.

Die Deutsche Flugsicherung DFS (2023a) beschreibt hierfür beispielsweise den Kontinuierlichen oder Stetigen Sinkflug folgend genauer. Im Verlauf eines kontinuierlichen Sinkflugs wird eine minimale Triebwerksleistung benötigt. Das Flugzeug sinkt, vorzugsweise im Gleitflug, stetig bis zur Landung, wodurch Treibstoff eingespart und Lärm reduziert wird. Optimal ist es, wenn der kontinuierliche Sinkflug bereits nach dem Verlassen der Reiseflughöhe eingeleitet wird. Diese Vorgehensweise ist jedoch nur unter bestimmten Bedingungen umsetzbar. Hierbei sind mögliche Einschränkungen wie ein hohes Verkehrsaufkommen zu beachten.

Abbildung 2 stellt die Vorgehensweise beim Kontinuierlichen Sinkflug grafisch dar.

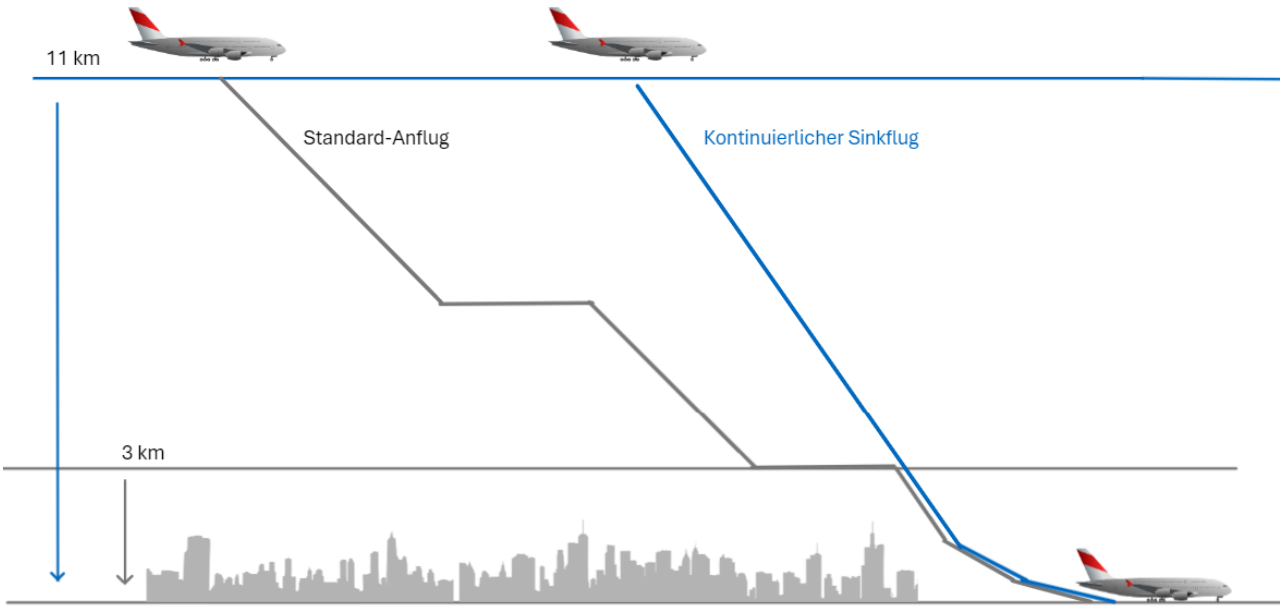


Abbildung 2: Kontinuierlicher Sinkflug

Quelle: Eigene Darstellung in Anlehnung an DFS (2023a)

Gegenteilig zum Kontinuierlichen Sinkflug, gibt es auch den Kontinuierlichen Steigflug, welcher im Grunde ident funktioniert (Eurocontrol, 2020). Obwohl es sich hierbei um zwei recht einfache operationale Maßnahmen des Flugverkehrs handelt, so hat kann dies bereits einen enormen Einfluss auf die Emissionen des Flugverkehrs haben (Eurocontrol, 2020). Eurocontrol (2020) bestätigt, dass bei häufiger Anwendung von Kontinuierlichen Sink- und Steigflügen innerhalb Europa 340.000 Tonnen Kerosin und über 1 Million Tonnen CO2 eingespart werden können. Dies resultiert wiederum in der Reduzierung der Lärmbelästigung für Bewohnerinnen und Bewohner rund um den Flughafen, als auch in Kostenersparungen von über 150 Millionen Euro für Fluglinien.

Bei den eben genannten und erklärten Beispielen handelt es sich allerdings nur um prominente Einzelfälle aus dem großen Begriff der Flugroutenoptimierung (DFS, 2023a). Es sollte aber dennoch ein Verständnis geben, dass es hierbei von den ANSPs sehr viel Spielraum und Möglichkeiten gibt, um den Luftraum auf operationaler Ebene effizienter zu machen und damit im Endeffekt auch Kerosin zu sparen, Lärmentwicklungen zu minimieren und damit auch das Klima und die Umwelt verstärkt zu schützen. Neben der FRA und dem Kontinuierlichen Sinkflug, nennt Eurocontrol (2023) weitere operationale Verfahren, den Luftraum effizienter zu machen:

- Kontinuierlicher Steigflug
- Curved Approach
- Airport collaborative decision-making (A-CDM)
- Performance-based Navigation (PBN)
- Systeme zum Umweltmanagement
- Advanced flexible use of airspace (AFUA) (Eurocontrol, 2023)

Neben den Methoden zur Flugroutenoptimierung gibt es selbstredend noch viele weitere, welche andere Ziele verfolgen und von den ANSPs angewendet werden. Austro Control (2019a) sowie die Civil Air Navigation Services Organisation (CANSO) (o. J.) beschreiben weitere prominente Beispiele

wie folgt. Hierbei ist die Rede von Systemen zur Unterstützung bei der Entscheidungsfindung oder auf Englisch Decision Support Systems. Diese Systeme können in den verschiedensten Bereichen der ANSPs zum Einsatz kommen. Beispielsweise am Flughafen selbst, dort ist häufig von Airport Collaborative Decision Making System (A-CDM) Systemen die Rede. Derartige Systeme finden Einsatz zur Hilfestellung bei der Koordination wenn mehr als ein Flugzeug gleichzeitig die Landebahn anfliegen möchte, oder diese verlassen. Das System soll dann Hilfestellung für die Entscheidungsfindung bieten, um die Situation so effizient und koordiniert wie möglich abzuwickeln. Diese Systeme finden, wie bereits erwähnt, in unterschiedlichen Bereichen Anwendung. Daher kann auch hier wieder die Namensgebung zwischen den ANSPs unterschiedlich sein. Nichtsdestotrotz ist ein häufig in Verwendung befindliches System das vorher erwähnte A-CDM.

CANSO (o. J.) beschreibt die Funktionsweise eines A-CDM folgendermaßen. Airport Collaborative Decision-Making (A-CDM) ist ein Programm zur Verbesserung der Effizienz von Flughafenbetrieben durch die Optimierung von Ressourcennutzung und die Vorhersagbarkeit von Ereignissen. Dies wird durch den Echtzeit-Austausch von Informationen zwischen Flughafenbetreibern, Fluggesellschaften, Bodenabfertigungsdiensten und der Flugsicherung erreicht. Erfolgreiche A-CDM-Implementierungen können die Effizienz, Kapazität und Umweltschutzleistung verbessern, was für alle Interessengruppen wichtig ist. Die Messung der A-CDM-Performance und die Erfüllung von Erwartungen sind entscheidend für den Return on Investment, sowohl für die anfänglichen Kosten als auch für die fortlaufende Optimierung am Flughafen. Vorteile für Fluggesellschaften wie verkürzte Rollzeiten und geringerer Treibstoffverbrauch sind bekannt, aber auch die Vorteile für die Flugsicherungsorganisationen sind erheblich.

Austro Control (2019a) beschreibt, dass durch die Inbetriebnahme eines A-CDM Systems, Einsparungseffekte in Höhe von 6.300 Tonnen CO2 über 30 Monate erfolgt sind. Dies entspricht einer Emissionseinsparung von ca. 2.500 Tonnen CO2 pro Jahr. Abbildung 3 zeigt wie durch ein A-CDM System die Abflüge anhand deren Bereitschaft besser koordiniert werden und dadurch Rollzeiten verkürzt und Verspätungen reduziert werden. Dies stellt einen von mehreren möglichen Einsatzbereichen eines A-CDM Systems dar.

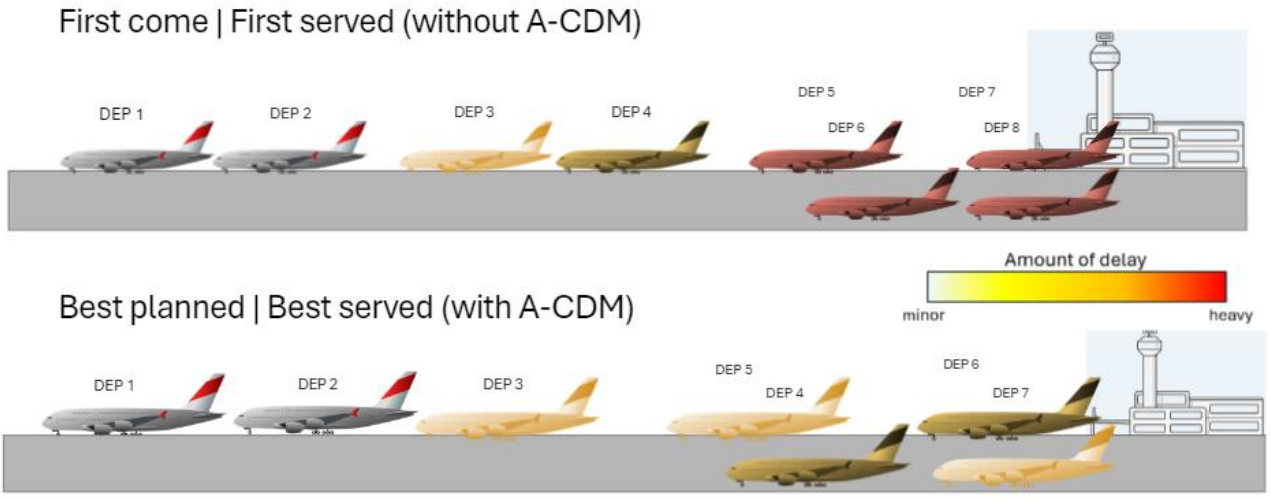


Abbildung 3: A-CDM System

Quelle: Eigene Darstellung in Anlehnung an Austro Control (2019a)

Einige bedeutende Möglichkeiten zur Reduzierung der negativen Umweltbelastungen, herbeigeführt durch den Flugverkehr, welche von Seiten der ANSPs genutzt werden können, wurden soeben genannt. Hierbei handelte es sich einerseits um den breiten Begriff der Flugroutenoptimierung, wobei zwei Praktiken näher beschrieben wurden, sowie technische Systeme zur Hilfestellung bei der optimalen Entscheidungsfindung im Sinne der Effizienz und Ressourcenschonung. In diesem Bereich wurde das A-CDM System näher beschrieben. Anzumerken ist allerdings, dass es sich hierbei in der Regel nicht um Standards handelt, sondern jeder ANSP individuell ist und selbst entscheiden kann, welche Maßnahmen implementiert und wie sie verwendet werden. Da es sich also um sehr individuelle Entscheidungsmöglichkeiten handelt, gibt es bei den Flugsicherungsorganisationen auch zahlreiche verschiedene Maßnahmen, Techniken und Systeme, welche zur Anwendung kommen. In diesem Teil wurden allerdings trotzdem einige Methoden beschrieben, welche tendenziell sehr häufig von ANSPs in Europa angewendet werden. Oft ist das Hauptziel der verwendeten Systeme und Maßnahmen eine erhöhte Effizienz am Boden und in der Luft. Diese Maßnahmen, egal ob rein operational oder mit Technologieunterstützung können, wie in diesem Kapitel durch Zahlen belegt, bereits enorm positive Auswirkungen auf CO2-Ausstoß, Treibstoffverbrauch, Lärmentwicklung sowie andere negative Effekte des Luftverkehrs haben. Die folgende Analyse im Rahmen dieser Masterarbeit wird weitere derartige Möglichkeiten, Systeme und Techniken zum Vorschein bringen.

Die in diesem Kapitel beschriebenen Grundlagen bezüglich der Umweltauswirkungen des Luftverkehrs sowie jene der ANSPs im allgemeinen Sinn, geben einen Überblick, über den Kontext in welchem sich diese Arbeit bewegt. Es wurde die Grundsituation vermittelt, dass der Flugverkehr einen signifikanten Einfluss auf die Umwelt und das Klima hat. ANSPs, welche es zur Hauptaufgabe haben den operativen Teil des Luftraumes zu koordinieren und damit für einen sicheren Luftraum zu sorgen, haben in der Ausführung ihrer Tätigkeit ebenfalls die Chance auf der operationalen Ebene, durch teilweise einfach umsetzbare Maßnahmen, großes im Sinne der Ressourcenschonung und damit dem Klima- und Umweltschutz zu bewirken. Nach der Klarstellung der Grundsituation und des Kontextes, werden nun im nächsten Abschnitt die technischen Grundlagen, welche für das weitere Vorgehen essentiell sind, vermittelt. Dabei ist von Web-Scraping und Text Mining die Rede.

3. Grundlagen

In diesem Kapitel werden die Grundlagen der in der Arbeit angewendeten Systeme näher erläutert. Die behandelten Hauptelemente stellen dabei das Web-Scraping sowie das Text Mining dar. Für das Web-Scraping wird einerseits die generelle Funktion sowie der Prozess erläutert. Darauffolgend werden alle Vorbereitungen beziehungsweise notwendige Grundlagen beschrieben, um Web-Scraping ganzheitlich zu verstehen sowie den Prozess effizient anzuwenden. Dieses Unterkapitel wird mit der Erläuterung der zu beachtenden Herausforderungen abgeschlossen. Auch das darauffolgende Kapitel des Text Mining beginnt mit einer allgemeinen Einführung und Beschreibung des Prozesses. Danach werden die gängigsten Methoden im Text Mining kurz beschrieben, sowie für die Arbeit anzuwendende Methoden identifiziert und beschrieben.

3.1. Web-Scraping

Wie bereits bekannt werden im Rahmen dieser Masterarbeit Informationen bezüglich aktuellen Praktiken, Maßnahmen und Systemen, welche von ANSPs in Europa eingesetzt werden, um die negativen Auswirkungen des Luftverkehrs auf die Umwelt und das Klima zu reduzieren, gesammelt. Die dafür benötigten Informationen werden in erster Linie von den Webseiten der europäischen ANSPs gesammelt. Die Sammlung der notwendigen Informationen für diese Recherche könnten also im Grunde folgendermaßen ablaufen. Es wäre möglich die zahlreichen Webseiten der europäischen ANSPs nach den gewünschten Inhalten zu durchsuchen und damit eine Datenbasis aufzubauen, welche im Anschluss strukturiert und analysiert wird. Da dies einen sehr ineffizienten und vor allem zeitlich langanhaltenden Prozess darstellt, wird hierfür die Technik des Web-Scrapings angewendet. Folgendes Kapitel wird einen Überblick über die relevanten Aspekte dieser Datenerhebungsmethode geben. Dies beinhaltet vor allem die Funktionalität und den Prozess, notwendige Vorbereitungen sowie die Herausforderungen.

3.1.1. Funktion und Prozess

Khder (2021) und Zhao (2017) beschreiben die Funktion des Web-Scrapings wie folgt. Beim Web-Scraping geht es im Grunde darum, Informationen wie beispielsweise Textinhalte, von Webseiten zu extrahieren und diese dann anschließend in einer Datenbank zu sichern. Die gesammelten Informationen können anschließend für Abfragen oder Analysen verwendet werden. Üblicherweise erfolgt die Extraktion von Webdaten mithilfe des Hypertext Transfer Protocol (HTTP) oder durch einen Webbrowser. Dieser Prozess kann entweder manuell durchgeführt werden oder automatisiert mit sogenannten Web-Crawlern. Aufgrund der enormen Menge an Informationen und Daten im World Wide Web (WWW), stellt das Web-Scraping eine sehr effiziente und leistungsstarke Technik zur Sammlung von Daten und Informationen dar.

Die gesamte Funktionalität des Web-Scrapings wird in Abbildung 4 übersichtlich dargestellt. Khder (2021) unterteilt den Prozess in folgende drei Schritte. Der erste Schritt beinhaltet die Webseitenanalyse. In diesem ersten Schritt wird sich mit dem Aufbau und der Struktur der Webseite vertraut gemacht. Dieser Schritt ist unerlässlich, um die Webseite im Anschluss für das Web-Scraping zu verwenden, und auch die gewünschten Informationen und Texte zu extrahieren. Der nächste Schritt beinhaltet das sogenannte Website Crawling. Im Rahmen dieser Arbeit wurde dies entweder mittels

der Python Bibliothek „Beautiful Soup“, oder der R Bibliothek „rvest“ durchgeführt. In diesem Schritt läuft ein Web-Crawler über die Webseite und sichert jene Text und Inhalte, welche dem gewünschten Format entsprechen beziehungsweise in den spezifizierten Komponenten der Webseite gespeichert sind. Im letzten Schritt geht es um die Datenorganisation, Sicherung, Säuberung sowie anschließenden Analysen oder Abfragen, um gewünschte Informationen zum Vorschein zu bringen.

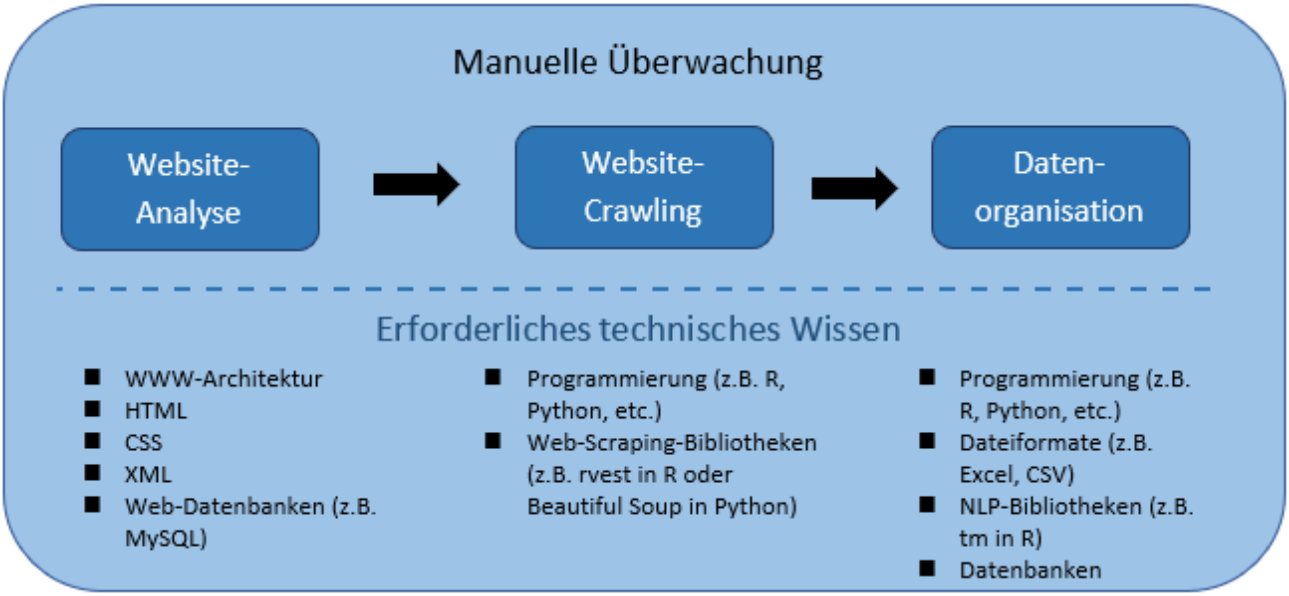


Abbildung 4: Übersicht Funktionalität Web-Scraping

Quelle: Eigene Darstellung in Anlehnung an Khder, 2021, S.148

Nach dieser ersten theoretischen Einführung in die Funktion des Web-Scrapings, lässt sich der spezifische Prozess des Web-Scrapings allerdings noch genauer darstellen. Khder (2021) unterteilt diesen in vier Phasen. Ablauf und Funktion gemäß Khder (2021) sowie die Prozessschritte in jeder einzelnen dieser vier Phasen wird folgend näher erläutert und ist in Abbildung 5 zusammengefasst.

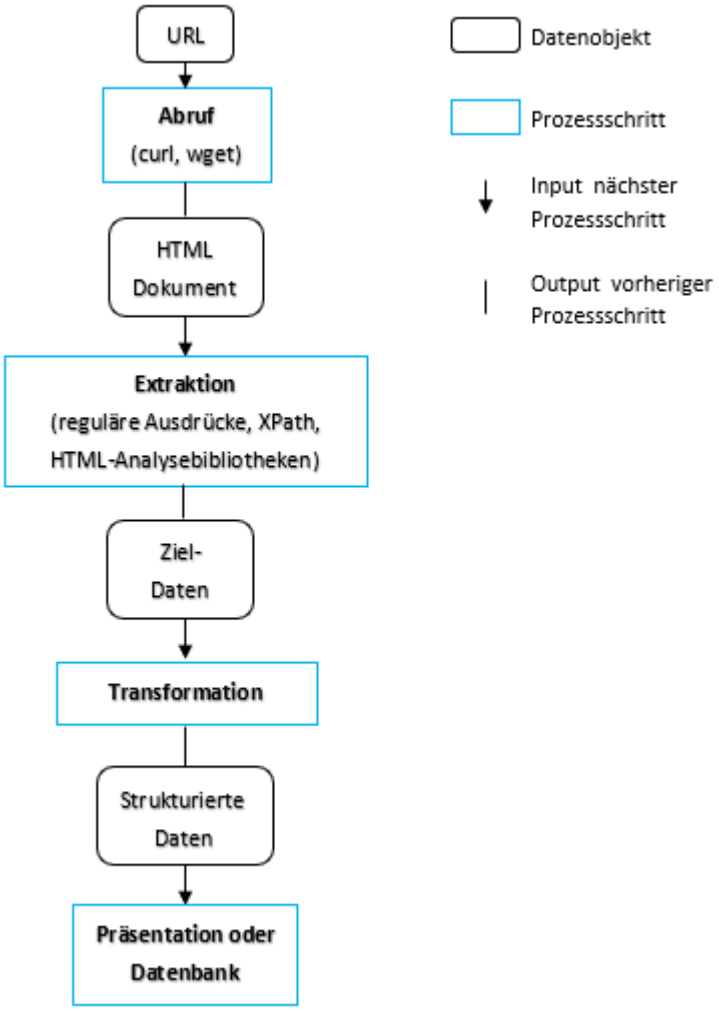


Abbildung 5: Web-Scraping-Prozess

Quelle: Eigene Darstellung in Anlehnung an Khder, 2021, S. 149

Der Prozess des Web-Scrapings beginnt mit der Uniform Resource Locater (URL) einer Webseite, auf welche man zugreifen möchte beziehungsweise von welcher Texte extrahiert werden sollen. Mit diesem ersten Schritt beginnt bereits die erste der vier Prozessschritte. Die vier Hauptphasen werden wie bereits erwähnt laut Khder (2021) folgendermaßen unterteilt:

- **Abruf:** In dieser Phase muss zunächst die gewünschte Webseite mit den relevanten Informationen aufgerufen werden. Dies erfolgt in der sogenannten Abrufphase über das HTTP-Protokoll, einem Internetprotokoll zum Senden und Empfangen von Anfragen von Webservern. Webbrowser verwenden ähnliche Methoden, um Inhalte auf Webseiten zu erhalten. In diesem Schritt können Bibliotheken wie „rvest“ von R oder „Beautifulsoup“ von Python oder ähnliche durch Senden einer http-GET-Anfrage an die Zieladresse (URL) verwendet werden, um die HTML-Seite als Antwort zu erhalten.
- **Extraktion**: Nach dem Aufrufen der HTML-Seite werden die relevanten und gewünschten Daten extrahiert. In dieser als Extraktionsphase bezeichnenden Stufe kommen reguläre Ausdrücke, HTML-Analysebibliotheken und XPath-Abfragen zum Einsatz. Im Grunde geht es hierbei um die Identifikation der spezifizierten Passagen und Textinhalten auf den Webseiten.

- **Transformation**: Da nun die relevanten Daten und Passagen auf der Webseite identifiziert wurden, können sie in ein strukturiertes Format für spätere Präsentation oder Speicherung umgewandelt werden.
- **Präsentation oder Datenbank**: Der finale Prozessschritt im Web-Scraping beinhaltet die Präsentation der gesammelten Informationen oder die Sicherung deren in eine Datenbank.

3.1.2. Aufbau HTML-Webseite

Um Web-Scraping nun für eine oder mehrere Webseiten anzuwenden, bedarf es noch dem Verständnis, wie eine HTML-Webseite aufgebaut ist. Dies ist einer der wichtigsten Schritte, um schlussendlich auch die richtigen und gewünschten Inhalte von einer Webseite zu sammeln (Mine & Mine, 2021). Es wird sich bei der folgenden Beschreibung nur auf die nötigsten Inhalte beschränkt, welche ausreichen um den Prozess zu erklären und zu verstehen.

Rybka (o.J.) beschreibt die Struktur eines HTML-Grundgerüsts wie folgt. Dies besteht aus den essenziellen Bausteinen, die erforderlich sind, um Inhalte im Webbrowser anzuzeigen. In einer HTML-Datei sind verschiedene Abschnitte enthalten. Der erste Abschnitt gibt dem Browser Anweisungen (Doctype), der zweite Abschnitt enthält Metadaten (Head), wie zum Beispiel den Titel der Webseite im <title>-Tag, und der dritte Abschnitt beinhaltet den eigentlichen Inhalt der Webseite (Body). Innerhalb der beiden Komponenten Head und Body können Informationen in verschiedenen Tags untergebracht werden. Textinformationen werde dabei häufig in Paragraph <p> oder Span -Tags gespeichert.

Für Überschriften gibt es beispielsweise eigene Elemente wie den <h1>-Tag oder um einen externen Link zu hinterlegen einen <a>-Tag (Mine & Mine, 2021). Jene Tags, in welchen Textinformationen ausgegeben werden, sind für die Zwecke dieser Arbeit besonders relevant. Mit diesem Wissen ist es bereits möglich, eine sehr einfache Webseite in HTML-Format zu gestalten. In Abbildung 6 wird der Code einer sehr einfachen, selbst erstellen HTML-Webseite gezeigt, sowie dessen anschließende Anzeige in einem Browser.

```
<!DOCTYPE html>
<html>
  <head>
    <title>Titel der Webseite</title>
  </head>
  <body>
   <h1>Aktuelle ANSP-Praktiken zur Reduzierung der Umweltbelastungen durch den Flugverkehr</h1>
     <p>In einem Paragraph(p) Ttag können Textinformationen im Body einer Webseite abgespeichert werden</p>
     <span>In einem span Tag können ebenfalls Texte gespeichert werden</span>
     <p>Mittels a Tags und dazugehörigem Attribut href können Links zu externen Seiten in Texten gespeichert werden.</p>
     <p>Hier geht's zur Webseite der <a href="https://www.austrocontrol.at/">Austrocontrol</a></p>
  </body>
</html>
```

Abbildung 6: Beispielaufbau einer einfachen HTML-Webseite

Quelle: Eigene Darstellung

Wie in der obigen Beschreibung und Abbildung 6 schon erahnt werden kann, ist im Rahmen dieser Arbeit besonders die HMTL-Komponente des Body wichtig, da hier in der Regel die Textinformationen und Paragraphen gespeichert werden (Rybka, o. J.).

Die eben erklärten Konzepte und Schritte beschreiben den Web-Scraping-Prozess umfassend. Es wurde aufgezeigt wie ausgehend von einer URL, gewünschte Informationen einer Webseite extrahiert werden können. Um den Prozess ganzheitlich zu verstehen und diesen im Anschluss auch praktisch anwenden zu können, bedarf es nun noch einem Vorgehen zur Identifikation der gewünschten HTML-Elemente und Komponenten innerhalb der Webseite.

3.1.3. Identifikation relevanter Komponenten

Um den Prozess des Web-Scrapings sowie die Identifikation der relevanten Komponenten und Passagen holistisch zu verstehen, wird folgend ein einfaches Beispiel grafisch dargelegt. Die zum Verständnis notwendigen Grundladen wurden hierfür bereits beschrieben. Genau wie in folgendem Beispiel dargestellt und erklärt, wurde das Web-Scraping auf die zahlreichen Webseiten der europäischen ANSPs im Rahmen dieser Arbeit angewendet. Davor ist es aber essentiell zu analysieren, wie eine Webseite aufgebaut ist und in welchen Komponenten die relevanten Informationen, im vorliegenden Fall Textinformationen, gespeichert sind. Diese Information kann auf mehrere Arten ermittelt werden (Wickham, o.J.). Einerseits indem der Seitenquelltext einer Webseite analysiert wird oder mittels sogenannten Selector-Gadgets, welche als Erweiterungen in Browsern heruntergeladen werden können (Wickham, o. J.). Diese erkennen die Komponenten automatisch und zeigen direkt an,

um welche es sich handelt (Wickham, o. J.). In der praktischen Anwendung stellte sich letztere als schnellere und effizientere Variante dar.

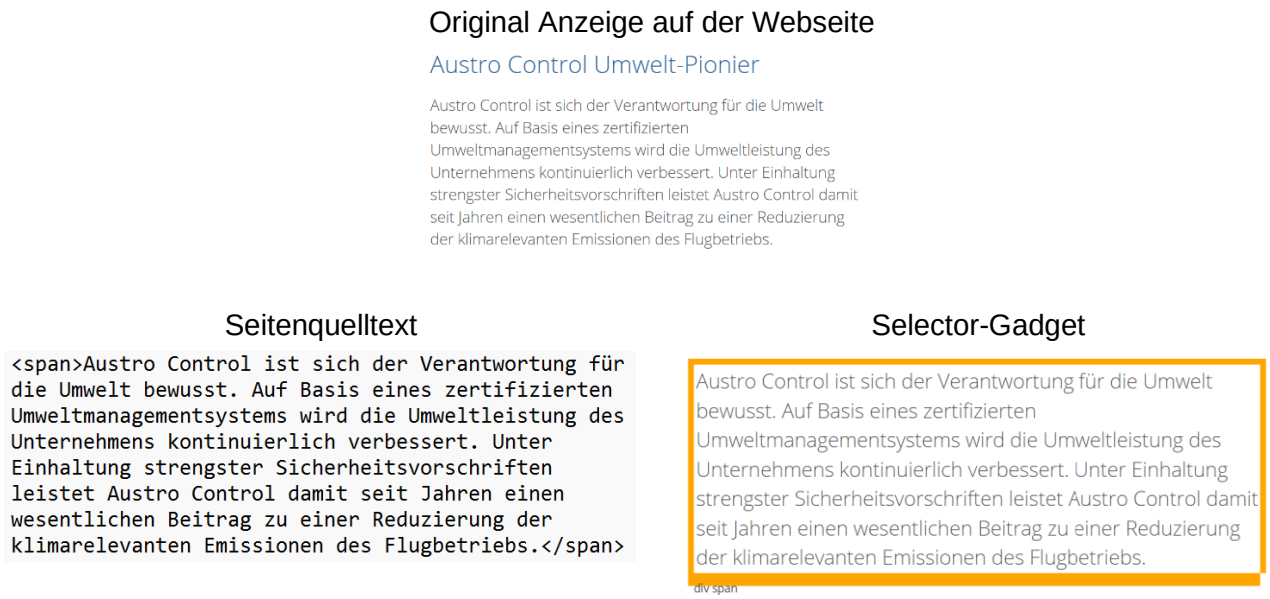


Abbildung 7: Identifikation relevanter Komponenten einer HTML-Webseite

Quelle: Eigene Darstellung

Wie in Abbildung 7 visuell dargestellt, kann durch die Anwendung beider Methoden herausgefunden werden, in welchen Tags die analysierten Textinformationen zu finden sind. Diese Information wird anschließend an den Web-Crawler übergeben, welcher anschließend die Webseite durchläuft und die Informationen aus diesen Komponenten extrahiert (Diouf et al., 2019; Khder, 2021; Mine & Mine, 2021; Zhao, 2017). In Rahmen des eben vorgestellten Beispiels, würde die Information, das es sich um einen -Tag handelt, and den Crawler übergeben. Dadurch wäre sichergestellt, dass die Textpassagen in der zuvor gezeigten Abbildung 7 vom Web-Crawler auf der entsprechenden Webseite identifiziert und extrahiert werden.

3.1.4. Herausforderungen

Trotz der zahlreichen Vorteile und der Möglichkeit zur effizienten und schnellen Datenbeschaffung im Web, gibt es bei der Anwendung von Web-Scraping auch einige Herausforderungen. Mine & Mine (2021) fassen die vier relevantesten Herausforderungen beim Web-Scraping folgendermaßen zusammen.

Die erste Herausforderung beim Web-Scraping besteht in der Schwierigkeit der Reproduzierbarkeit. Dies liegt hauptsächlich daran, dass die Daten auf Webseiten zumeist nicht statisch sind. Selbst wenn derselbe Code zu einem späteren Zeitpunkt auf dieselbe Webseite angewendet wird, ist es durchaus realistisch, unterschiedliche Ergebnisse zu erhalten. Dies liegt meist daran, dass die Quellendaten auf der Zielwebseite aktualisiert oder verändert wurden. Ein weiterer Grund für die mangelnde Reproduzierbarkeit ist, dass die Struktur von Webseiten im Laufe der Zeit verändert werden kann. Dies kann zur Folge haben, dass ein Scraping-Script, welches zuvor für die Datenextraktion funktioniert hat, möglicherweise nicht mehr funktioniert. Daher sind laufende Anpassungen am Code bei wiederholten Scraping-Prozessen essentiell.

Die zweite Herausforderung wie von Mine & Mine (2021) dargestellt, befasst sich mit der Datenqualität und dem Umgang mit fehlenden Daten. Es ist wichtig zu betonen, dass dies keine spezifische Herausforderung des Web-Scrapings ist, sondern vielmehr eine allgemeine Problematik, wenn mit großen Mengen an Daten gearbeitet wird. Fehlende Daten stellen eine allgegenwärtige Herausforderung dar, da sie die Genauigkeit und Zuverlässigkeit von Analysen beeinträchtigen.

Die dritte und für das Web-Scraping spezifische Herausforderung stellen alle Herausforderungen dar, welche die automatisierte Sammlung von Inhalten auf Webseiten erschweren oder blockieren. Webseiten können beispielsweise eine IP-Adresse sperren oder die Zugriffshäufigkeit begrenzen, mit welcher dieselbe IP-Adresse auf eine Webseite zugreifen darf. Diese Sperren werden häufig festgelegt, wenn der Anschein besteht, dass eine IP-Adresse für bösartige oder übermäßige Anfragen verwendet wird. Darüber hinaus wird die Zugriffshäufigkeit auf Webseiten in manchen Fällen generell begrenzt, um im allgemeinen für mehr Sicherheit vor bösartigen Zugriffen zu sorgen. Dies kann das Scraping-Script komplett blockieren oder stark verlangsamen, da die Zugriffshäufigkeiten respektiert werden müssen. CAPTCHAs (Completely Automated Public Turing Tests to Tell Computers and Humans) sind eine weitere gängige Sicherheitsmaßnahmen, die dem Web-Scraping den Zugang zu Webseiten erschweren. Diese erfordern manuelle Interaktion zur Lösung einer Herausforderung, bevor gewünschte Inhalte zugänglich gemacht werden. Es kann sich bei CAPTCHAs beispielweise um Bilderkennung, textuelle- oder auditive Rätsel handeln.

Die vierte und abschließende Hauptaufgabe, wie von Mine & Mine (2021) erläutert, betrifft die mangelnde Kontrolle im World Wide Web (WWW). Dieser Herausforderung beinhaltet sowohl die Verfügbarkeit von Webseiten als auch die begrenzte Einflussnahme auf deren Inhalte. Im Hinblick auf die Verfügbarkeit kann es vorkommen, dass Webseiten oder Server zeitweise nicht erreichbar sind, was zu Schwierigkeiten beim Abrufen von Daten führen kann. Besonders bei benötigten Datenextraktionen in Echtzeit kann dies eine Herausforderung darstellen. Die mangelnde Kontrolle über die Inhalte bezieht sich darauf, dass es in den meisten Fällen schwierig ist, für die Richtigkeit und Aktualität der extrahierten Daten zu garantieren.

3.1.5. Zusammenfassung

Im vorliegenden Kapitel wurden die Grundlagen des Web-Scrapings behandelt. Es wurde hierbei auf die Funktionalität und den spezifischen Prozess näher hingewiesen. Der eigentliche Prozess beim Web-Scraping wird auf effiziente und in der Regel schnelle Weise von einem Programm durchgeführt (Khder, 2021). Im Fall der vorliegenden Arbeit wurde dies hauptsächlich durch die „rvest“ Bibliothek von R sichergestellt. Auf Seiten der Anwenderin oder des Anwenders ist es neben des Schreibens des eigentlichen Scraping-Scripts besonders wichtig für eine ausreichende Vorbereitung des Prozesses zu sorgen. Dies beinhaltet sich über die Struktur und den Aufbau der zu verwendeten Webseiten zu informieren, da dies eine der wichtigsten Anforderungen für das Script darstellt (Khder, 2021; Mine & Mine, 2021). Hierfür muss verstanden werden, wie eine HTML-Webseite aufgebaut ist, und in welchen Komponenten die relevanten Informationen gefunden werden können. Abgeschlossen wurde das Kapitel mittels Erklärung der gängigsten Herausforderungen beim Web-Scraping. Hier wurden laut Literatur vor allem die Herausforderungen der Reproduzierbarkeit, der Datenqualität, dem Umgang mit fehlenden Werten sowie die mangelnde Kontrolle im Internet und über die Inhalte auf Webseiten genannt. Web-Scraping bietet eine sehr effiziente und schnelle Weise, einen Zugriff auf

eine große Menge an Daten zu erlangen. Nichtsdestotrotz gibt es Herausforderungen sowie auch speziell Vorbereitungen, die berücksichtigt werden müssen, um das Web-Scraping optimal einzusetzen (Khder, 2021; Mine & Mine, 2021). Web-Scraping wird im Rahmen der vorliegenden Arbeit dazu dienen, die notwendige Textdatenbasis effizient zu kreieren. Anschließend wird diese mittels diversen Methoden und Techniken des Text Minings analysiert. Was unter diesem Begriff genau zu verstehen ist behandelt das nächste Kapitel.

3.2. Text Mining

Gemäß Talib et al., (2016) handelt es sich beim Text Mining um einen Prozess der Extraktion von interessanten und nicht trivialen Mustern aus einer großen Menge von Textdokumenten. Es gibt verschiedene Techniken und Werkzeuge, um aus Textinhalten wertvolle Informationen zu extrahieren. Die Auswahl der richtigen und geeigneten Text-Mining-Techniken ist dabei essentiell, um Zeit und Aufwand für die Extraktion der relevanten oder gewünschten Informationen zu minimieren.

Da es für die Zielerreichung dieser Arbeit notwendig ist aus einer großen Menge an Textinformationen, relevante Einblicke zum Vorschein zu bringen, werden dafür bestimmte Text-Mining-Methoden zur Hilfe herangezogen. Das folgende Kapitel stellt eine übersichtliche Einführung in die Thematik und die Methoden des Text Minings dar. Das Kapitel ist folgendermaßen aufgebaut. Begonnen wird mit einer generellen Einführung zu Text Mining, sowie den gängigsten Methoden und den Zielen des Text Minings. Im Anschluss wird eine Literaturrecherche durchgeführt, um schlussendlich jene Methoden zu identifizieren, welche sich für die Erreichung der Ziele der vorliegenden Masterarbeit am besten eignen.

3.2.1. Der Prozess des Text Mining

Auch Hippner & Rentzmann (2006) beschreiben die Herausforderung im Text Mining als jene, in einem Text sprachlich wiedergegebene Informationen für maschinelles Lernen oder maschinelle Analyse zu erschließen. Dies spiegelt sich laut den Autoren auch im Prozess des Text Minings wider. Dieser beinhaltet zwar Ähnlichkeit zu einem klassischen Data-Mining-Prozess, unterscheidet sich allerdings grundlegend in der Datenaufbereitung. Beim Text Mining ist es notwendig, eine zusätzliche linguistische Datenverarbeitung durchzuführen, um die fehlende Struktur der Daten zu rekonstruieren. Der zumeist iterative Prozess wird von Hippner & Rentzmann (2006) in Abbildung 8 veranschaulicht.

Abbildung 8: Text-Mining-Prozess

Quelle: Eigene Darstellung in Anlehnung an Hippner & Rentzmann, 2006, S. 288

Wie in Abbildung 8 zu sehen ist, teilen Hippner & Rentzmann (2006) den Prozess des Text Minings in sechs Stufen ein. Diese werden folgend erläutert.

Aufgabendefinition: in diesem ersten Schritt geht es darum, die Problemstellung sowie die Ziele des Vorhabens klar zu definieren (Hippner & Rentzmann, 2006). Wie auch schon vorher von Talib et

al. (2016) erwähnt, handelt es sich dabei um einen besonderen wichtigen Schritt, da je nach Problemstellung und verfolgtem Ziel, die passenden Text-Mining-Methoden gewählt werden müssen. Nur mit den passenden Methoden zu einer konkreten Problemstellung und einem definierten Ziel, kann mit qualitativ hochwertigen Resultaten gerechnet werden.

Dokumentselektion: Ausgehend von den definierten Zielen, geht es in diesem Schritt gemäß Hippner & Rentzmann (2006) darum, die relevanten Textdokumente zu wählen. Hierbei kann es sich um allerlei Dokumenttypen beziehungsweise Datenbasen handeln, wie beispielsweise Webseiten, E-Mails, Formulare, Berichte, oder ähnliche. Im Rahmen dieser Arbeit werden Webseiten und die darin enthaltenen Textinformationen der ANSPs in Europa als zu analysierende Dokumente gewählt.

Dokumentaufbereitung: In diesem Schritt handelt es sich laut Talib et al. (2016) um die Aufbereitung der Textdokumente für die anschließend anzuwendenden Methoden und Analysen. In der englischen Sprache wird hierbei oft von Data- oder Text Cleaning gesprochen. In diesem Schritt geht es im Grunde darum, Anomalien in den Dokumenten zu erkennen und zu entfernen. Dadurch wird sichergestellt, dass die eigentliche Essenz des vorliegenden Textes erfasst wird. Dabei werden beispielsweise Stopp-Wörter entfernt, Wortstämme identifiziert sowie Daten indiziert. Es muss allerdings auch beachtet werden, dass unterschiedliche Text-Mining-Methoden die Daten in der Regel in unterschiedlichen Formen benötigen. Dies führt zu unterschiedlichen Ansätzen der Dokumentenaufbereitung, basierend auf den verfolgten Zielen einer Analyse sowie den verwendeten Methoden.

Text-Mining-Methoden: Nachdem die Textdokumente in die gewünschte beziehungsweise notwendige Form gebracht wurden, können nun die spezifischen Text-Mining-Methoden darauf angewendet werden. Klassische Beispiele hierfür sind unter anderen die Klassifikation, Segmentierung oder Abhängigkeitsanalyse. (Hippner & Rentzmann, 2006). Auf gängige und relevante Methoden, vor allem in Bezug zur vorliegenden Zielsetzung dieser Arbeit, wird im nächsten Kapitel näher eingegangen.

Interpretation und Evaluation: Talib et al. (2016) sowie auch Hippner & Rentzmann (2006) beschreiben diesen vorletzten Schritt im Prozess folgendermaßen. In diesem Schritt werden die extrahierten Muster und Erkenntnisse analysiert, um ihre Relevanz und Bedeutung für das definierte Problem zu bewerten. Dies beinhaltet die Überprüfung der interpretierten Ergebnisse im Kontext des ursprünglichen Ziels, um sicherzustellen, dass die gewonnen Informationen den Anforderungen entsprechen. Diese Phase ermöglicht es, etwaige Verbesserungen oder Anpassungen am Text-Mining-Modell vorzunehmen, um die Qualität der Ergebnisse zu optimieren.

Anwendung der Ergebnisse: Hippner & Rentzmann (2006) beschreiben die Anwendung der Ergebnisse im Text-Mining-Prozess als jenen Schritt, in welchem die gewonnen Erkenntnisse in konkrete Handlungen oder Entscheidungen umgesetzt werden. Wurden die Ziele im vorherigen Schritt erreicht, beinhaltet dieser letzte Schritt nun die Integration der identifizierten Muster und Informationen in relevante Geschäftsprozesse oder Systeme, um im besten Fall einen praktischen Nutzen aus den Analysen zu generieren.

3.2.2. Methodenüberblick im Text Mining

Bevor nun im nächsten Schritt relevante und passende Methoden für die vorliegende Problemstellung identifiziert werden, gibt es in diesem Kapitel einen zusammenfassenden Überblick über Metho-

den, welche im Text Mining generell zu Verfügung stehen. Talib et al. (2016) unterteilt die im Text Mining zur Verfügung stehenden Methoden und Techniken in fünf Kernbereiche ein. Um einen ganzheitlichen Überblick über die im Text Mining angewandten und zur Verfügung stehenden Methoden zu erhalten, werden im Folgenden alle fünf Kernbereiche des Text Minings, welche von verschiedenen Autoren und Autorinnen als besonders aktuell, relevant oder häufig verwendet betrachtet werden, zusammenfassend erläutert.

Information Extraction (IE) ist laut den Autoren Talib et al. (2016) eine Technik, die das Ziel verfolgt, sinnvolle Informationen aus großen Textmengen zu extrahieren. IE-Systeme werden verwendet, um spezifische Attribute und Entitäten aus dem Dokument zu extrahieren und ihre Beziehung herzustellen. Der extrahierte Korpus wird in einer Datenbank für weitere Verarbeitung gespeichert. Zusammenfassend fokussieren sich diese Methoden darauf, spezifische, strukturierte Informationen aus einem Textdokument zu extrahieren.

Information Retrieval (IR) beschreiben Wuttke (2022) sowie Talib et al. (2016) als einen Prozess der Extraktion relevanter und zugehöriger Muster anhand gegebener Mengen von Wörtern oder Phrasen. IR-Systeme wenden unterschiedliche Algorithmen an, um beispielsweise das Verhalten der Nutzerin oder des Nutzers zu verfolgen und dementsprechend relevante Daten zu suchen. Sowohl Google als auch Yahoo Suchmaschinen verwenden IR-Systeme häufig, um anhand einer Phrase im Web relevante Dokumente zu extrahieren. Dadurch liefern sie der Nutzerin oder dem Nutzer relevante und angemessene Informationen, die seinen und ihren Bedürfnissen entsprechen.

Die eben beschriebenen Techniken werden die relevantesten für die vorliegende Problemstellung der Arbeit darstellen, da es hier um die Extraktion von Informationen aus einer Datenbasis geht. Um einen ganzheitlichen Überblick über gängige Text-Mining-Techniken zu erhalten, werden folgend auch die restlichen Kernfunktionalitäten des Text Mining zusammenfassend beschrieben.

Natural Language Processing (NLP) ist laut Allayhari et al. (2017) ein Teilgebiet der Informatik, künstlichen Intelligenz sowie der Linguistik. NLP zielt darauf ab, die natürliche Sprache mithilfe von Computern zu verstehen. Viele derartiger Algorithmen nutzen NLP-Techniken, wie zum Beispiel Part-of-Speech-Tagging (POS), syntaktisches Parsing und andere Arten von linguistischer Analysen. Derzeit aktuelle Anwendungsgebiete von NLP sind beispielsweise Chatbots, virtuelle Assistenten, automatisiertes Übersetzen von Texten oder das automatisierte Erstellen von Textzusammenfassungen.

Classification bezieht sich laut Allahyari et al. (2017) auf das überwachte Lernen, oder in englischer Sprache Supervised Learning. Hierbei wird eine Funktion abgeleitet oder ein Klassifikator aus Trainingsdaten gelernt, um Vorhersagen für nicht gesehene oder zukünftige Daten zu treffen. Im Text Mining handelt es sich hierbei um den Prozess der Zuordnung von Textdokumenten zu vordefinierten Kategorien oder Klassen. Hierbei ist es das Ziel, Muster in den Textdaten zu identifizieren und dadurch eine automatisierte Zuordnung der Textdokumente in bestimmte Klassen oder Kategorien vorzunehmen. Ein alltägliches Beispiel für die soeben beschriebene Themenzuordnung ist die automatisierte E-Mail-Spamerkennung.

Clustering stellt gemäß Talib et al. (2016) einen unüberwachten, auf Englisch Unsupervised Prozess zur Einteilung von Textdokumenten in Gruppen durch die Anwendung verschiedener Clustering Algorithmen dar. In einem sogenannten Cluster werden ähnliche Begriffe oder Muster gruppiert, die

aus verschiedenen Dokumenten extrahiert wurden. Bekannte Clustering Algorithmen oder Techniken beinhalten beispielsweise hierarchisches, verteiltes, dichtebasiertes, zentroidbasiertes oder k-meam-Cluster.

Nach diesem kurzen Überblick über jene Text-Mining-Methoden, welche laut Literatur besonders relevant oder als aktuell angesehen werden, stellt es sich der nächste Abschnitt zum Ziel, die passenden Methoden für die vorliegende Problemstellung zu identifizieren und in Folge dann auch im Rahmen der Arbeit effizient zu nutzen. Vorab ist allerdings bereits bekannt, dass die Hauptaufgabe der zu verwendenden Methoden darin bestehen wird, relevante Informationen aus einer großen Menge an Textinhalten zu gewinnen. Aus diesem Grund ist bereits zu diesem Zeitpunkt absehbar, dass die zuvor definierten Kategorien des Information Retrievals sowie der Information Extraction in diesem Kontext von Bedeutung sein werden. Allerdings handelt es sich hierbei erst um Überbegriffe beziehungsweise Kategorien. Die konkreten Techniken werden daher mithilfe einer Literaturanalyse identifiziert.

3.2.3. Literaturanalyse – Identifikation passender Methoden

Auf der Suche in der Literatur nach ähnlichen Problemstellungen beziehungsweise ähnlichen Vorgehensweisen, wurde unter anderen das Paper von Holubliev & Simishko (2021) mit dem Titel Web-Scraping and Text Mining of Ukranian News Articles About Ecology gefunden. In dieser Arbeit setzten es sich die beiden Autoren zum Ziel ukrainische Nachrichtenwebseiten zu analysieren, deren Texte zu extrahieren und anschließend diese auf Ökologie-Themen zu untersuchen. Die Ergebnisse dieser Arbeit sollen dann von Organisationen oder auch Privatpersonen genutzt werden können, die sich mit umweltbezogenen Projekten beschäftigen. Die Stude zeigt darüber hinaus die Entwicklung von umweltbezogenen Themen im Zeitverlauf der letzten Jahre. Dadurch können wichtige Ereignisse extrahiert und deren Auswirkungen auf die Medienberichterstattung beobachtet werden. Die Ziele dieser Arbeit unterscheiden sich zwar von der hier vorliegenden, nichtsdestotrotz ist das Vorgehen und die Problemstellung sehr ähnlich. Zuerst werden die Informationen von den Webseiten extrahiert und diese dann nach bestimmten Informationen durchsucht, um anschließend Aussagen darüber, in diesem Fall die Ökologie, treffen zu können. Aus diesem Grund ist die Methodologie dieser Arbeit von großem Interesse, um herauszufinden wie die Autoren die nützlichen Informationen aus den Texten herausfilterten.

Die Autoren Holubliev & Simishko (2021) starteten ihre Arbeit bei der Identifikation glaubwürdiger Nachrichtenquellen. Hier wurde besonders darauf geachtet, eine hohe Anzahl an seriösen Webseiten zu finden, um den Effekt der Diversifikation zu nutzen. Nach der Identifikation der Webseiten wurde ein Scraping-Script geschrieben, welches die Textinformationen der Webseiten extrahiert. Anschließend wurden relevante Passagen und Artikel mittels einer Liste von spezifischen Schlüsselwörtern oder Keywords gefunden. Dadurch ergab sich der Vorteil, dass nicht der gesamte extrahierte Text analysiert werden musste, sondern nur jene Passagen, welche mindestens einen dieser relevanten Schlüsselbegriffen beinhaltetet. Die Autoren erstellten diese Liste mit Schlüsselwörtern basierend auf globalen, ökologischen Problemen und weiteren Fachbegriffen.

Zusammengefasst kann also aus der Arbeit von Holubliev & Simishko (2021) mitgenommen werden, dass hier die relevanten Inhalte mittels Keywords identifiziert wurden. Diese stellen aktuelle Probleme oder Fachbegriffe im jeweiligen Bereich, in diesem Fall der Ökologie, dar. Auch für das Ziel die-

ser Arbeit stellte die Keyword-Analyse eine passende Methode dar, um die relevanten Inhalte der Webseiten und anderer Dokumente zu identifizieren und anschließend zu analysieren.

Eine weitere Arbeit von den Autoren Lee & Lee (2018) wurde analysiert. Sie hat den Titel Measuring Contribution of Spatial Information to Environmental Research Using Text Mining Techniques. Die beiden Autoren machten es sich dabei zum Ziel die Trends in der umweltbezogenen Forschung quantitativ zu analysieren. Lee & Lee (2018) beabsichtigten zu analysieren, wie viele wissenschaftliche Arbeiten in Korea sich mit Umweltthemen beschäftigen. Dabei lag der Fokus besonders auf der Verwendung von Informationen, die räumliche Aspekte, wie zum Beispiel geografische Positionen oder Koordinaten enthalten. Auch in dieser Arbeit wurden für die Analyse vorerst relevante Suchbegriffe, passend für die entsprechende Domäne identifiziert, um anschließend zahlreiche Paper danach zu durchsuchen. Auf diese Weise konnte ein schneller Eindruck gewonnen werden, wie viele Paper in Korea das gefragte Thema behandeln. Darüber hinaus haben die Autoren Lee & Lee (2017) weitere Analysen durchgeführt. Sie haben analysiert in welchem Kontext, welche Schlüsselwörter zusammen auftreten. Dieser Prozess wird generell unter dem Begriff des Pattern Minings verstanden. Das Pattern Mining ermöglichte es zu berechnen, wie hoch die Wahrscheinlichkeit ist, dass ein Beitrag oder ein Paper ein bestimmtes Thema im gefragten Bereich behandelt, wenn zum Beispiel die Schlüsselwörter A und B gemeinsam darin vorkommen. Mit diesen Berechnungen können dann im Endeffekt Vorhersagen getroffen werden und bedingte Wahrscheinlichkeiten zu den jeweiligen Papers abgeleitet werden.

Aus der Arbeit von Lee & Lee (2018) ergeben sich wichtige Eckpunkte, die für die Durchführung der vorliegenden Masterarbeit relevant sind. Ähnlich wie im zuvor vorgestellten Paper geht es in der Arbeit von Lee & Lee (2018) darum, Informationen aus Texten zu extrahieren, die nicht manuell durchsucht werden können. Dies wurde erneut mithilfe einer Schlüsselwortsuche durchgeführt. Zusätzlich wurde im Paper von Lee & Lee (2018) auch die Methodik des Pattern Minings auf die Textdokumente angewendet. Diese Methode unterstützt die Schlüsselwortsuche, indem sie eine numerische Größe liefert, die die Wahrscheinlichkeit ausdrückt, dass sich ein Paper mit einem bestimmten Thema befasst, wenn eines oder mehrere Schlüsselwörter darin vorkommen.

Die Arbeit von Modaphotala & Issac (2009) legt den Fokus darauf, Umweltberichte von Unternehmen zu analysieren, insbesondere im Hinblick auf ökonomische, ökologische und soziale Leistungen. Die Umweltberichte der Unternehmen wurden dabei von deren Webseiten gesammelt, heruntergeladen und anschließend als Basis für das Text Mining verwendet. Die Methodologie ist auch in diesem Paper ähnlich zu den vorherigen. Auch im Paper von Modaphotala & Issac (2009) wurde die Schlüsselwortsuche als Methode des Text Minings eingesetzt, um relevante Informationen aus den Berichten der Unternehmen zu extrahieren und anschließend zu analysieren. Das Ergebnis durch die Textanalyse sowie die Schlüsselwortsuche war es, die quantitative Analyse von Umweltberichten von Unternehmen zu erleichtern, um Einblicke in deren wirtschaftliche, ökologische und sozialen Leistungen zu gewinnen. Ebenfalls wie im zuvor vorgestellten Paper von Lee & Lee (2018) wurde anschließend Pattern Mining angewendet, um Muster und Zusammenhänge zwischen Schlüsselwörtern und Themen in den Textdaten zu identifizieren.

Durch eine Literaturanalyse mit dem Fokus auf ähnlichen Problemstellungen und Zielen wurden zahlreiche wissenschaftliche Arbeiten gefunden. Drei Paper wurden dabei näher betrachtet, da sie die

größten Gemeinsamkeiten zur aktuellen Problemstellung aufweisen. Nach der näheren Analyse der drei Paper kann bedenkenfrei festgehalten werden, dass eine Schlüsselwortsuche beim vorliegenden Fall als Methode mit dem vielversprechendsten Potential scheint. Es geht hierbei darum Informationen aus einem Text zu extrahieren, wobei man im ersten Schritt noch nicht gänzlich sicher ist, wonach eigentlich gesucht wird (Holubliev & Simishko, 2021). Aus diesem Grund bietet die Schlüsselwortsuche für die vorliegende Problemstellung eine optimale Lösung, da auch hier zu Beginn nicht eindeutig festgelegt werden kann, nach welchen Methoden und Techniken der ANSPs überhaupt gesucht wird. Im weiteren Schritt wurde aus der Analyse abgeleitet, dass durchaus auch das Pattern Mining Sinn machen kann. Hierbei können Zusammenhänge und Muster zwischen Texten und Schlüsselwörtern identifiziert werden (Lee & Lee, 2018; Modapothala & Issac, 2009). Nichtsdestotrotz hat sich das Pattern Mining auch in den vorgestellten Papern meist erst im Laufe der Arbeit ergeben, da es je nach erhaltenen Ergebnissen Sinn machen kann, oder weniger (Lee & Lee, 2018; Modapothala & Issac, 2009). Aus diesem Grund wird im Rahmen dieser Arbeit auf jeden Fall eine Schlüsselwortsuche angewendet, um in einem ersten Schritt zu klären, ob relevante Informationen überhaupt in den Texten vorhanden sind. Die hiermit identifizierte Text-Mining-Methode der Schlüsselwortsuche wird im folgenden Kapitel genauer beschrieben, sowie deren Herausforderungen geklärt, um eine optimale Verwendung für die vorliegende Forschung zu garantieren.

3.2.4. Keyword-Analyse

Aufgrund des soeben festgestellten möglichen Potentials eine Keyword-Analyse im Rahmen dieser Arbeit zu verwenden, wird jene in diesem Kapitel genauer beschrieben. Ziel ist es, einen generellen Überblick über den Prozess einer Keyword- oder Schlüsselwortanalyse zu geben, sowie die Herausforderungen zu definieren. Mit dieser Klarstellung sollte es in Folge möglich sein, eine Keyword-Analyse im Rahmen der hier vorliegenden Problemstellung effizient durchzuführen, um schlussendlich auch die gewünschten Ergebnisse zu erhalten.

Wie bereits im vorhergehenden Kapitel angesprochen, ist die Keyword-Analyse im Bereich des Information Retrival im Text Mining anzusiedeln (Talib et al., 2016). Gemäß Talib et al. (2016) verfolgt die Keyword-Analyse im Text Mining das Ziel der Informationsextraktion. Dieser Prozess beinhaltet die Verwendung einer vordefinierten Menge an Schlüsselwörtern oder Keywords, welche im Anschluss dazu verwendet werden, spezifische Informationen aus einem Text zu extrahieren. Dies kann eine spezifische Wortsuche darstellen, bei der das Interesse darauf liegt, festzustellen, ob ein bestimmtes Wort überhaupt in einem Textdokument vorhanden ist. Alternativ kann das Keyword natürlich auch als Richtungsweiser dienen, um die Suche nach relevanten Texten zu präzisieren. In diesem Fall werden bestimmte Sätze, Passagen oder sogar ganze Textdokumente aufgrund des Vorkommens der Schlüsselwörter identifiziert.

3.2.4.1. Prozess einer Keyword-Analyse

Den Prozess einer Keyword-Analyse haben Noh et al. (2015) genauer beschrieben. In Abbildung 9 wird der Prozess visuell dargestellt. Hier sind die Stufen beziehungsweise Schritte einer Keyword-Analyse sowie weitere vier Faktoren, die den Prozess und das Ergebnis beeinflussen, zu sehen.

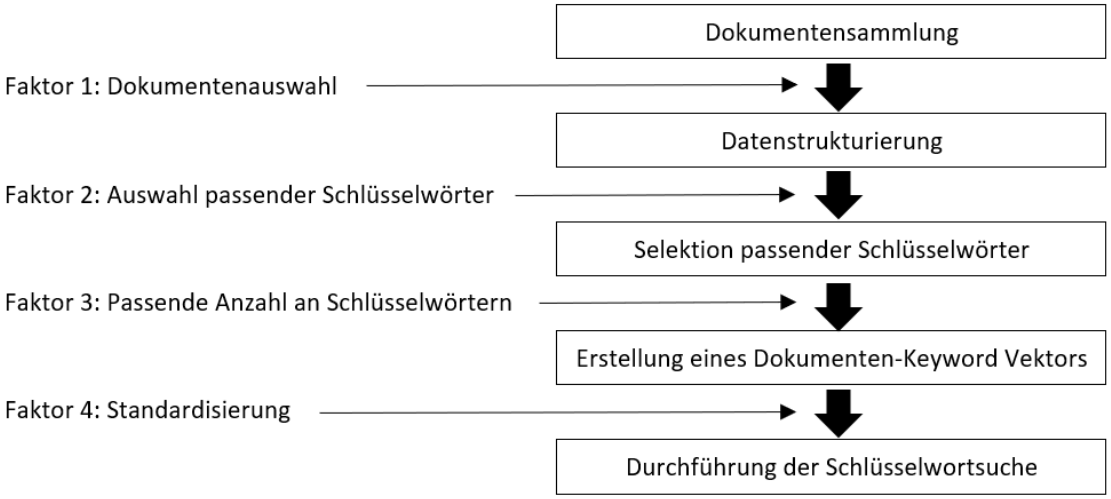


Abbildung 9: Prozess einer Keyword-Analyse

Quelle: Eigene Darstellung in Anlehnung an: Noh et al., 2015, S. 4350

Wie Abbildung 9 basierend auf Noh et al. (2015) zu erkennen ist, teilen die Autoren die Keyword-Analyse in fünf Schritte ein. Diese fünf Schritte unterliegen wiederum vier Faktoren oder Herausforderungen.

Die fünf Phasen einer Keyword-Analyse beinhalten gemäß Noh et al. (2015) die Sammlung der Dokumente, die Transformation in ein strukturiertes Format, die Selektion der passenden Schlüsselbegriffe beziehungsweise Keywords, das Erstellen eines Vektors bestehend aus der Häufigkeit an vorkommenden Schlüsselbegriffen und Textdokumenten, sowie die abschließende Durchführung der Schlüsselwortsuche.

Neben den eben erklärten fünf Prozessschritten, ist der Erfolg einer Keyword-Analyse laut Noh et al. (2015) auch stark von vier weiteren Faktoren beziehungsweise Herausforderungen abhängig. Diese vier Faktoren sind maßgeblich für die Qualität der Endergebnisse verantwortlich, weshalb es auch im Prozess dieser Arbeit besonders wichtig ist, diese vier Faktoren zu berücksichtigen und entsprechend umzusetzen Die Erfolgsfaktoren oder Herausforderungen bei einer Keyword-Analyse werden folgend näher erläutert.

3.2.4.2. Erfolgsfaktoren und Herausforderungen

Der erste Faktor stellt gemäß Noh et al. (2015) die Dokumentenauswahl dar. Dieser erste Faktor behandelt also die Wichtigkeit, die passenden Ausgangsdokumente für eine Analyse auszuwählen. Die Autoren betonen hier, dass hinterfragt werden muss wo und wie man zu den gewünschten Dokumenten kommt, wie diese aufgebaut sind und wie die notwendige Information in den Dokumenten vorhanden ist, beispielsweise als Text oder in Grafiken. Diese Gegebenheiten beeinflussen den weiteren Prozess stark, da es für unterschiedliche Dokumente unterschiedliche Vorgehensweisen und Ansätze benötigt werden.

Noh et al. (2015) definieren den zweiten Faktor als die Herausforderung der Auswahl von passenden Schlüsselwörtern. Diese können auf verschiedene Arten zusammengestellt werden. Es ist beispielsweise eine gängige Vorgehensweise, Wörter die statistisch häufig in Dokumenten vorkommen, als

relevant zu betrachten. Eine Alternative dazu ist es, Wörter und Begriffe auszuwählen, die gut zu den Hauptthemen des Dokuments passen. Im Allgemeinen sind Wörter, die häufig in Texten vorkommen, mit hoher Wahrscheinlichkeit repräsentative Schlüsselbegriffe. Allerdings gilt es hier eine Einschätzung zu treffen, denn zu häufig erscheinende Wörter und Begriffe sind in der Regel zu allgemein. Natürlich können die Begriffe auch über andere Arten wie eine Textanalyse erhoben werden. Interviews oder Absprachen mit Fachleuten und Domänenexperten sind eine weitere, häufig angewandte Methode.

Auch die Anzahl an ausgewählten Schlüsselbegriffen und Keywords ist laut Noh et al. (2015) ein entscheidender Faktor bei der Durchführung einer Keyword-Analyse. Fällt die Entscheidung auf zu viele Begriffe, ist es wahrscheinlicher, dass zahlreiche allgemeine Wörter miteinbezogen werden. Werden hingegen zu wenige verwendet, besteht die Gefahr, dass nur in bestimmten Dokumenten vorkommende Schlüsselwörter gefunden werden, was es schwierig macht, die Gesamteigenschaften der Dokumente effektiv darzustellen. Es gilt also auch diesen Faktor wiederum abzuwiegen und gegebenenfalls eine Entscheidung zu treffen oder unterschiedliche Varianten auszuprobieren und zu vergleichen.

Den vierten und letzten Faktor beschreiben Noh et al. (2015) wie in Abbildung 9 ersichtlich, als die Herausforderung der Standardisierungsmethode. Die Herausforderung besteht darin, wie die abgeleiteten Termvektoren standardisiert werden sollen. Diese Vektoren repräsentieren die Häufigkeit der ausgewählten Schlüsselwörter in den Dokumenten. Eine unangemessene Standardisierung kann zu Verzerrungen in den Analyseergebnissen führen. Es gibt verschiedene Standardisierungsmethoden, darunter keine Standardisierung, Skalierung auf einen Bereich von 0–1 oder boolesche Methoden. Die Wahl der Methode beeinflusst die Ergebnisse von Clusteranalysen, und es ist wichtig, die Auswirkungen auf die Analyseleistung zu berücksichtigen. Es kann also festgehalten werden, dass dieser Faktor stark davon abhängt, welche Analysen mit den Ergebnissen durchgeführt werden sollen. Diese letzte Herausforderung kann also durchaus entfallen, da die Standardisierung keine zwingende Herausforderung darstellt.

3.2.5. Zusammenfassung

Im aktuellen Kapitel wurde das Text Mining näher vorgestellt und beschrieben. Relevante Aspekte, welche für die vorliegende Problemstellung von Bedeutung sind, wurden ebenfalls näher erläutert. Dies beinhaltete in erster Linie den Begriff Text Mining zu erläutern sowie den Prozess des Text Minings näher zu betrachten und in einzelnen Schritten darzulegen. Im nächsten Abschnitt wurden die Methoden, welche im Text Mining Anwendung finden, näher beschrieben. Es wurde zeitnah ersichtlich, dass sich die vorliegende Forschung im Bereich der Information Extraction oder Information Retrieval bewegt. Um dies genauer einschränken zu können und eine Wahl für eine oder mehrere passende Methoden zu gewährleisten, wurde darauffolgend eine Literaturanalyse durchgeführt. Dabei wurde besonders Wert darauf gelegt, Literatur mit vergleichbaren Problemstellungen und Zielen zu finden, um mögliche Text-Mining-Methoden abzuleiten. Die Literaturanalyse brachte dabei zum Vorschein, dass sich für eine erste Analyse beziehungsweise Identifikation von relevanten Textpassagen eine Schlüsselwortsuche oder Keyword-Analyse anbietet. Diese kann verwendet werden, um relevante Wörter, Sätze oder Passagen aus Textdokumenten zu filtern. Aus diesem Grund stellt die Keyword-Analyse einen passenden ersten Schritt dar, um ANSP Techniken und Praktiken zur Redu-

zierung der Umweltbelastungen in den Textdaten zu suchen. Im letzten Schritt wurde die Keyword-Analyse näher beschrieben. Es wurde auch hier wiederum auf den Prozess eingegangen sowie noch etwas spezifischer auf die Herausforderungen bei der Durchführung einer Keyword-Analyse. Die vier beschriebenen Faktoren und Herausforderungen sind bei der Anwendung als besonders relevant anzusehen, da sie die Qualität der Ergebnisse stark beeinflussen können. Bei diesen vier Faktoren handelt es sich um die richtige Dokumentenauswahl, die passende Schlüsselwortidentifikation, die richtige Anzahl an Begriffen und Wörtern, sowie abschließend die passende Standardisierungsmethode zu wählen. Es muss allerdings festgehalten werden, dass der Prozess einer Keyword-Analyse sowie die eben erwähnten Faktoren, von der Zielsetzung einer Arbeit beeinflusst werden. Aus diesem Grund können auch Faktoren vernachlässigt werden oder der Prozess einer Keyword-Analyse an Gegebenheiten angepasst. Nach der soeben erfolgten Identifikation passender Text-Mining-Methoden, wird im nächsten Abschnitt das konkrete Vorgehen im Rahmen dieser Arbeit vorgestellt.

4. Methodik

In Abbildung 10 ist die angewandte Methodik in Rahmen dieser Arbeit dargestellt, sowie folgend die einzelnen Schritte näher beschrieben. Neben der Beschreibung der genauen Methodik, werden in diesem Kapitel auch mögliche Herausforderungen diskutiert, welche während der Durchführung der einzelnen Schritte auftreten können.

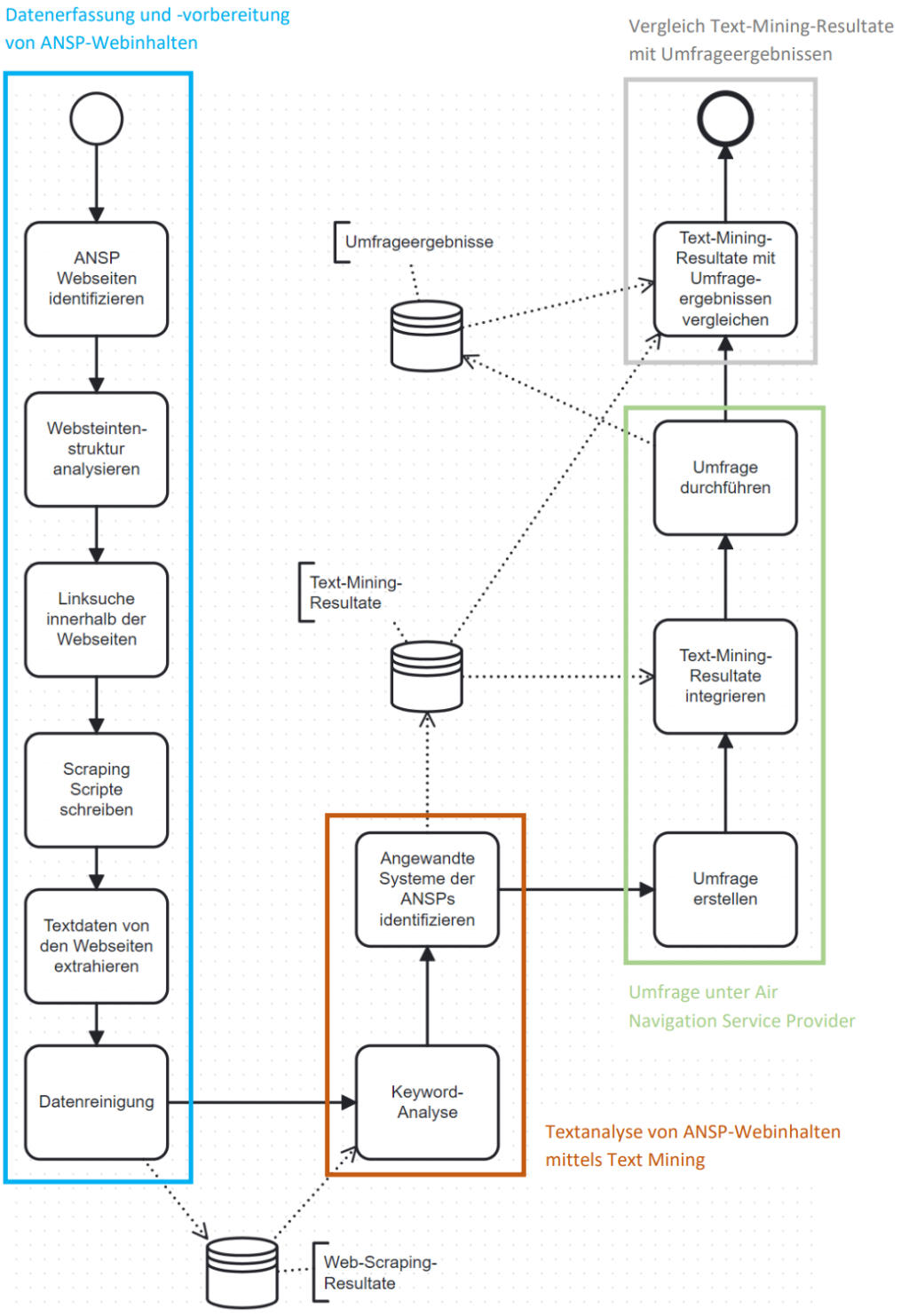


Abbildung 10: Methodologie

Quelle: Eigene Darstellung

Um die Methodologie einfacher zu veranschaulichen und zu erklären, wird sie in vier Schritte unterteilt, wobei jeder dieser Schritte, weitere spezifische Aufgaben und Tätigkeiten beinhaltet. Diese vier Schritte sowie die darin enthaltenen Aufgaben werden folgend näher beschrieben.

4.1. Datenerfassung und -vorbereitung von ANSP-Webinhalten

Wie bereits bekannt, startet der Prozess bei den Webseiten der europäischen ANSPs. Diese werden mittels einer Internetrecherche gesucht und als Ausgangspunkt der Forschung verwendet. Im ersten Schritt geht es hierbei darum, alle relevanten Links auf der eigentlichen Webseite zu identifizieren. Wie in den Grundlagen bereits erklärt, kann eine Webseite viele verschiedene Tags enthalten, darunter auch Tags welche einen Text mit einem Link zu einer anderen Seite versehen. Dies ist ebenfalls der Fall für unterschiedliche Seiten auf der gleichen Webseite. Aus diesem Grund ist es relevant, alle Links, welche auf die Zielwebseite verweisen zu extrahieren, um diese später für das Scraping zu verwenden. Herausforderungen stellen hierbei Links zu externen Webseiten dar. Es ist auf Webseiten sehr häufig der Fall, dass auch Links in der HTML-Seite eingebettet sind, welche zu externen Webseiten wie zum Beispiel Facebook, YouTube oder ähnlichen führen. Diese Seiten sind für die hier vorliegende Analyse irrelevant und sollten daher nicht als Basis für das Web-Scraping verwendet werden.

Im zweiten und nächsten Schritt werden, wie bereits vorher erwähnt, diese Links als Ausgangsbasis für das Web-Scraping verwendet. Dafür müssen die Seiten zuerst manuell analysiert werden, um ihre Struktur und den Aufbau zu verstehen. Wurde dies erledigt, können die relevanten Passagen aus den jeweiligen Links mittels des Scraping-Scripts extrahiert und gespeichert werden. Bei diesen Passagen handelt es sich rein um textuelle Informationen. Herausforderungen im eigentlichen Scraping-Schritt sind jene, dass die verschiedenen Webseiten in unterschiedlichen Sprachen teilweise eine unterschiedliche Anzahl an Informationen und Textpassagen bereitstellen. Es sollte nach Möglichkeit hierbei kein Informationsverlust aufgrund der Anzeige der Webseiten in englischer Sprache stattfinden. Die nächste Herausforderung ist, dass die abgerufenen Webseiteninformationen unstrukturiert sind. Bevor die anschließende Analyse durchgeführt werden kann, müssen diese Informationen strukturiert werden. Die dritte Herausforderung besteht in der Richtigkeit der extrahierten Textinformationen. Da die Informationen von einer Webseite gewonnen wurden, kann nicht von Beginn an zweifelsfrei sichergestellt werden, ob diese auch tatsächlich der Wahrheit entsprechen. Der Herausforderung dieser Sicherstellung der Richtigkeit der Inhalte wird sich zu einem späteren Zeitpunkt im Prozess noch genauer angenommen.

Bevor es nun aber mit der eigentlichen Analyse der extrahierten Informationen weitergeht, muss die gesammelte Datenbasis gesäubert werden. Es wurden bis zu diesem Punkt textuelle Informationen von den europäischen ANSP-Webseiten gesammelt. Hier steht außer Frage, dass auch zahlreiche irrelevante oder redundante Informationen in den Daten vorhanden sein können. Diese gilt es zu reduzieren, um den Zeitaufwand und die Effizienz der anschließenden Analysen zu optimieren. Wie allerdings bereits im theoretischen Teil zum Text Mining angesprochen wurde, gilt es hierbei die Reinigung der Datenbasis beziehungsweise deren Aufbereitung mit den anschließend zu verwendenden Methoden in Einklang zu bringen.

Abschließend wird versucht, sich einen Überblick über die gesammelten Daten zu verschaffen. Dadurch kann ein erster Eindruck gewonnen werden, welche Informationen in der gesammelten Datenbasis vorhanden sind. Dabei wird natürlich ein besonderer Fokus darauf gelegt, herauszufinden, ob relevante beziehungsweise gesuchte Informationen überhaupt in den Daten vorkommen. Um dies in einem ersten und sehr einfachen Schritt herauszufinden, bieten sich zahlreiche Analysen an. Beispielsweise wird analysiert wie viele Links und Paragraphen gesamt extrahiert wurden. Darüber hinaus kann ebenfalls investigiert werden, wie viele der extrahierten Paragraphen einen Umweltbezug in deren Inhalt aufweisen. Dadurch kann also auch in diesem Fall festgestellt werden, ob eine weitere Analyse fokussierend auf die Praktiken der ANSPs zum Umweltschutz vielversprechend ist, oder eher nicht. Zeigt der erste Eindruck der Textinformationen, dass die fortlaufende Analyse vielversprechend ist, da beispielsweise gesehen werden kann, dass Themen im Kontext des Umweltschutzes durchaus vertreten sind, kann der folgende Versuch die spezifischen Methoden und Techniken der ANSPs zu identifizieren, als vielversprechend eingestuft werden.

4.2. Textanalyse von ANSP-Webinhalten mittels Text Mining

Der zweite Schritt in der Analyse beinhaltet die Methoden des Text Minings. In diesem Schritt werden einerseits die zuvor definierten Text-Mining-Methoden auf die Datenbasis angewandt, und diese dann im Anschluss mittels Statistik und Grafiken näher beschrieben. Wie bereits aus dem Kapitel der Grundlagen bekannt, wurde mittels einer Literaturrecherche die Methode der Keyword-Analyse beziehungsweise Schlüsselwortsuche identifiziert, da sie einen vielversprechenden Ausgangspunkt für die vorliegende Problemstellung darstellt. Die Identifizierung passender Schlüsselwörter wird hierbei die größte Herausforderung darstellen. Sind diese Wörter allerdings einmal definiert, kann die Schlüsselwortsuche beginnen und anschließend die Ergebnisse ausgewertet werden, und mittels einfacher statistischen Darstellungen präsentiert. Nichtsdestotrotz sind dies noch nicht finale beziehungsweise alle Ergebnisse, welche im Rahmen dieser Arbeit zum Vorschein kommen werden. Die Identifikation der eigentlichen und spezifischen Techniken und Systemen, derzeit angewendet von den europäischen ANSPs, um den Luftverkehr umweltfreundlicher zu gestalten, stellt das gewünschte Endergebnis des Text Minings dar.

4.3. Umfrage unter Air Navigation Service Provider

Der dritte und vorletzte Schritt in der Methodologie behandelt die Erstellung sowie Durchführung einer Umfrage. Die Ergebnisse aus dem vorherigen Schritt werden hierbei in den Fragebogen integriert, und von Vertretern der europäischen ANSPs bewertet. Mittels des Fragebogens werden aber nicht nur die Ergebnisse aus dem zweiten Schritt überprüft, sondern auch ein Gesamtüberblick über die derzeitige Situation geschaffen. Dies beinhaltet zu hinterfragen, welche Methoden und Praktiken derzeit zur Reduzierung der negativen Umweltauswirkungen angewendet werden, welche Herausforderungen es bei der Implementierung von Methoden und Praktiken gibt, sowie die zukünftigen Intentionen in diesem Bereich zu hinterfragen. Durch die Integration der soeben genannten Punkte in den Fragebogen wird ein holistischer Überblick über die aktuelle Situation in diesem Kontext von der Gegenwart über in die Zukunft abgefragt.

4.4. Vergleich Text-Mining-Resultate mit Umfrageergebnissen

Den abschließenden Schritt dieser Masterarbeit stellt ein Vergleich zwischen den Text-Mining-Resultaten und den Ergebnissen der Umfrage dar. Ziel des Web-Scrapings beziehungsweise der Masterarbeit ist es, den aktuellen Stand abzuleiten, welche Maßnahmen, Methoden oder Techniken derzeit von den europäischen ANSPs angewendet werden, um die negativen Auswirkungen des Luftverkehrs auf die Umwelt zu reduzieren oder zu vermeiden. Dieser Stand wurde in einem ersten Anlauf durch das Web-Scraping und Expertenmeinungen abgeleitet, sowie folgend auch mithilfe der Durchführung der Umfrage. Somit ist sichergestellt, dass die Umfrage den tatsächlichen aktuellen Stand darstellt. In diesem letzten Schritt geht es also um die Analyse, inwiefern dieser aktuelle Stand vom Web-Scraping abgeleitet wurde. Sollten zwischen den Ergebnissen Differenzen aufgetreten sein, werden die Gründe für diese ebenfalls in diesem Schritt hinterfragt.

5. Datenerfassung und -vorbereitung von ANSP-Webinhalten

Der nun folgende Hauptschritt der Datenerhebung umfasst alle notwendigen Handlungen, um eine analysebereite Datenbasis zu generieren. Er stellt somit den ersten der vier Hauptaktionen in der soeben vorgestellten Methodologie dar. Bevor die Informationen der Webseiten extrahiert und zusammengetragen werden können, bedarf es allerdings einiger Vorbereitungen. Der Fokus liegt im folgenden ersten Schritt darauf, die zu verwendenden Webseiten der ANSPs zu identifizieren sowie hinsichtlich deren Aufbau und Struktur zu analysieren.

5.1. Datenidentifikation und Strukturanalyse

Der erste Schritt umfasst die Identifikation der europäischen ANSP Webseiten sowie eine Analyse des Aufbaus der einzelnen Webseiten. Dies ist erforderlich, um anschließend die richtigen Passagen und Inhalte der Webseiten an das Scraping-Script übergeben zu können. Die Identifikation der europäischen ANSPs mit deren Webseiten wurde mittels einer Internetrecherche durchgeführt. Die anschließende Analyse der Struktur und des Aufbaus der Webseiten wurde, wie in den Grundlagen erläutert, entweder über die Investigation des Seitenquelltextes oder mithilfe des Selector-Gadets durchgeführt (Wickham, o. J.). Diese Erstanalyse verschaffte somit einen Überblick über den Aufbau der Webseiten und dadurch Einsicht, in welchen Tags der HTML-Webseiten die relevanten Textinformationen gespeichert sind. Dies ist eine essenzielle Information, welche dem Scraping-Script anschließend übergeben wird. Mithilfe dieser Informationen werden im Anschluss die richtigen Passagen innerhalb der spezifizierten Webseiten identifiziert und extrahiert. Das Ergebnis der Identifikation der Webseiten sowie der Analyse der Struktur sind in Tabelle 2 ersichtlich. Alle für das Web-Scraping verwendete Webseiten sind in der folgenden Tabelle aufgelistet. Die wichtigste Information für das Scraping ist in der letzten Spalte zu finden. Hier wird vermerkt, in welchen Tags die Textinformationen auf den einzelnen Webseiten gespeichert werden.

Tabelle 2: Verwendete ANSP Webseiten

ANSP	Homepage	Land	Tags
Albcontrol	http://www.albcontrol.al/	Albanien	Paragraph
ANA LUX	https://ana.gouvernement.lu/	Luxemburg	Paragraph
ANS CR	www.rlp.cz	Tschechische Republik	Paragraph
ARMATS	www.armats.am	Armenien	Paragraph
Austro Control	www.austrocontrol.at	Österreich	Paragraph, Span
Avinor Flysikring	www.avinor.no	Norwegen	Paragraph
BHANSA	https://www.bhansa.gov.ba	Bosnien Herzegowina	Paragraph
BULATSA	www.bulatsa.com	Bulgarien	Paragraph
Croatia Control	www.crocontrol.hr	Kroatien	Paragraph
DFS	www.dfs.de	Deutschland	Paragraph
EANS	www.eans.ee	Estland	Paragraph
ENAIRE	www.enaire.es	Spanien	Paragraph

ENAV	www.enav.it	Italien	Paragraph
Fintraffic ANS	www.fintraffic.fi/en/ans	Finnland	Paragraph
HungaroControl	www.hungarocontrol.hu	Ungarn	Paragraph
AirNav Ireland	www.airnav.ie/	Irland	Paragraph
LFV	www.lfv.se	Schweden	Paragraph
LGS	www.lgs.lv	Lettland	Paragraph
LPS	www.lps.sk	Slowakei	Paragraph
LVNL	www.lvnl.nl	Niederlande	Paragraph
MATS	www.maltats.com	Malta	Paragraph
NATS	www.nats.aero	United Kingdom	Paragraph
NAVIAIR	www.naviair.dk	Dänemark	Paragraph
PANSA	www.pansa.pl	Polen	Paragraph
Sakaeronavigatsia	www.airnav.ge	Georgien	Paragraph
Skeyes	www.skeyes.be	Belgien	Paragraph
Skyguide	www.skyguide.ch	Schweiz	Paragraph
Slovenia Control	www.sloveniacontrol.si	Slowenien	Paragraph, Div
SMATSA	http://www.smatsa.rs	Serbien und Montenegro	Paragraph
BELAERONAVIGATSIA	www.ban.by	Belarus	Paragraph
ISAVIA	www.isavia.is	Island	Paragraph
MNAV PORTUGAL	www.nav.pt	Portugal	Paragraph

Nach dieser ersten Analyse kann nun die obenstehende Tabelle für das weitere Web-Scraping verwendet werden, da sie dafür alle nötigen Informationen enthält. Wie Tabelle 2 zeigt, speichert der Großteil der Webseiten die Textinformationen in Paragraph, also <p>-Tags. Dies stellt einen vorteilhaften Faktor für das Scraping-Script dar, da dies bedeutet, dass das Script einmal verfasst wird und anschließend ohne hohen Anpassungsaufwand auf so gut wie alle Webseiten der europäischen ANSPs angewendet werden kann. Dies führt wiederum zu einem höheren Automatisierungsgrad, im Vergleich zu einem Script das händisch für jede Webseite spezifiziert werden muss. In diesem Fall bedarf es also nur bei zwei der 32 Webseiten einer Anpassung beziehungsweise einer Erweiterung der zu extrahierenden Passagen um die identifizierten Tags. Bei der österreichischen Austro Control handelt es sich um die Elemente und <p>, bei der slowenischen Slovenia Control um die Tags <p> und . Dies sind daher alle Anpassungen, welche im Prozess berücksichtigt werden müssen. Bei den restlichen Webseiten sollte es ausreichend sein die Paragraph-Tags zu extrahieren.

Es gilt allerdings anzumerken, dass die obenstehende Tabelle nur jene Webseiten von europäischen ANSPs enthält, welche auch für den Zweck der Arbeit und das Web-Scraping sinnvoll beziehungsweise möglichen sind. Teilweise waren Webseiten nicht auf dem aktuellsten Stand, in der Zielsprache Englisch nicht verfügbar oder überhaupt nicht erreichbar. Aus diesen Gründen ist es naheliegend, derartige Webseiten vom Web-Scraping-Prozess auszuschließen. Während des eigentlichen

Scraping-Prozesses kamen mehrere Webseiten zum Vorschein, welche aufgrund unterschiedlicher Gegebenheiten nicht verwendet werden konnten. Beispielsweise aufgrund von diversen Fehlermeldungen während des Prozesses oder aufgrund von einer unverhältnismäßigen Komplexität der Webseitenstruktur. Die zusammenfassende Tabelle jener ANSPs, welche aus den genannten Gründen unverwendet bleiben, ist untenstehend einzusehen.

Tabelle 3: Fehlerhafte Webseiten

ANSP	Homepage	Land	Fehler
AZANS	www.azans.az	Aserbaidschan	Webseite nicht erreichbar
DCAC Cyprus	www.mcw.gov.cy/dca	Zypern	Komplexe Struktur und inkonsistente Sprache
DSNA	www.aviation-civile.gouv.fr	Frankreich	Keine englische Übersetzung
MOLDATSA	www.moldatsa.md	Moldawien	Authentifizierungs- und Zugriffsfehler
Oro Navigacija	www.ans.lt	Litauen	Keine englische Übersetzung
UkSATSE	www.uksatse.ua	Ukraine	Konvertierungsfehler im Scraping-Script
ROMATSA	www.romatsa.ro	Rumänien	Zahlreiche Links nicht erreichbar(Fehler 404) und viele Seiten ohne Informationsgehalt
HCAA	http://www.hcaa.gr/	Griechenland	Komplexe Struktur
EUROCONTROL	https://www.eurocontrol.int/	Mehrere	Blockierung der Scraping Zugriffe

Zusammengefasst bleiben also 32 geeignete Webseiten zur weiteren Analyse übrig. Wie dieser Prozess funktioniert und wie die Datenbeschaffung im Detail aussieht, wird im nächsten Abschnitt erläutert.

5.2. Web-Scraping

Wie bereits während der Beschreibung der Methodologie bekannt wurde, bedarf es bei der Datenbeschaffung einiger Vorbereitung. Dies beinhaltet speziell die Aufbereitung der zu verwendenden Links sowie der Klarstellung des Aufbaus der einzelnen Webseiten, um die gesuchten Textpassagen möglichst effizient zu extrahieren. Nachdem die Webseiten der europäischen ANSPs identifiziert sind, kann mit der Sammlung der relevanten Links begonnen werden.

5.2.1. Link-Scraping

In einem ersten Schritt geht es hierbei darum, alle relevanten Links auf den Webseiten der ANSPs in Europa zu identifizieren. Diese Links werden dann anschließend für die Extrahierung der Textinformationen verwendet. Dieser Schritt ist besonders relevant, da von den verwendeten Links anhängig ist, auf welche Webseiten und damit auch auf welche Textinformationen das Scraping-Script zugrei-

fen kann. Durch fehlende Links innerhalb einer Webseite können wichtige Informationen verloren gehen.

5.2.1.1. Zielsetzung

Wie bereits in den Grundlagen angesprochen, werden Links in einer HTML-Webseite in einem <a>-Tag mit dem Attribut „href“ gespeichert (Mine & Mine, 2021). Dies bedeutet also, dass in einem ersten Schritt alle verfügbaren Links auf der Homepage jedes ANSPs mittels dieses Tags und Attributs extrahiert werden, und anschließend diese extrahierten Links als Ausganspunkt genommen werden, um wiederum nach weiteren Links auf den Webseiten zu suchen. Warum diese Vorgehensweise nötig ist, kann in Abbildung 11 eingesehen werden. Hier wir die Link-Struktur innerhalb einer HTML-Webseite am Beispiel der österreichischen Flugsicherungsorganisation Austro Control dargestellt.

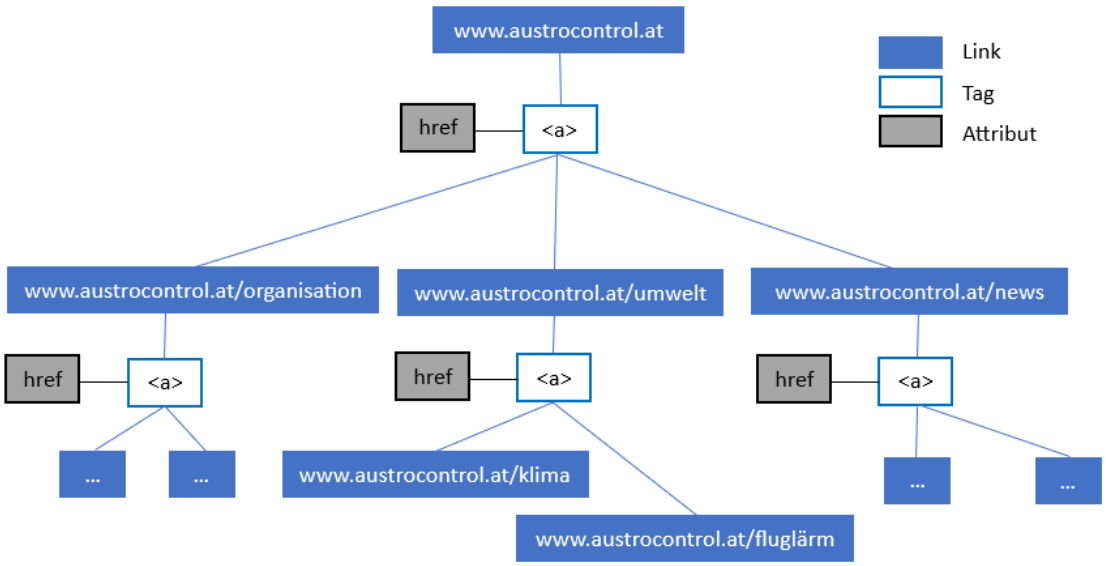


Abbildung 11: Link-Struktur einer HTML-Webseite

Quelle: Eigene Darstellung

Wie in Abbildung 11 zu sehen ist, wird für jeder Webseite von der Basis-URL ausgegangen und dort nach den entsprechenden Tags und Attributen gesucht. Dies identifiziert wiederum weitere Links, welche dann erneut als Ausgangs-URL dienen, um erneut nach weiteren Links auf der spezifischen Webseite zu suchen. Dieser Prozess wird solange wiederholt, bis keine neuen Links mehr gefunden werden können. Durch diese Vorgehensweise wird garantiert, alle Links unter einer Webseite zu finden, und nicht nur jene, welche unter dem Ausgangslink, also der Homepage, zu finden sind. Zu beachten war hierbei, dass auch nur Links beziehungsweise URLs mit derselben Domäne verwendet werden. Im Fall des soeben vorgezeigten Beispiels würde dies bedeuten, dass nur Webseiten mit der Domäne „www.austrcontrol.at“ extrahiert werden. Dadurch wird sichergestellt, dass keine Links zu externen Webseiten wie beispielsweis YouTube, Facebook oder ähnlichem verwendet wurden.

5.2.1.2. Funktionsbeschreibung

Der Code hierfür wurde in der Programmiersprache R verfasst unter der Verwendung der Bibliothek „rvest“ (RDocumentation, o.J.). Rvest wird direkt in der Programmierumgebung R installiert, und wird zur Datenextrahierung aus HTML-Webseiten verwendet (RDocumentation, o.J.). Die Bibliothek bietet

die Funktionalität, Webseiten zu lesen und anschließend Texte, Tabellen oder Bilder von Webseiten zu identifizieren und zu extrahieren (RDocumentation, o.J.). Es können mit diesem Package also strukturierte Daten aus HTML-Code einer Webseite extrahiert werden und anschließend in die Programmierumgebung geladen, und für weitere Analysen und Visualisierungen verwendet werden (RDocumentation, o. J.). Der Code für das Link-Scraping sieht folgendermaßen aus.

Listing 1: Link-Scraping-Script

```
# Funktion um alle Links innerhalb einer Domäne zu erhalten
getAllUniqueLinksWithinDomain <- function(url, domain, unique_url_list) {
  print(paste("Neue Anfrage wird gesendet", url))

  # HTML-Inhalt der URL lesen
  page <- read_html(url)

  # Alle 'a' Elemente mit 'href' Attribut finden
  links <- page %>% html_nodes("a[href]")

  #Überprüfen ob Links vorhanden sind
  if (length(links) > 0) {
    # Iteration über alle gefundenen Links
    for (link in links) {
      #Extraktion der 'href' Attribute
      href <- link %>% html_attr("href")

      #Wenn die Domain passt, hinzufügen der Domain
      if (startsWith(href, "/") && domain == "https://www.austrocontrol.at/en") {
        href <- paste0("https://www.austrocontrol.at", href)
      }

      #Überprüfen, ob der Link zur angegebenen Domain gehört
      if (startsWith(href, domain)) {
        if (!(href %in% unique_url_list)) {
          print("Nicht in der Liste.")
          print(href)
          unique_url_list <- c(unique_url_list, href)

          #Bestimmte Links werden direkt aussortiert
          if (!grepl(".pdf|.mp4|\\?|.png", href)) {
            unique_url_list    <-    getAllUniqueLinksWithinDomain(href,    domain,
unique_url_list)
          }
        }
      }
```

```
    }
  }
  return(unique_url_list)
}
```

Wie Listing 1 darstellt, dient die Funktion „getAllUniqueLinksWithinDomain“ dazu, alle eindeutigen Links innerhalb einer bestimmten Domäne zu extrahieren. Der Funktion werden hierfür folgende drei Parameter übergeben.

- url: Die URL, von der aus der Prozess gestartet wird. Hierbei handelt es sich um die Homepage der spezifischen europäischen ANSPs.
- domain: Die Ziel-Domäne innerhalb welcher die Links gesammelt werden sollen.
- unique_url_list: Eine zu Beginn leere Liste, in der die eindeutigen Links gesammelt werden und anschließend als Ergebnis von der Funktion zurückgegeben werden.

Die Funktion durchläuft die folgenden Schritte. Zunächst gibt die Funktion eine Meldung aus, dass eine neue Anfrage gesendet wird, und zeigt die aktuelle URL an. Anschließend liest sie den HTML-Inhalt der angegebenen URL und findet alle 'a'-Elemente mit einem 'href'-Attribut auf der Seite.

Die Funktion überprüft, ob die gefundenen Links innerhalb der angegebenen Domain liegen. Wenn dem so ist und ein gefundener Link noch nicht in der Liste der eindeutigen Links enthalten ist, wird er zur Liste hinzugefügt. Die Funktion nutzt anschließend die identifizierten Links als neue Ausgangslinks, um auch auf diesen Webseiten nach weiteren Links zu suchen, welche den spezifizierten Kriterien entsprechen.

Um die Liste der eindeutigen Links zu optimieren, filtert die Funktion bestimmte Arten von Links heraus, wie beispielsweise E-Mail-Schutzlinks, PDFs, MP4, Bilder und bestimmte Dateiformate. Derartige Links wurden im Scraping Prozess häufig angetroffen, weshalb sie von Beginn an exkludiert werden.

Abschließend gibt die Funktion die aktualisierte Liste der eindeutigen Links zurück. Ihr Hauptziel besteht darin, sämtliche eindeutige Links innerhalb der angegebenen Domäne zu sammeln, wobei sie rekursiv die verlinkten Seiten durchgeht und dabei bestimmte Arten von Links ausschließt, um eine präzise Liste zu generieren.

5.2.1.3. Anwendung und Ergebnis

Untenstehend ist die beispielhafte Anwendung sowie der Aufruf dieser Funktion ersichtlich. In diesem Beispiel wurde die Funktion auf die Webseite „https://www.austrocontrol.at/en“ angewendet. Es handelt sich hierbei um die englischsprachige Webseite der österreichischen Flugsicherungsorganisation Austro Control. In Listing 2 sind die übergebenen Parameter an die Funktion ident. Natürlich scheint dies in erster Linie redundant, ermöglicht aber für andere Anwendungsfälle ein erhöhten Grad an Flexibilität. Es ermöglicht dem Benutzer oder der Benutzerin, verschiedene Start-URLs und Domänen zu verwenden, wenn es der spezifische Anwendungsfall erfordert. Dies war durchaus auch bei zahlreichen Webseiten der europäischen ANSPs nötig.

Listing 2: Anwendung Link-Scraping-Script

```
#Parameter für die Funktion
base_url <- "https://www.austrocontrol.at/en"
domain <- "https://www.austrocontrol.at/en"

#Funktionsaufruf
unique_links <- getAllUniqueLinksWithinDomain(base_url, domain, character(0))

#Filterung und Löschung bestimmter Linktypen
unique_links<-
unique_links[!grepl("mp4|pdf|problems|png|jpg|twitter|facebook|linkedin|instagram
", unique_links)]

#Speicherung der Links in einer Textdatei in einem Verzeichnis
file_path <- "Pfad zum Verzeichnis zur Speicherung"
writeLines(unique_links, file_path)
```

Zusammenfassend sammelt die oben gezeigte Beispielanwendung alle eindeutigen Links von der Webseite https://www.austrocontro.at/en, filtert bestimmte Link-Typen aus dem Endergebnis heraus und speichert die bereinigte Liste an Links in einer Textdatei ab. Das Ergebnis im Link-Scraping stellt also ein Textdokument dar, mit all den identifizierten Links zu einer Webseite innerhalb der angegebenen Domäne. Diese Textdatei wird im nächsten Schritt verwendet, um die Textinhalte dieser Links zu extrahieren.

Nachdem das Link-Scraping nun auf die knapp 40 Webseiten der europäischen ANSPs angewendet wurde, werden die identifizierten URLs im nächsten Schritt als Basis für die Extraktion der Textinformationen verwendet.

5.2.2. Text-Scraping

Dieser Schritt befasst sich mit der Aufgabe, die eigentlichen Textdaten für eine anschließende Analyse zur Verfügung zu stellen. Wie bereits in den Grundlagen und zu Beginn des aktuellen Kapitels geklärt, bedarf es hierfür einer Investigation der Webseiten, um den Aufbau der Seiten zu verstehen, sowie die dafür benötigten Links innerhalb der entsprechenden Domäne. Mit diesem bereits zuvor gesammeltem Wissen, werden in diesem Kapitel die Textpassagen extrahiert.

5.2.2.1. Zielsetzung

Das Ziel für den Schritt des Text-Scrapings lässt sich kurz und prägnant zusammenfassen. Die Methode soll ein Script oder Codestück bereitstellen, dass die relevanten Tags der HTML-Webseiten, sowie die zuvor identifizierten Links als Parameter übergeben bekommt. Anschließend werden die relevanten Textabschnitte extrahiert und in die Programmierumgebung, in diesem Fall R, geladen. Dies sollte möglichst eindeutig erfolgen, was bedeutet, dass so gut wie möglich sichergestellt werden muss, dass idente Textpassagen nur einmal vom Scraping-Script erkannt und extrahiert werden. Dies stellt sicher, dass die Ergebnisse wiederum eindeutig und nicht verfälscht sind und spart ebenfalls Zeit und Aufwand bei der Säuberung der Textdokumente.

Um die Funktionsweise des Scraping Algorithmus besser zu illustrieren und zu verstehen, wird im Folgenden Beispiel angenommen, dass eine Webseite die Textpassagen in einen <p>, also Paragraph-Tag, speichert. Gemäß Diouf et al. (2019), würde der Scrpaing Algorithmus hier folgendermaßen funktionieren. Die Funktion durchläuft den Aufbau der HTML-Webseite von oben nach unten durch. Alle relevanten Passagen, in diesem Fall die <p>-Tags, werden extrahiert und in die Programmierumgebung geladen. Von dort aus können die Texte je nach Bedarf gespeichert werden und weitere Analysen durchgeführt (Diouf et al., 2019). Die Funktion ist in Abbildung 12 grafisch dargestellt. Bei Identifizierung eines <p>-Tags wird der Inhalt dessen extrahiert und gespeichert.

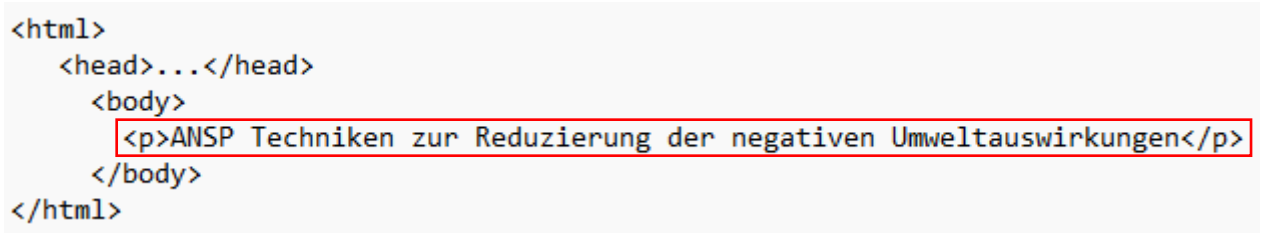


Abbildung 12: Funktionalität Text-Scraping-Script

Quelle: Eigene Darstellung

5.2.2.2. Funktionsbeschreibung

Auch der Code für das Text-Scraping wurde mittels der Programmiersprache R unter der Verwendung des Packages rvest programmiert. Das Script kann in Listing 3 eingesehen werden.

Listing 3: Text-Scraping-Script

```
#Text-Scraping Funktion
scrape_text <- function(url) {
  tryCatch({
    webpage <- read_html(url)
    p_elem <- webpage %>% html_nodes("p") %>% html_text()

    #Konventierung
    p_elem <- unlist(p_elem)

    #Whitespaces entfernen
    p_elem <- trimws(p_elem)

    return(p_elem)
  }, error = function(e) {
    cat(paste("Error scraping", url, ":", conditionMessage(e), "\n"))
    return(NULL)
  })
}
```

Wie in Listing 3 dargestellt, bekommt die Funktion „scrape_text“ einen Parameter übergeben.

- url: diese Information beinhaltet eine URL einer Webseite, von welcher der Text extrahiert wird

Auch die Funktionalität dieses Codes beruht wiederum auf der Beschreibung der RDocumentation (o.J) sowie der verwendeten „rvest“ Bibliothek. Die Funktion beginnt damit, die HTML-Struktur der angegebenen URL mithilfe der „read_html“ Funktion zu laden. Anschließend werden die <p> Paragraph-Tags auf der geladenen Webseite identifiziert. Mithilfe von „html_text“ werden die Textinhalte der identifizierten Tags extrahiert und in der Variable „p_elem“ gespeichert. Abschließend wird dann noch der extrahierte Text in einen Vektor umgewandelt, um die Handhabung zu erleichtern. Dies geschieht mit Hilfe der „unlist“ Funktion. Anschließend werden dann noch mit der Funktion „trimws“ überflüssige Leerzeichen in den Textdokumenten entfernt. Die bereinigten und extrahierten Textelemente werden als das Ergebnis der Funktion zurückgegeben. Die Funktion ist robust gegenüber möglichen Fehlern beim Scrapen. Mit „tryCatch“ wird sichergestellt, dass das Script bei auftretenden Fehlern nicht abbricht, sondern stattdessen eine beschreibende Fehlermeldung ausgegeben wird. Die Fehlermeldung enthält Informationen über die fehlerhafte URL und die Art des aufgetretenen Fehlers.

5.2.2.3. Anwendung und Ergebnisse

In Listing 4 kann der Code und dessen genauere Beschreibung für die Anwendung der eben erklärten Funktion eingesehen werden. Der Beispielaufruf der Funktion wird wiederum mit einem Beispieldokument, den Links der österreichischen Austro Control, dargestellt. Das Textdokument mit den darin enthaltenen Links wurde bei jedem Aufruf entsprechend für die spezifischen ANSPs angepasst.

Listing 4: Anwendung Text-Scraping-Script

```
#Einlesen der Links je ANSP aus einer Textdatei
setwd(Pfad zum Arbeitsverzeichnis mit den Link-Textdokumenten)
uniqueLinks <- readLines("5_austria_links.txt")

#Initiierung Ergebnisliste
unique_p_elements <- character(0)

# Schleife durch alle Links im Dokument
for (url in uniqueLinks) {
  p_elems <- scrape_text(url)

  #Überprüfen, ob Text erfolgreich extrahiert wurde
  if (!is.null(p_elems)) {
    unique_p_elements <- unique(c(unique_p_elements, p_elems))
  }
}
# Kombinieren der Paragraphen zu einem Gesamtdokument
all_content <- paste(unique_p_elements, collapse = "\n")
```

Zu Beginn des Codestückes wird das Verzeichnis, in welchem die Dokumente mit den Links gespeichert sind, festgelegt. Die Links werden dann aus den jeweiligen Textdateien, im Beispielfall am ANSP Austro Control aus Österreich, eingelesen und in der Variable „uniqueLinks“ gespeichert. Im nächsten Schritt wird zunächst eine leere Ergebnisliste namens „unique_p_elements“ erstellt. In dieser Liste werden die extrahierten Paragraphen von den verschiedenen Webseiten gespeichert. Anschließend wird eine Schleife durchlaufen, die jeden Link aus der eingelesenen Liste nacheinander durchläuft. Die Funktion „scrape_text“ wird auf jeden dieser Links angewendet, und die extrahierten Paragraphen werden in der Variable „p_elems“ gespeichert. Im vorletzten Schritt wird überprüft ob der Text erfolgreich extrahiert wurde. Falls ja, werden die extrahierten Textstücke zur Liste „unique_p_elements“ hinzugefügt, wobei Duplikate durch die „unique“ Funktion vermieden werden sollten. Schlussendlich werden alle eindeutigen Paragraphen zu einem Gesamtdokument kombiniert, wobei Zeilenumbrüche zwischen den Paragraphen eingefügt werden.

Zusammenfassend kann gesagt werden, dass diese Code-Anwendung einen iterativen Prozess der Textextraktion von verschiedenen Webseiten repräsentiert, wobei die Ergebnisse zu einem kohärenten Gesamtdokument je europäischem ANSP kombiniert werden. Diese Dokumente werden anschließend als Grundlage für weiterführende Analysen und Forschung im Rahmen der Masterarbeit dienen.

Zu den Ergebnissen kann also nochmals wiederholt werden, dass es sich hierbei wie auch zuvor bei den Links, um Textdokumente handelt. In diesem Fall befinden sich alle extrahierten Textpassagen, getrennt durch Zeilenumbrüche, im Textdokument. Für jeden der 32 ANSPs ist ein derartiges Textdokument zur weiteren Analyse vorhanden.

Wie die gesammelte Datenbasis nun im Detail aussieht und welche Schlüsselmerkmale sie aufweist, wird im nächsten Abschnitt systematisch beleuchtet. Hierbei wird ein genauerer Blick auf die Datenzusammensetzung und die Struktur der Informationen geworfen, sowie Aussagen über die Datenqualität und etwaige Herausforderungen abgeleitet.

5.3. Datenbeschreibung

Wie schon in den Vorbereitungen erwähnt, wurden die Textinformationen von 32 europäischen ANSP Webseiten erfolgreich zusammengetragen. Die Dateien sind natürlich aufgrund der unterschiedlichen Webseiten und deren Informationszugang sehr unterschiedlich. Nichtsdestotrotz wird der Abschnitt der Datenbeschreibung dafür verwendet, um einen Überblick über die Datenbasis zu erhalten. Dies beinhaltet statistische Zusammenfassungen über die extrahierten Daten, eine Erklärung der Struktur und des Formats der Daten, sowie eine Aussage über Die Datenqualität zu treffen. Letzterer Punkt kann folgend Aufschluss darüber geben, inwiefern und ich welchem Ausmaß eine Datenreinigung relevant sein wird.

5.3.1. Datensatzübersicht

In diesem Abschnitt wird ein genereller Überblick über die Datenbasis gegeben. Es geht hierbei also darum einen ersten Eindruck über die gesammelten Daten zu erlangen. Vorrangig wird hier eine statistische Auswertung der Daten vorgenommen. Die Datensatzübersicht ist entscheidend, um einen schnellen Überblick über die grundlegenden Charakteristika der gesammelten Daten zu erhalten.

Im ersten Schritt kann hierbei aufgezeigt werden, wie viele Webseiten überhaupt pro europäischem ANSP verwendet wurden. Abbildung 13 zeigt die Anzahl an Webseiten, also extrahierten Links, für jeden ANSP sowie die dazugehörige Anzahl an extrahierten Paragraphen.

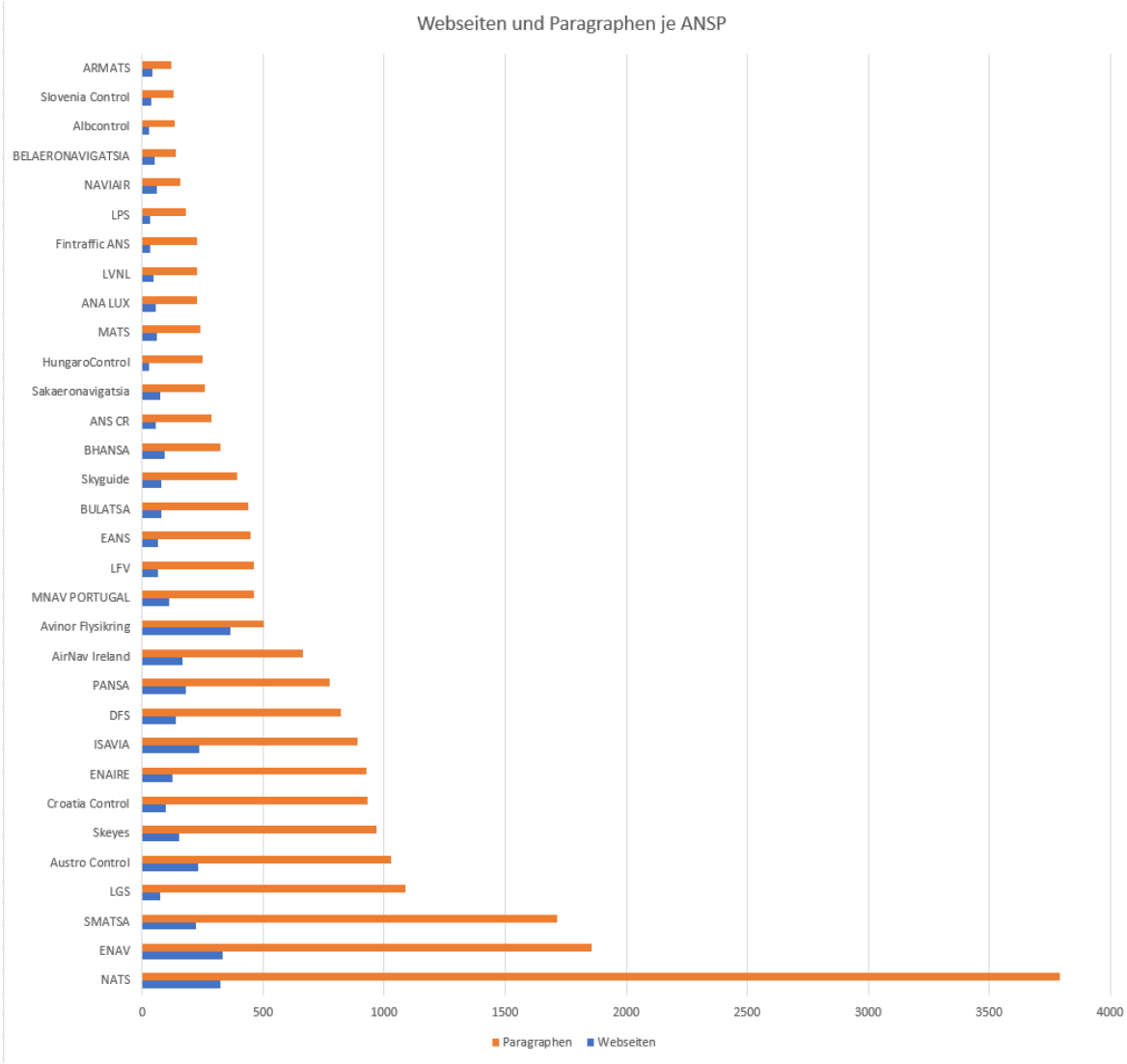


Abbildung 13: Anzahl gesammelte Links und Paragraphen je Webseite

Quelle: Eigene Darstellung

Insgesamt wurden 21.042 Paragraphen beziehungsweise entsprechende Tags von den 32 Webseiten extrahiert. Innerhalb der 32 Homepages wurden gesamt 3.746 Links beziehungsweise weiterführende Webseiten analysiert und deren Textinformationen extrahiert. Diese Informationen lassen also auf eine Textdichte pro Link von circa 5,62 Paragraphen schließen.

Besonders bemerkenswert ist die Flugsicherungsorganisation NATS aus dem United Kingdom, der mit über 3.500 gesammelten Paragraphen und 323 extrahierten Links den größten Anteil in der Datenbasis einnimmt. Dies könnte auf einen besonders umfangreichen Informationsgehalt pro Link oder auf eine spezifische Struktur der Webseite hinweisen. Des Weiteren deutet die hohe Anzahl an extrahierten Paragraphen und Links bei NATS darauf hin, dass pro Webseite viele einzelne Paragraphen relevante Informationen enthalten könnten. Dies könnte auf eine besonders detaillierte Informationsstruktur der Webseite hindeuten. Im Sinne der gewonnen Paragraphen liegt also die NATS klar

im Spitzenfeld. Auch die 323 extrahierten Links sind eindrucksvoll, auch wenn dies kein Einzelfall im Vergleich zu anderen ANSPS darstellt. Die Ergebnisse der NATS werden also noch weiteren manuellen Analysen unterzogen, um sicherzustellen, dass es sich hierbei nicht um mögliche Fehler oder ähnliches im Scraping Prozess handelt. Die restlichen ANSPs ergeben aber keinen Grund zur Sorge. Natürlich unterscheiden sich Anzahl der extrahierten Paragraphen und Links teilweise sehr, dies ist aber größtenteils rückzuführen auf unterschiedliche Webseitenstrukturen sowie unterschiedliche Informationsauskunft.

Abbildung 14 repräsentiert die durchschnittlichen Paragraphen je Webseite sowie der Gesamtdurchschnitt an Paragraphen je Link über alle ANSPs einzusehen.

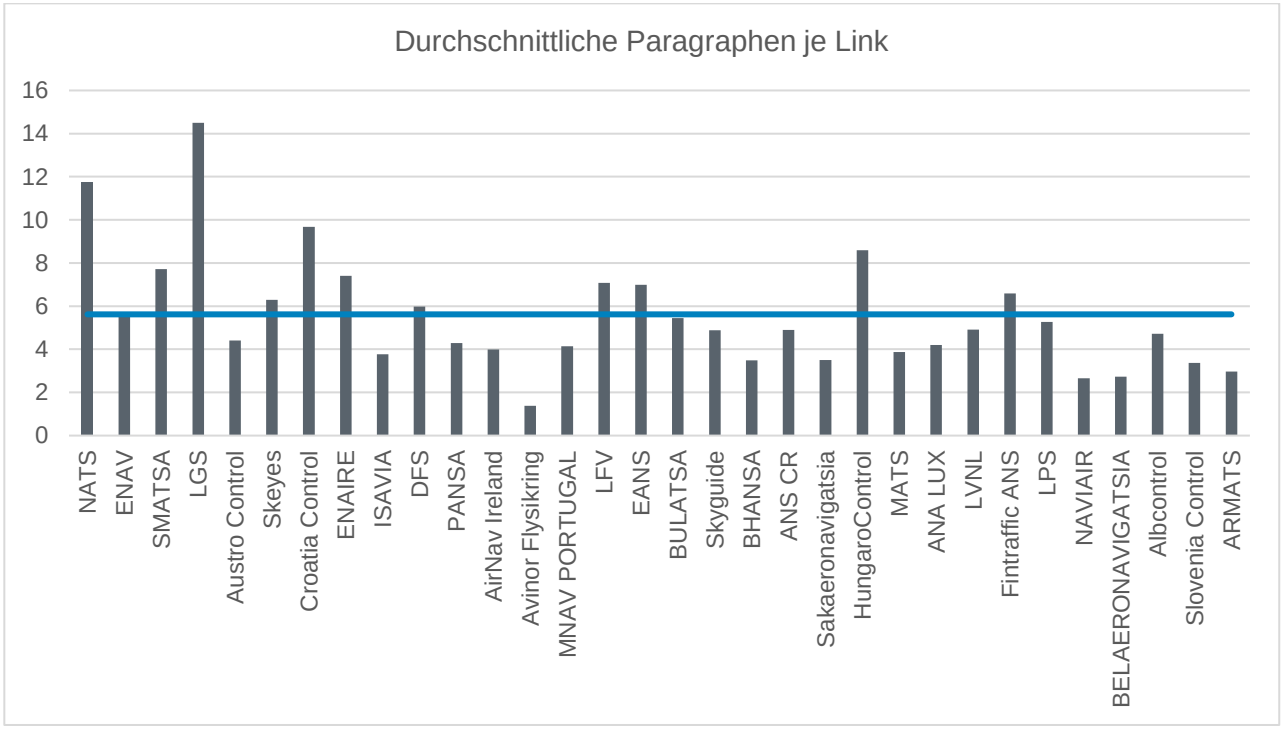


Abbildung 14: Durchschnittliche Paragraphen je Link und Gesamtdurchschnitt

Quelle: Eigene Darstellung

Ein genauer Blick auf Abbildung 14 zeigt, dass die durchschnittliche Anzahl der Paragraphen pro Webseite je nach ANSP variiert. Besonders hervorstechend ist hierbei wieder der ANSP NATS sowie auch der lettische ANSP LGS. Mit einer durchschnittlichen Anzahl an Paragraphen von 11,75 bei NATS sowie von 14,50 bei LGS, zeigt dies einen signifikanten Ausschlag nach oben in diesen Zahlen. Dies steht im Kontrast zu den anderen ANSPs, bei denen die durchschnittliche Anzahl pro Webseite sehr viele näher am Gesamtdurchschnitt von etwa 5,50 liegt. Diese Unterschiede in der durchschnittlichen Verteilung könnten auf unterschiedliche Strukturen der Webseiten oder Informationspraktiken der entsprechenden ANSPs hinweisen. Des Weiteren ist auffallend dass das norwegische ANSP mit nur 1,38 Paragraphen deutlich unter dem Durchschnitt der restlichen ANSPs liegt. Dies kann auf unterschiedliche Aspekte hindeuten. Beispielsweise wiederum auf eine knappe Informationsstruktur, spezialisierte Inhalte oder vermehrt großen, aber dafür einzelnen Paragraphen auf der Webseite.

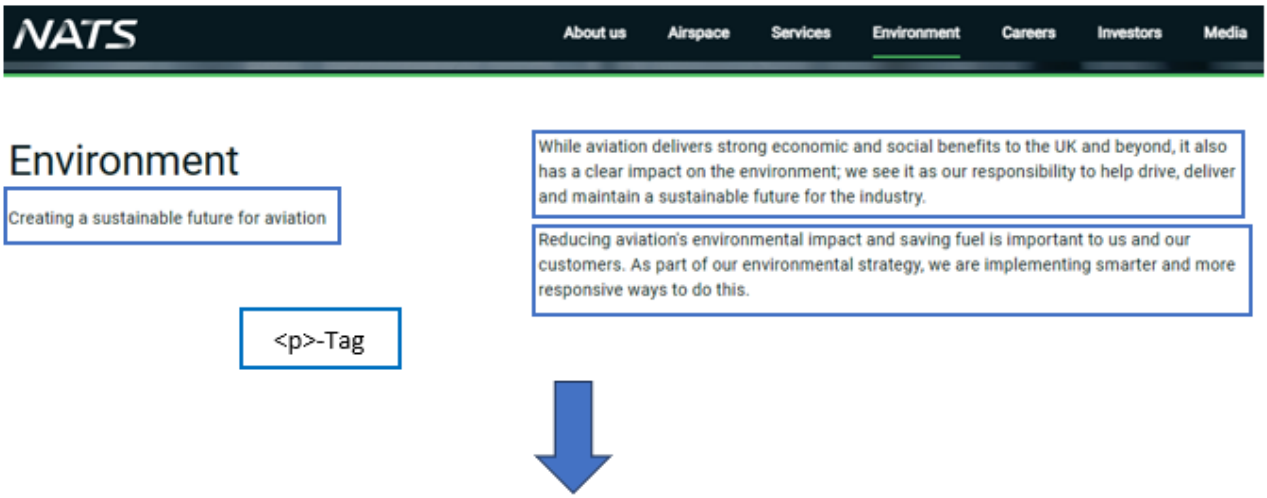
Aufbauend auf den Erkenntnissen aus der Datensatzübersicht wird nun die Struktur sowie das Format der gesammelten Daten analysiert.

5.3.2. Struktur und Format

Die Struktur der erhobenen Daten wurde im Verlauf dieses Prozesses weitgehend vordefiniert. Der Algorithmus speicherte die Textinformationen in einfachen Textdokumenten, wobei jeder extrahierte Paragraph einem Absatz im Dokument entspricht. Der Zeilenumbruch zwischen den einzelnen Paragraphen wurde zwecks der Lesbarkeit gewählt. Dies erleichtert die Übersicht über die Qualität oder die Anzahl der gewonnen Informationen. Zudem wurde darauf geachtet, möglichst wenig Textinformationen zu verlieren.

Im Gegensatz zu gängigen Textextraktionspraktiken, bei denen des Öfteren aus Gründen der Effizienz Sonderzeichen, Zahlen, sogenannte Stopp-Wörter oder kurze Wörter entfernt werden, wurde in diesem Fall bewusst darauf verzichtet (Mine & Mine, 2021). Die Begründung liegt darin, dass wichtige Informationen verloren gehen könnten. Insbesondere besteht die Gefahr, dass relevanten Wörter für die anschließende Schlüsselwortsuche aussortiert werden. Ein konkretes Beispiel hierfür ist das Wort "CO2". Ebenso könnte die Abkürzung "ghg" (Treibhausgase), aufgrund ihrer geringen Buchstabenanzahl, bei üblichen Textextraktionsverfahren aussortiert werden, was in diesem Fall vermieden wurde.

Nachfolgend wird exemplarisch die Datenstruktur und das Format der extrahierten Daten dargestellt. Wie bereits erwähnt, entspricht jeder Paragraph einer Zeile im Textdokument, und sämtliche Informationen wurden ohne Ausnahmen inkludiert. Die Visualisierung verdeutlicht die Extrahierung von den Texten auf der Webseite in ein einfaches Textformat. In diesem Fall wurde beispielhaft ein kurzer Ausschnitt der Umweltseite der Flugsicherungsorganisation NATS aus dem United Kingdom zur Veranschaulichung gewählt. Dies kann in Abbildung 15 eingesehen werden.




```
Creating a sustainable future for aviation
While aviation delivers strong economic and social benefits to the UK and beyond, it also has a clear
impact on the environment; we see it as our responsibility to help drive, deliver and maintain a
sustainable future for the industry.
Reducing aviation's environmental impact and saving fuel is important to us and our customers. As part of
our environmental strategy, we are implementing smarter and more responsive ways to do this.
```

Abbildung 15: Datenstruktur der extrahierten Inhalte

Quelle: Eigene Darstellung

Die Textinformationen wurden wie in Abbildung 15 veranschaulicht, in einer sehr einfachen Struktur und Format von den Webseiten extrahiert. Im nächsten Schritt wird die Datenqualität der extrahierten Inhalte bewertet. Anschließend wird es möglich sein eine Einschätzung über den Aufwand, welcher für die Säuberung der Dokumente nötig sein wird, abzuleiten.

5.3.3. Datenqualität

In einer ersten Analyse der Datenqualität wird nun vor allem auf die Vollständigkeit, Genauigkeit sowie die Konsistenz der Datenbasis geachtet. Nachdem die Daten von den 32 verfügbaren Webseiten erfolgreich gesammelt wurden, geht es nun darum, einen ersten Eindruck von den Daten zu gewinnen. Es muss allerdings vorweg genommen werden, dass aufgrund unterschiedlicher Webseiten und damit unterschiedlicher Strukturen, hier unterschiedliche Ergebnisse der Textinformationen auftreten. Jede der 32 Webseiten und damit auch die Ergebnisse des Scrapings ist individuell und musste daher im Prozess auch individuell validiert werden.

Eine manuelle Investigation und Vergleich der Zielwebseiten sowie der Ergebnisse hat ergeben, dass die Vollständigkeit der Daten gegeben ist. Dies hat herausgestellt, dass alle gewünschten Informationen zur Gänze extrahiert wurden, und keine Lücken aufgetreten sind. Die Vollständigkeit der Informationen wurde vor allem durch den erfolgreichen und genauen Extraktionsprozess der Links gewährleistet. Dieser hat sichergestellt, alle relevanten Webseiten innerhalb einer Domäne zu identifizieren. Somit wurden auch alle Informationen vom anschließenden Scraping Code gefunden und erfolgreich extrahiert.

Ebenso wie mit der Vollständigkeit steht es um die Genauigkeit der Daten. Die Genauigkeit der gewonnen Daten wurde ebenfalls durch eine manuelle Investigation und Vergleich mit den Original-

webseiten sichergestellt. Hierbei wurden keine Abweichungen zwischen den Informationen auf den Webseiten und dem finalen Textdokument erkenntlich. Abweichungen könnten hier allerdings vor allem in Fällen von dynamisch generierten Inhalten auftreten. Allerdings war zum Zeitpunkt der Datenvalidierung die Genauigkeit der Daten gegeben.

Auch die Struktur der Zieldaten erwies sich als konsistent über alle verwendeten Webseiten hinweg. Dies ist vor allem einer genauen Analyse der Strukturen noch vor dem Scraping Prozess zurückzuführen. Daher war es sichergestellt, dass alle extrahierten Daten dieselbe, leicht verständliche und gut strukturierte Form aufwiesen. Dies sollte besonders in der fortführenden Analyse von Vorteil sein.

Zusammenfassend kann über die Datenqualität der gesammelten Informationen keine negativen oder überraschenden Aussagen getroffen werden. Vollständigkeit, Genauigkeit sowie die Klarheit und die Struktur sind in allen 32 Fällen durch effiziente Scraping-Scripts sowie einer genauen Vorbereitung noch vor dem eigentlichen Prozess gegeben. Natürlicherweise muss festgehalten werden, dass durch die manuelle Validierung auch der ein oder andere Fehler in den Resultaten entdeckt wurde. Hierbei handelte es sich aber zumeist um das extrahieren von irrelevanten Inhalten aus den Webseiten. Informationen wie beispielsweise Kontaktdaten werden hierunter verstanden. Derartige Informationen sind für eine folgende Analyse selbstredend irrelevant, wurden aber dennoch aufgrund gegebener Webseitenstrukturen extrahiert. Irrelevante Informationen geht es nun in einem nächsten Schritt zu identifizieren und so gut wie möglich aus der Datenbasis auszuschließen. Ein weiterer Faktor ist das Vorkommen identischer Paragraphen auf unterschiedlichen Webseiten. Das Scraping-Script hat diese nicht in allen Fällen als ident erkannt, und es wurden dadurch teilweise idente Paragraphen von verschiedenen Webseiten extrahiert. Die Identifikation sowie folgender Ausschluss derartiger Informationen, steigern die Qualität der Ergebnisse sowie die Effizienz von anschließenden Analysen.

5.4. Datenreinigung

Die Struktur und das Format der gesammelten Daten wurden zu diesem Punkt bereits mehrfach erwähnt. In diesem Schritt geht es nun darum, Unreinheiten in den Daten zu identifizieren und anschließend aus der Textbasis zu entfernen. Eine gewissenhafte Datensäuberung stellt sicher, dass in einer nachfolgenden Analyse keine Zeit und Aufwand dafür verschwendet wird, irrelevante Daten zu untersuchen. Bevor dies aber in die Tat umgesetzt werden kann, müssen die Anomalien und Unreinheiten in der extrahierten Textbasis identifiziert werden.

5.4.1. Identifikation von Unreinheiten

Bereits bei der ersten Übersicht der Daten, um eine erste Aussage über die Datenqualität treffen zu können, wurde auf einige Anomalien in den Textdaten hingewiesen. Die Identifikation der Anomalien in den Textdaten basierte hauptsächlich auf einer manuellen Investigation der extrahierten Inhalte. Diese Identifikation konzentrierte sich besonders auf das Vorhandensein von zwei Arten von Unreinheiten in der Textbasis, da diese vermehrt vorgekommen sind. Einerseits handelt es sich hierbei um das Vorhandensein sehr ähnlicher Paragraphen oder Duplikaten. Allerdings wurde wie bereits bekannt das Abspeichern von Duplikaten im Scraping-Script weitestgehend berücksichtigt. Nichtsdestotrotz wurden einige Paragraphen von den Webseiten extrahiert, welche aus verschiedenen Gründen beinahe ident sind. Ein Beispiel hierfür, welches häufig Schuld daran hatte, warum ähnliche

Textpassagen extrahiert wurden, ist in Abbildung 16 veranschaulicht. In den meisten Fällen handelte es sich hierbei sozusagen um eine Vorschau auf eine verlinkte Webseite und bereits vorgezeigte Textinhalte. Dies bedeutet wiederum, dass ähnliche aber nicht idente Textpassagen auf zwei unterschiedlichen Links aufscheinen und damit vom Scraping-Script extrahiert werden. Da dies recht häufig der Fall war, würde dies für eine anschließende Analyse einen erheblichen Mehraufwand darstellen, wobei keinerlei neue Informationen zum Vorschein kommen. Aus diesem Grund musste ein Weg gefunden werden, wie diese sehr ähnlichen Textpassagen nur einmal in der Datenbasis abgespeichert werden.

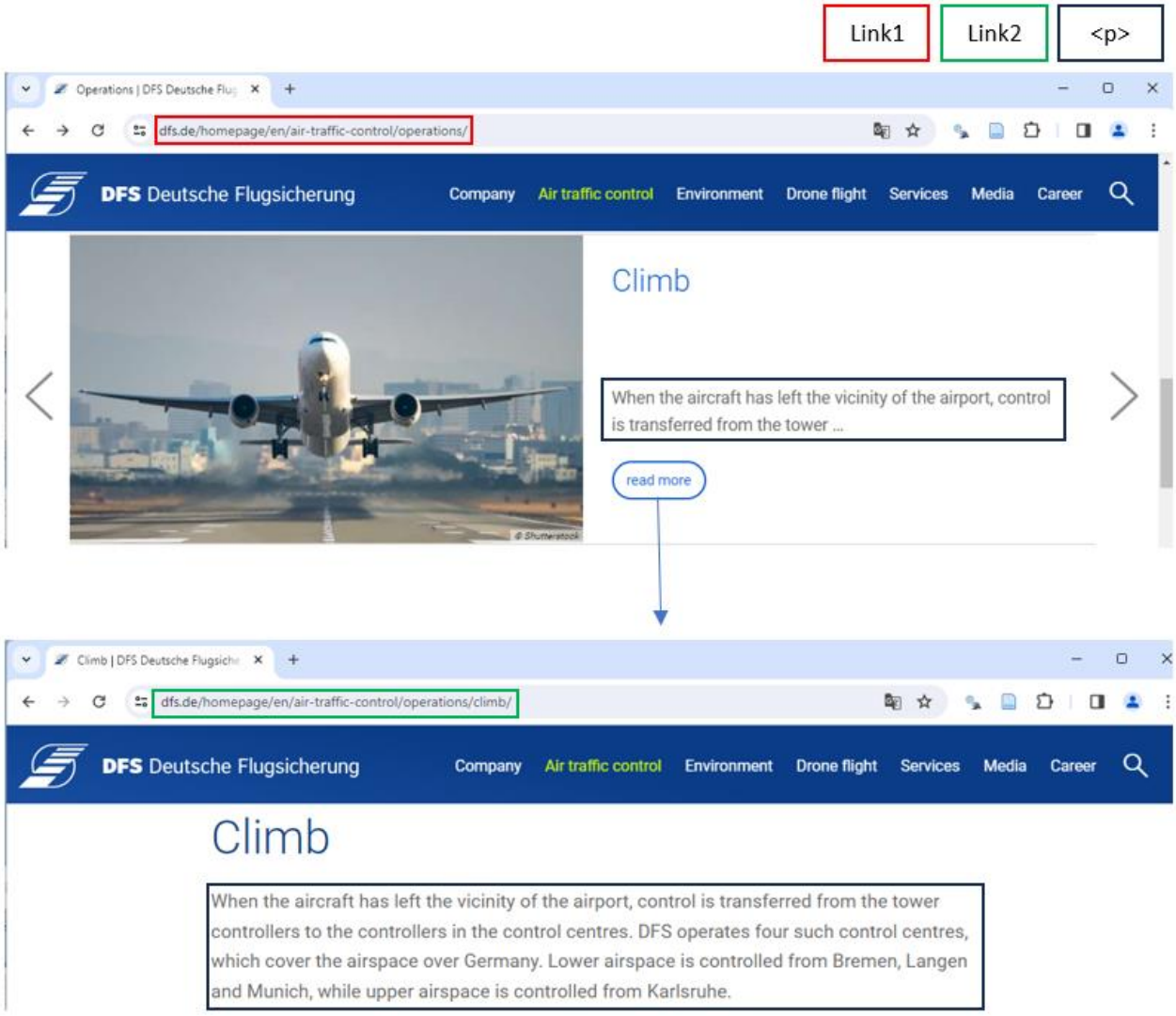


Abbildung 16: Ähnliche Paragraphen auf unterschiedlichen Webseiten
Quelle: https://www.dfs.de/homepage/en/air-traffic-control/operations/

In Abbildung 16 handelt es sich um Ausschnitte von der Webseite der deutschen Flugsicherungsorganisation namens DFS. Hier wird grafisch die Ausgangssituation dargelegt und veranschaulicht, wie es im Scraping Prozess dazu kommen kann, dass ähnliche Paragraphen mehrmals in der Datenbasis gespeichert werden. Wird auf der ersten Webseiten (Link1) auf den Button „read more“ geklickt, so wird der gesamte Beitrag unter einem neuen Link (Link2) angezeigt. Es wird also auf zwei unterschiedlichen Webseiten beziehungsweise Links beinahe dieselbe Information in einem gleichen

HTML-Tag gespeichert, was zur Extrahierung beider Inhalte führt. Das Ergebnis in der finalen Textdatei sieht wie in Abbildung 17 dargestellt aus.

```
When the aircraft has left the vicinity of the airport, control is transferred from the tower ...
When the aircraft has left the vicinity of the airport, control is transferred from the tower controllers
to the controllers in the control centres. DFS operates four such control centres, which cover the
airspace over Germany. Lower airspace is controlled from Bremen, Langen and Munich, while upper airspace
is controlled from Karlsruhe.
```

Abbildung 17: Idente Paragraphen in den Textdaten

Quelle: Eigene Darstellung

In Abbildung 17 kann nun gut eingesehen werden, dass es sich um ähnliche, allerdings nicht idente Paragraphen handelt, weshalb das Scraping-Script diese auch nicht als ident feststellt, und damit mehrmals abspeichert. Selbstredend macht es aber dennoch keinen Sinn derartige Paragraphen, wie den ersten und unvollständigen in Abbildung 17, in der Datenbasis zu behalten. Der Aufwand einer folgenden Analyse steigt damit an, wobei keinerlei neue Erkenntnisse aus diesem Paragraphen gewonnen werden können. Dies bedeutet, dass für derartige Fälle eine Lösung zum Ausschluss der entsprechenden Paragraphen gefunden werden muss. Wie in Abbildung 17 erkennbar, handelt es sich hierbei zumeist um unvollständige Paragraphen, welche zu Beginn ident mit dem vollständigen Paragraphen sind. Häufig werden derartige unvollständigen Paragraphen mit Phrasen wie beispielsweise „…“, „weiterlesen“, oder „mehr Information“ beendet.

Die zweite Anomalie, welche bei der manuellen Validierung der Daten gefunden wurde, bezieht sich auf das Vorhandsein irrelevanter Informationen in der extrahierten Textbasis. Hierbei handelte es sich zumeist um Kontaktinformationen wie beispielsweise E-Mailadressen, Telefonnummern oder ähnliches. Darüber hinaus auch Seitenüberschriften, welche im entsprechenden HTML-Element auf der Webseite gespeichert wurden. Ein weiteres Beispiel für derartige Informationen, wie sie häufig aufgetreten sind, sind Namen oder ähnliches. Ein zusammengetragener Ausschnitt von einzelnen Beispielen derartiger Informationen ist in Abbildung 18 veranschaulicht. Das Beispiel beruht auf den Scraping Ergebnissen der Webseite https://www.crocontrol.hr/, dem ANSP des Landes Kroatien.

```
Phone: +385 1 6259 589
Fax: +385 1 2020 338
About Us
Activities
Boris Mrkonja
Copyright
Media
Disclaimer
```

Abbildung 18: Irrelevante Informationen in den Textdaten

Quelle: Eigene Darstellung

Die Identifikation derartiger Informationen beruht vorrangig auf deren Format. Darüber hinaus kommt hierbei wiederum die Struktur der Daten zu Vorteil, da alle extrahierten Inhalte eine Zeile im Textdokument bilden. Dies hilft bei der Identifikation irrelevanter Daten, da sie auch bei manueller Analyse, auf den ersten Blick aufgrund ihrer Länge oder des Formats auffallen. Wie in Abbildung 18 abgebil-

det, handelt es sich hierbei um Paragraphen, bestehend aus einem, oder sehr wenigen Wörtern, oder aus einer Folge an Zeichen und Zahlen. Selbstredend kann aus Paragraphen wie diesen ohnehin keine relevanten Erkenntnisse entnommen werden. Darüber hinaus erschweren sie eine Analyse, da zahlreiche irrelevante Informationen entweder manuell oder durch ein Algorithmus analysiert werden müssen. Aufgrund dessen werden derartige Textabschnitte für eine anschließende Analyse aus der Datenbasis entfernt.

5.4.2. Methoden zur Datensäuberung

Nach der Identifikation der Unreinheiten in den Daten geht es nun um die Bereinigung dieser. Bei der Bereinigung beider identifizierter Anomalien, welche im vorigen Kapitel im Detail geschildert wurden, war die Struktur der beschafften Daten von großem Vorteil. Die Speicherung in einzelnen Paragraphen getrennt voneinander durch eine Leerzeile im Textdokument, ließ einen einfachen Vergleich der Daten zu. Dieser Punkt war vor allem für die Validierung nach der Bereinigung bedeutend, auf welche im nächsten Kapitel eingegangen wird.

Um die ähnlichen aber nicht identen Paragraphen in der Datenbasis zu identifizieren und diese zu entfernen, wurde ein einfacher Vergleich aller Textelemente miteinander vorgenommen. Die Abspeicherung der Elemente in einzelne Paragraphen machte es einfach möglich, alle Paragraphen alphabetisch zu sortieren. Das Ergebnis ist nun, dass ähnliche Paragraphen untereinander stehen und somit auch für eine manuelle Validierung sofort ins Auge fallen. Natürlich wurde sich mit der Sortierung sowie der Identifikation ähnlicher Paragraphen wiederum mit der Programmiersprache R geholfen. Das Script weißt dabei folgende Funktionalität auf. Sollte es an irgendeiner Stelle vorkommen, dass sich idente Paragraphen in der Datenbasis befinden, so werden diese auf jeden Fall vom Code identifiziert. Um ähnliche aber nicht idente Paragraphen zu finden, werden die ersten fünf Wörter der Paragraphen miteinander verglichen. Sind die ersten fünf Wörter ident, so wird jener Paragraph identifiziert, welcher eine kürzere Gesamtlänge aufweist. So können auf einfache Weise die Fälle mit „…“ oder „weiterlesen“ entdeckt werden. Die gesamte Funktionalität des Algorithmus wird folgend genauer erklärt, sowie der Code veranschaulicht.

Der Algorithmus liest eine Sammlung von Paragraphen aus einer gegebenen Textdatei ein. Hierbei handelt es sich um die Textdateien der 32 europäischen ANSPs. Anschließend werden die Paragraphen alphabetisch sortiert, dies stellt sicher, dass sich ähnliche beziehungsweise idente Paragraphen untereinander befinden. Die Identität wird durch einen Vergleich der ersten fünf Wörter sichergestellt. Wenn die Paragraphen komplett identisch sind, unabhängig von ihrer Länge, werden sie erkannt. Mithilfe dieser Funktionalität wird sichergestellt, dass auch tatsächlich idente Informationen aus der Datenbasis entfernt werden. Zur Klassifikation als ähnliche Paragraphen müssen die ersten fünf Wörter übereinstimmen. Danach wird der kürzere der beiden Paragraphen identifiziert und in einer Variable gespeichert. Dies passiert mit allen ähnlichen Paragraphen des jeweiligen Textdokuments. Die Speicherung der ähnlichen Paragraphen in eine gemeinsam Variable, bietet die Möglichkeit zur manuellen Überprüfung vor der endgültigen Löschung aus der Datenbasis. Nach einer manuellen Kontrolle der identifizierten Paragraphen werden diese aus den Daten entfernt.

Listing 5: Identifikation identer und ähnlicher Paragraphen

```
# Einlesen der Paragraphen aus dem Textdokument
setwd(Pfad zum Arbeitsverzeichnis mit den Textdokumenten)
paragraphs <- readLines("Beispiel-Textdatei")

# Funktion zur Überprüfung der Identität der ersten 5 Wörter
check_first_words_identity <- function(paragraph1, paragraph2) {
  words1 <- strsplit(paragraph1, "\\s")[[1]][1:5]
  words2 <- strsplit(paragraph2, "\\s")[[1]][1:5]

  # Identität überprüfen
  identical(words1, words2)
}

# Identifizieren der identen Paragraphen
identified_paragraph_indices <- numeric(0)

# Schleife durch alle Paragraphen im Dokument, beginnend ab dem zweiten Paragra-
phen
for (i in 2:length(paragraphs)) {
  # Überprüfen, ob die Identität der ersten 5 Wörter besteht
  if (check_first_words_identity(paragraphs[i-1], paragraphs[i])) {
    # Speichern des Index des vorherigen Paragraphen
    identified_paragraph_indices <- c(identified_paragraph_indices, i-1)
  }
}

# Ausgabe der identen bzw. des kürzeren Paragraphen
identified_paragraphs <- paragraphs[identified_paragraph_indices]
print(identified_paragraphs)

# Löschen der identifizierten Paragraphen
filtered_paragraphs <- paragraphs[-identified_paragraph_indices]
filtered_paragraphs
```

Dieser Algorithmus ermöglicht eine effiziente Identifikation und Entfernung von Duplikaten oder ähnlichen Inhalten in einer Sammlung von Paragraphen. In der Praxis hat sich ein Vergleich der ersten fünf Wörter als am vielversprechendsten herausgestellt. Abbildung 19 zeigt ein visuelles Beispiel, welches die Funktionalität des Algorithmus veranschaulicht. Im ersten Schritt wird eine Beispieldatei an identen und ähnlichen Paragraphen gezeigt, zusammengetragen aus der realen Datenbasis. Anschließend wird veranschaulicht, welche Paragraphen der Algorithmus zur potentiellen Entfernung aus der Datenbasis identifiziert hat.

```
Environmental and climate protection concerns everyone.
Environmental and climate protection is cruicial for all of us.
When the aircraft has left the vicinity of the airport, control is transferred from the tower ...
When the aircraft has left the vicinity of the airport, control is transferred from the tower controllers
to the controllers in the control centres. DFS operates four such control centres, which cover the airspace
over Germany. Lower airspace is controlled from Bremen, Langen and Munich, while upper airspace is
controlled from Karlsruhe.
Phone: +385 1 6259 589
Phone: +385 1 6259 589
Phone: +385 1 6259 589
```

Abbildung 19: Anwendung Identifikation identer und ähnlicher Paragraphen

Quelle: Eigene Darstellung

Anhand der soeben beschriebenen Funktionalität des Algorithmus kann in Abbildung 19 festgestellt werden, dass der dritte Paragraph sowie zwei der drei Telefonnummern zur Löschung aus den Daten identifiziert werden sollten. Die ersten beiden Paragraphen in Abbildung 19 sind anhand der definierten Kriterien nicht als ähnliche Paragraphen zu klassifizieren und werden daher auch nicht vom Algorithmus als derartige wahrgenommen. Die folgenden zwei Paragraphen in Abbildung 19 stellen das zuvor vorgestellte Beispiel der ähnlichen aber nicht identen Paragraphen dar. In diesem Fall sollte der erste der beiden Paragraphen, aufgrund seiner geringeren Wortanzahl im Vergleich zum vollständigen Absatz, identifiziert werden. Die folgenden drei Telefonnummern sind ident und mehrfach vorhanden. Hier ist es relevant, dass zwei davon identifiziert werden um Duplikate zu entfernen. In Abbildung 20 sind die korrekt identifizierten Paragraphen des Algorithmus zu sehen.

```
> print(identified_paragraphs)
[1] "When the aircraft has left the vicinity of the airport, control is transferred from the tower ..."
[2] "Phone: +385 1 6259 589"
[3] "Phone: +385 1 6259 589"
```

Abbildung 20: Ergebnis Identifikation identer und ähnlicher Paragraphen

Quelle: Eigene Darstellung

Nach einer anschließenden manuellen Validierung der identifizierten Abschnitte, werden diese aus der Datenbasis entfernt und somit das Nicht-Vorhandensein von Duplikaten sowie irrelevanten oder redundanten Informationen sichergestellt.

Die erste Art von Anomalien in der Textdatenbank wurde nun identifiziert und konnte erfolgreich aus der Datenbank entfernt werden. Vorhandene Duplikate beziehungsweise sehr ähnliche Informationen wurden aus der Datenbasis entfernt. Bei der zweiten Art handelt es sich nun zumeist um irrelevante Informationen wie vorher angesprochen. Telefonnummer, andere Kotaktdaten, Überschriften oder generell sehr kurze Paragraphen, in welchen kaum Information vorhanden ist, stellen diese zweite Herausforderung dar. Auch bei der Identifikation dieser Informationen ist ein Algorithmus, geschrieben in R, hilfreich. Dieser identifiziert die Abschnitte in der Datenbasis aufgrund ihrer Länge.

Die Funktionalität dieses Scripts ist vergleichbar mit jenem zuvor. Zu Beginn werden die Paragraphen aus einer Textdatei eingelesen. Die folgende Funktion „identify_short_paragraphs“ durchläuft anschließend jeden Paragraphen in der übergebenen Textdatei und prüft, ob der Paragraph gesamt aus fünf oder weniger Wörtern besteht. Durch eine manuelle Analyse einiger der Textdateien wurde offensichtlich, dass fünf Wörter hierfür einen guten Parameter darstellt. Paragraphen bestehend aus fünf oder weniger Wörtern, beinhalten kaum relevante Information. Besteht ein Absatz aus lediglich

bis zu fünf Wörtern, wird der Paragraph zu einer Liste hinzugefügt. Schlussendlich gibt die Funktion die identifizierte Liste kurzer Paragraphen zurück. Bevor diese aus der Originaldatei entfernt werden, werden sie nochmals durch manuelle Analyse auf Korrektheit überprüft. Ist dies gegeben, werden die identifizierten Paragraphen aus der Originaldatei gelöscht.

Listing 6: Identifikation kurzer Paragraphen

```
# Einlesen der Paragraphen aus dem Textfile
setwd(Pfad zum Arbeitsverzeichnis mit den Textdokumenten)
paragraphs <- readLines("Beispiel-Textdatei")

# Funktion zur Identifikation kurzer Paragraphen
identify_short_paragraphs <- function(paragraphs) {
  # Initialisierung der Variable für kurze Paragraphen
  short_paragraphs_indices <- numeric(0)

  # Schleife durch alle Paragraphen
  for (i in seq_along(paragraphs)) {
    word_count <- length(strsplit(paragraphs[i], "\\s")[[1]])

    # Überprüfen auf fünf oder weniger Wörter und hinzufügen zur Liste
    if (word_count <= 5) {
      short_paragraphs_indices <- c(short_paragraphs_indices, i)
    }
  }

  return(short_paragraphs_indices)
}

# Funktionsaufruf
short_paragraph_indices <- identify_short_paragraphs(paragraphs)

# Ausgabe der identifizierten Paragraphen zur Kontrolle
print(paragraphs[short_paragraph_indices])

# Löschen der identifizierten Paragraphen in eine neue Datei
filtered_paragraphs <- paragraphs[-short_paragraph_indices]
```

Dieser einfache Code ermöglicht es im Grunde, kurze Paragraphen aufgrund deren Wortanzahl in einer gegebenen Sammlung zu identifizieren. Anschließend werden diese validiert und gelöscht. Um die Funktionalität der Funktion wieder besser zu illustrieren, ist in Abbildung 21 ein Beispielauszug aus den Textdokumenten gezeigt. Auf den ersten Blick ist ersichtlich, dass in diesem Auszug Informationen enthalten sind, welche für die Zielsetzung der vorliegenden Masterarbeit irrelevant sind. Nichtsdestotrotz befinden sie sich derzeit noch in der Datenbasis, da sie von der vorherigen Funktion nicht aufgespürt wurden.

```
When the aircraft has left the vicinity of the airport, control is transferred from the tower controllers
to the controllers in the control centres. DFS operates four such control centres, which cover the airspace
over Germany. Lower airspace is controlled from Bremen, Langen and Munich, while upper airspace is
controlled from Karlsruhe.
Media
Phone: +385 1 6259 589
```

Abbildung 21: Anwendung Identifikation kurze Paragraphen

Quelle: Eigene Darstellung

Die soeben vorgestellte Funktion „identify_short_paragraphs“ identifiziert richtigerweise folgende Paragraphen, zur folgenden manuellen Validierung sowie Löschung aus der Datenbasis.

```
> print(paragraphs[short_paragraph_indices])
[1] "Media"                    "Phone: +385 1 6259 589"
```

Abbildung 22: Ergebnis Identifikation kurzer Paragraphen

Quelle: R Studio, Eigene Darstellung

Mithilfe der eben vorgestellten Funktionen wurde sichergestellt, so viel irrelevante Informationen wie möglich aus den Textdateien zu entfernen. Dies führt wiederum zu besseren Analyseergebnissen und vermindertem Zeitaufwand für die Durchführung einer Analyse. In einem letzten Schritt im Rahmen der Datensäuberung wurde erneut eine umfassende Kontrolle der gesäuberten Daten vorgenommen. Dies sollte nochmals einen Eindruck schaffen, wie erfolgreich die Funktionen waren und wie es um die Datenqualität nach der Säuberung steht.

5.4.3. Validierung der gesäuberten Daten

Die Validierung der nun gereinigten Daten wurde mittels eines Vergleiches der Originaldaten angestellt. Somit ist sichergestellt, dass keine relevanten Informationen während des Prozesses verloren gingen. Sichergestellt wurde dies, indem die zu entfernenden Informationen und Paragraphen nicht automatisch aus der Datenbasis entfernt wurden, sondern zuerst in eine eigene Liste gespeichert wurden. Diese Liste wurde anschließend einer manuellen Kontrolle unterzogen. Mit diesem Schritt im Prozess, wurde sichergestellt, dass keine relevante Information unabsichtlich aus der Datenbasis entfernt wird.

Ein zweiter Schritt, die Datenqualität zu erhalten ist der eben angesprochene Vergleich der Daten nach dem Reinigungsprozess zu den Originaldaten. Diese Dokumente werden erneut manuell verglichen und stichprobenartig geprüft, ob die relevanten Informationen auch in den gesäuberten Dateien noch vorhanden sind. Auch die Identifikation relevanter Paragraphen für den Vergleich wurden mittels einer manuellen Überprüfung identifiziert. Hierbei wurden keine Anomalien oder Fehler in den gereinigten Daten festgestellt. Die Datenqualität ist daher nach dem Prozess als höher einzustufen, da die zuvor erklärten Funktionen nur irrelevante Informationen aus den Textdokumenten herausgefiltert haben. Die Textdokumente bestehen nun aus den von den Webseiten gesammelten Paragraphen, ohne Duplikate oder ähnlichen irrelevanten Inhalten. Auch der Informationsgehalt der Dokumente wurde mittels der Löschung sehr kurzer Textabschnitte erhöht.

Nach Abschluss der Datensäuberung sowie der Validierung der Daten nach dem Prozess, kann die Datenbasis nun folgend dafür verwendet werden, die eigentlichen Analysen durchzuführen. Durch die gewissenhafte Säuberung der Textdokumente sind die Daten nun bestens auf eine Analyse vor-

bereitet. In einem ersten Schritt wird wie schon im Rahmen dieser Masterarbeit angesprochen, eine Schlüsselwortsuche oder sogenannte Keyword-Analyse durchgeführt.

6. Textanalyse von ANSP-Webinhalten mittels Text Mining

Im folgenden Abschnitt wird die durchgeführte Text-Mining-Analyse sowie auch deren Ergebnisse beschrieben. Wie zuvor angesprochen, handelt es sich bei der ersten anzuwendenden Methode aus dem Text Mining um eine Keyword-Analyse. Die Vorbereitung für eine derartige Analyse, der Prozess an sich sowie die Ergebnisse werden folgend beschrieben.

6.1. Keyword-Analyse

Wie schon in den Grundlagen erwähnt, liegt das Ziel einer Keyword-Analyse oder Schlüsselwortsuche laut Talib et al. (2016) darin, spezifische Informationen aus einem Textkorpus zu extrahieren. Es geht im Grunde um die Prüfung, ob bestimmte Informationen in einer Datenbasis vorhanden sind (Talib et al., 2016). Einige Herausforderungen der Schlüsselwortsuche wurden bereits von Noh et al. (2015) in den Grundlagen beschrieben. Eine sehr bedeutende Herausforderung stellt dabei die Auswahl der Keywords beziehungsweise Schlüsselwörter dar. Diese Terme und Begriffe entscheiden schlussendlich über Erfolg oder Misserfolg einer Keyword-Analyse. Die Identifikation und die Auswahl der anzuwendenden Wörter stellt den ersten Schritt einer Keyword-Analyse dar.

6.1.1. Auswahl der Keywords

Die Definition und Auswahl der Keywords erfolgte in Zusammenarbeit mit einem Experten auf diesem Gebiet. Dadurch sollte sichergestellt werden, aktuelle und in der Domäne verwendete Begriffe zu erhalten. Alternativ hätte man diese auch über eine Analyse der Datenbasis in der entsprechenden Thematik herausfiltern können. Da jedoch alle ANSPs unterschiedlich sind und auch in ihrer sprachlichen Ausdrucksweise variieren, wurde die Expertise eines Domänenexperten hinzugezogen. Hierbei stand Herr Emir Ganić, PhD, von der Fakultät für Transport und Verkehrstechnik an der Universität Belgrad in Serbien zur Verfügung. In dieser Absprache ging es vor allem darum, herauszufinden, welche aktuelle Methoden, Praktiken und Systeme die europäischen ANSPs verwenden, um genau nach diesen Schlüsselwörtern in der Datenbasis zu suchen. Wie im Gespräch mit Herrn Ganić erfahren, erschwert die Individualität der europäischen ANSPs die Auswahl passender Keywords. Zahlreiche Methoden und Praktiken, insbesondere Systeme beispielsweise zur automatisierten Luftverkehrsplanung oder zur unterstützen Entscheidungsfindung, werden von den ANSPs unter unterschiedlichen Namen verwendet. Somit ist davon auszugehen, dass auch in den Paragraphen der Datenbasis derartige Systeme unter unterschiedlichsten Namen erwähnt und erklärt werden. Aus diesem Grund wurden im Rahmen des Gesprächs jene Schlüsselwörter und Begriffe erarbeitet, welche die ANSPs mit einer hohen Wahrscheinlichkeit verwenden, wenn sie über derartige Systeme oder andere verwendete Praktiken und Techniken sprechen. Somit wurden nicht spezifische Systeme und Techniken als Begriffe verwendet, sondern allgemeinere Begriffe welche im Rahmen der Beschreibung dieser sehr wahrscheinlich auftreten. Dieser Ansatz hat zwei bedeutende Vorteile im Gegensatz zu der Verwendung vordefinierter Begriffe. Mithilfe dieser mehr allgemeinen Begriffen werden auf mehrere Stellen im Text hingewiesen, welche relevante Informationen beinhalten. Somit ist auch sichergestellt, dass im Endeffekt Systeme und Praktiken von unterschiedlichen ANSPs identifiziert werden, obwohl diese mit unterschiedlichen Namen referiert werden. Des Weiteren wird durch die Verallgemeinerung der Begriffe so ein gesamtheitliches Stimmungsbild der Datenbasis gewonnen und es kann besser abgeschätzt werden, inwiefern die Umweltthematik generell in der Daten-

bank repräsentiert ist. Dies ist ein erster und wichtiger Hinweis, ob relevante beziehungsweise gesuchte Informationen überhaupt in den Daten vorhanden sind.

Nach dieser Festlegung auf die Kriterien der Keywords, wurden folgende Keywords erarbeitet. Die Begriffe sind insgesamt in sieben Kategorien aufgeteilt und können der untenstehenden Tabelle entnommen werden. Die sieben verwendeten Kategorien sind in Tabelle 4 einerseits in deren englischen Bezeichnung, als auch in deren deutschen Übersetzung zu sehen. Die Keywords wurden in englischere Sprache erarbeitet und auch in dieser für folgende Analysen verwendet, weshalb hier von einer Übersetzung abgesehen wurde. Das teilweise vorkommende Zeichen „*“ am Ende von Begriffen steht für einen Platzhalter. Im Falle des Begriffes „annoy*“ aus der Kategorie Noise könnte dies also beispielsweise für annoying, annoyed oder jegliche andere Endung des definierten Wortstammes „annoy“ stehen.

Tabelle 4: Keywords

Noise/Lärm
sleep disturb*, sleep-disturb*, annoy*, sound*, noise monitor*, noise measur*, balanced approach, noise, noise management, noise abatement, noise level, noise certificate, noise charges, noise footprint, noise map, noise issue, noise complaint, noise action plan, noise modelling, noise mapping, noise exposure, noise perception, noise prediction, noise problem, noise regulation, noise tolerance, acoustic, sound insulation, noise insulation
Air Quality/Luftqualität
air quality, emission*, GHG, Greenhouse, air pollution, CO2, carbon dioxide
Fuel Consumption/Kraftstoffverbrauch
fuel consumption, reduc* fuel burn, fuel consumption redu*
Sustainability/Nachhaltigkeit
sustainability, sustainable future, sustainable aviation
Environment/Umwelt
environment, environmental management, environmental impact, environmental
Operational Efficiency/Operationale Effizienz
Free route airspace, fra, continuous descent, continuous climb, cdo, cda, cco
Reports/Berichte
environ* report, sustain* report

Die sieben unterschiedlichen Kategorien der Keywords wurden gewählt, um auch einen Eindruck zu gewinnen, in welchem Rahmen beziehungsweise welche Ziele mit den angewendeten Techniken oder Systemen der ANSPs verfolgt werden. Die Kategorie „Reports“ wurde gewählt, um auf mögliche weitere Datenquellen wie beispielsweise Umweltberichte oder Stellungnahmen, welche auf den Webseiten genannt werden, zu identifizieren. Diese sind häufig nicht textuell auf den Webseiten vorhanden, sondern zumeist downloadbar in PDF Format. Deshalb dient diese Kategorie der Identifikation derartiger Informationen, da auch diese wichtige Informationen beinhalten können.

Nach der Identifikation der Schlüsselbegriffe durch die Hinzuziehung der Expertenmeinung von Herrn Ganić, geht es im nächsten Schritt um die eigentliche Keyword-Analyse innerhalb der gesammelten Daten. Im nächsten Abschnitt wird die Methodik der durchgeführten Schlüsselwortsuche erläutert.

6.1.2. Methodik

Der Prozess der Keyword-Analyse beinhaltet folgende Schritte. Zur effizienten Durchführung wurde wiederum ein R-Script verfasst. Dieses hatte es zum Ziel den Prozess der Schlüsselwortsuche zu automatisieren. Das Ziel des Scripts ist es, die Daten für die Keyword-Analyse entsprechend vorzubereiten und diese dann anschließend nach den zuvor identifizierten Begriffen zu durchsuchen. Dabei wird jedes der 32 Textdokumente einzeln durchsucht und die Ergebnisse anschließend festgehalten. Somit können schlussendlich die Resultate aller 32 europäischen ANSPs miteinander verglichen werden. Die Hauptfunktionalität des Scripts liegt also darin, die Textdokumente einmal zu durchlaufen und das Vorkommen der Schlüsselwörter zuverlässig zu zählen.

Folgend wird die gewünschte Funktionalität des R-Scripts beschrieben sowie anschließend der Code erklärt. Der verwendete Code befindet sich unterhalb der Erklärung in Listing 7 und Listing 8 abgebildet.

Der verwendete Code für die Keyword-Analyse besteht aus zwei Funktionen. Die erste Funktion „preprocess_text“ dient dazu, die Textdokumente entsprechend vorzubereiten und in eine Form zu bringen, um die Schlüsselwörter verlässlich im Text aufzuspüren. Die Funktion nimmt einen Text als Übergabeparameter entgegen und führt folgende Schritte daran aus. Zuerst wird der gesamte Text in Kleinbuchstaben umgewandelt. Dadurch wird sichergestellt, dass keine Schlüsselwörter aufgrund von Klein- oder Großbuchstaben übersehen werden. In einem zweiten Schritt werden die Satzzeichen aus dem Textdokument entfernt. Dies ist besonders relevant, wenn Schlüsselwörter ohne Platzhalter am Satzende stehen. Dadurch würde der gesuchte Begriff aufgrund des direkt anschließenden Satzzeichens nicht identifiziert. Im nächsten Schritt werden Bindestriche aus dem Textdokument entfernt. Besonders auffällig war dies bei beim Schlüsselwort „CO-2“. Wurde dieses Wort auf diese Weise geschrieben, so würde es nicht als Keyword „co2“ entdeckt. Deshalb wurden Bindestriche aus dem Text entfernt. Im letzten Schritt der Funktion werden überflüssige Leerzeichen entfernt. Befindet sich am Satzende ein Leerzeichen oder zwischen zwei Wörtern mehr als ein Leerzeichen, werden diese entfernt. Der nun vorbereitete Text wird anschließend von der Funktion zurückgegeben und an die zweite Funktion übergeben.

Listing 7: Funktion Datenvorbereitung Keyword-Analyse

```
preprocess_text <- function(text) {
  text <- tolower(text)  #Umwandlung in Kleinbuchstaben
  text <- gsub("[[:punct:]]", "", text)  #Entfernung Satzzeichen
  text <- gsub("-", "", text)  #Bindestriche entfernen
  text <- gsub("\\s+", " ", text)  #Überflüssige Leerzeichen entfernen
  return(text)
}
```

Die zweite Funktion „count_keywords_with_wildcards” dient dazu, die Anzahl der Vorkommen von Schlüsselwörtern in einem Text zu zählen, wobei es einerseits möglich ist nach vollständigen

Schlüsselbegriffen zu suchen, als auch nach Schlüsselbegriffen mit Wildcards oder Platzhaltern. Die Funktion nimmt zwei Übergabeparameter entgegen. Bei diesen Parametern handelt es sich einerseits um den zu durchsuchenden Text sowie eine Liste an Schlüsselwörtern. Zu Beginn der Funktion wird der Eingabetext mittels der zuvor definierten Funktion in die gewünschte Form gebracht. Danach durchläuft die Funktion eine Schleife für jedes Schlüsselwort in der übergebenen Liste von Schlüsselwörtern. Anschließend wird überprüft, ob das aktuelle Schlüsselwort eine Wildcard enthält oder nicht. Wenn ja, wird das bei der Suche dementsprechend berücksichtigt. Andernfalls wird das Schlüsselwort als vollständiger Ausdruck behandelt. Die Funktion verwendet anschließend „regmatches“ und „gregexpr“, um alle Übereinstimmungen des Schlüsselbegriffs im vorbereiteten Text zu identifizieren. Die Anzahl an Übereinstimmungen für jedes Schlüsselwort wird gezählt und in dem Vektor „keyword_counts“ gespeichert. Die Funktion erstellt schlussendlich noch ein eigenes Datenframe um die Ergebnisse darin abzuspeichern. Dieses Datenframe besteht aus zwei Spalten. Die verwendeten Schlüsselwörter sowie deren jeweiligen Häufigkeiten sind darin enthalten. Dieses erstellte Datenframe wird im letzten Schritt von der Funktion zurückgegeben.

Listing 8: Funktion Keyword-Analyse

```
#Funktion zur Zählung der Schlüsselwörter
count_keywords_with_wildcard <- function(text, keywords) {
  #Texte vorbereiten mit der definierten Funktion
  text <- preprocess_text(text)
  keyword_counts <- numeric(length(keywords))

  for (i in seq_along(keywords)) {
    keyword <- keywords[i]

    #Überprüfen ob es sich um ein Wildcard-Keyword handelt oder nicht
    if ("*" %in% strsplit(keyword, "")[[1]]) {
      keyword_pattern <- gsub("\\*", ".*", keyword)
    } else {
      keyword_pattern <- paste0("\\b", keyword, "\\b")
    }

    #Überprüfen ob das reguläre Ausdrucksmuster gültig ist
    if (length(grep(keyword_pattern, "")) > 0) {
      stop("Invalid regular expression pattern: ", keyword_pattern)
    }

    #Verwendung von regmatches und gregexpr um übereinstimmende Teile zu extrahieren
    keyword_matches <- regmatches(text, gregexpr(keyword_pattern, text, ignore.case = TRUE))

    #Anzahl der Übereinstimmungen zählen
```

```
    keyword_counts[i] <- sum(sapply(keyword_matches, length))
  }

  #Ergebnis in ein eigenes Datenframe speichern und zurückgeben
  result <- data.frame(Keyword = keywords, Count = keyword_counts)
  return(result)
}
```

Zusammengefasst ermöglicht die Funktion „count_keywords_with_wildcards“ die Zählung von Schlüsselwörtern in einem Text, wobei auch Wildcards unterstützt werden, um verschiedene Formen der Schlüsselwörter zu berücksichtigen. Alle Vorbereitungen sind nun getroffen, um die Keyword-Analyse auf die Textdokumente der 32 europäischen ANSPs anzuwenden. Die Ergebnisse werden im nächsten Abschnitt präsentiert.

6.1.3. Ergebnisse

Die folgend präsentierten Ergebnisse der Keyword-Analyse werden einen Eindruck geben, inwiefern das Thema Umweltschutz in den extrahierten Texten der 32 ANSP-Webseiten vorhanden ist. Darüber hinaus gibt es einen Einblick, inwiefern eine weitere Analyse nach spezifischen Techniken oder Systemen sinnvoll ist. Mithilfe von Abbildung 23 kann ein erster Eindruck über die Vorkommnisse der Keywords in den unterschiedlichen Kategorien gewonnen werden. Klar zu sehen ist, dass die Kategorie Umwelt mit über 1000 absoluten Häufigkeiten den Spitzenreiter der Kategorien darstellt. Gefolgt von den Kategorien Luftqualität und Operationale Effizienz.

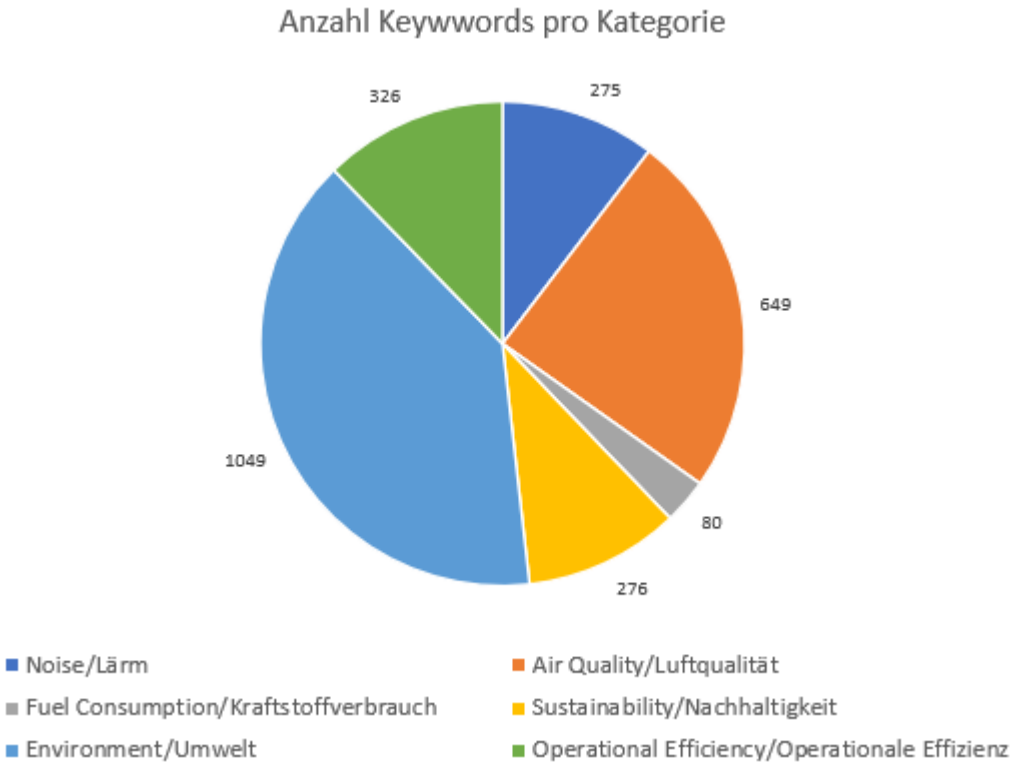
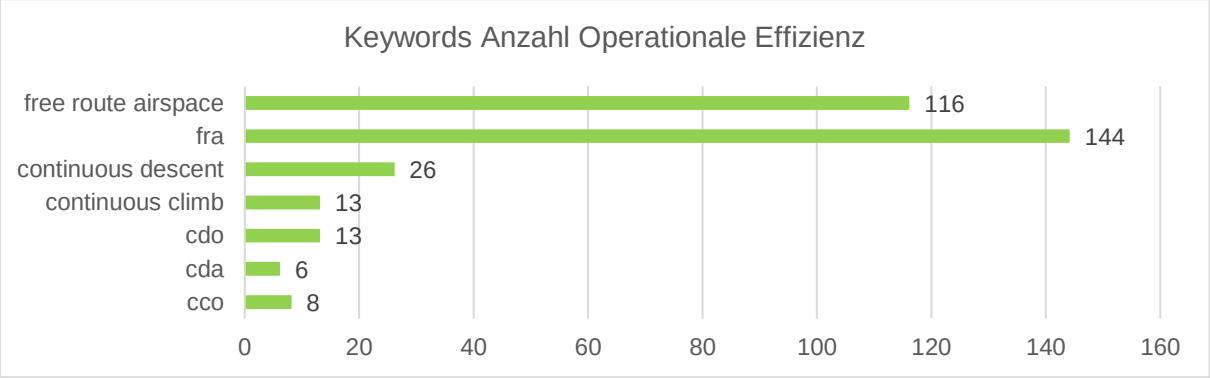
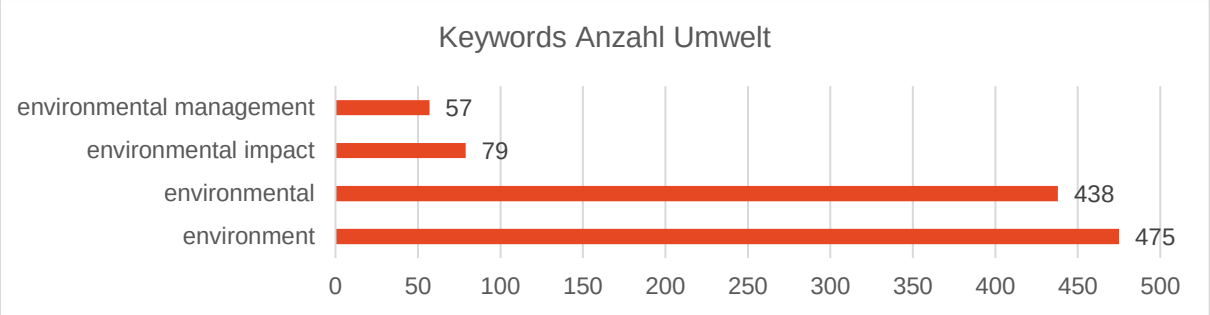


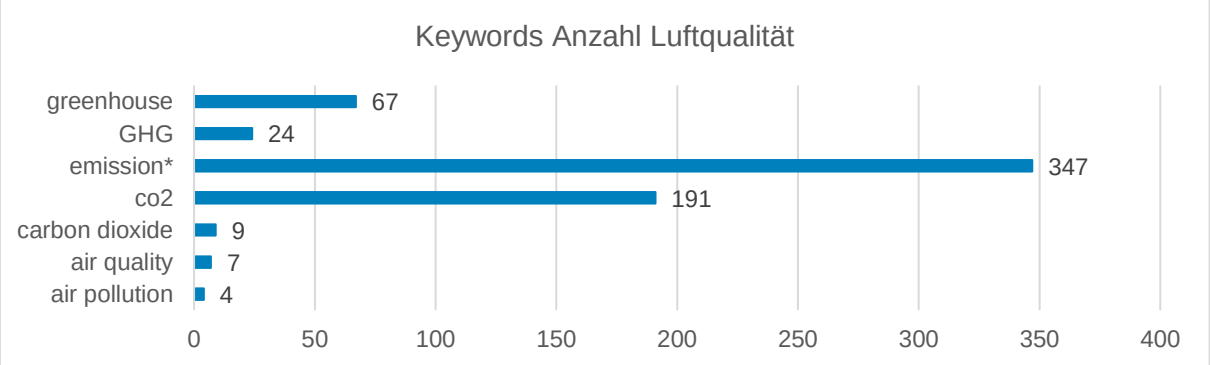
Abbildung 23: Absolute Häufigkeiten der Keywords pro Kategorie

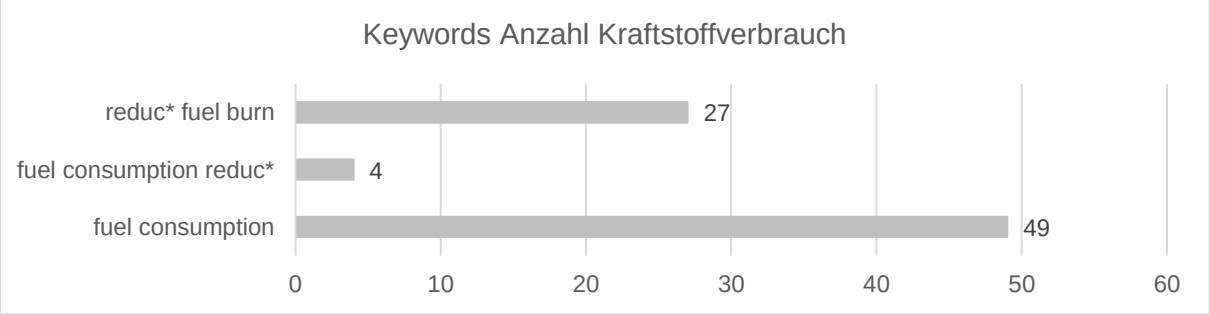
Quelle: Eigene Darstellung

Um über alle verwendeten Keyword Kategorien einen ganzheitlichen Eindruck zu erlangen, sind in der Abbildung 24 mehrere Balkendiagramme veranschaulicht, welche die Verteilung der Anzahl der Keywords innerhalb jeder Kategorie präsentieren. Somit sind die Top-Keywords je Kategorie auf den ersten Blick ersichtlich und somit auch abzuschätzen, in welchem Kontext in anschließenden Analysen mit den meisten relevanten Informationen gerechnet werden kann.









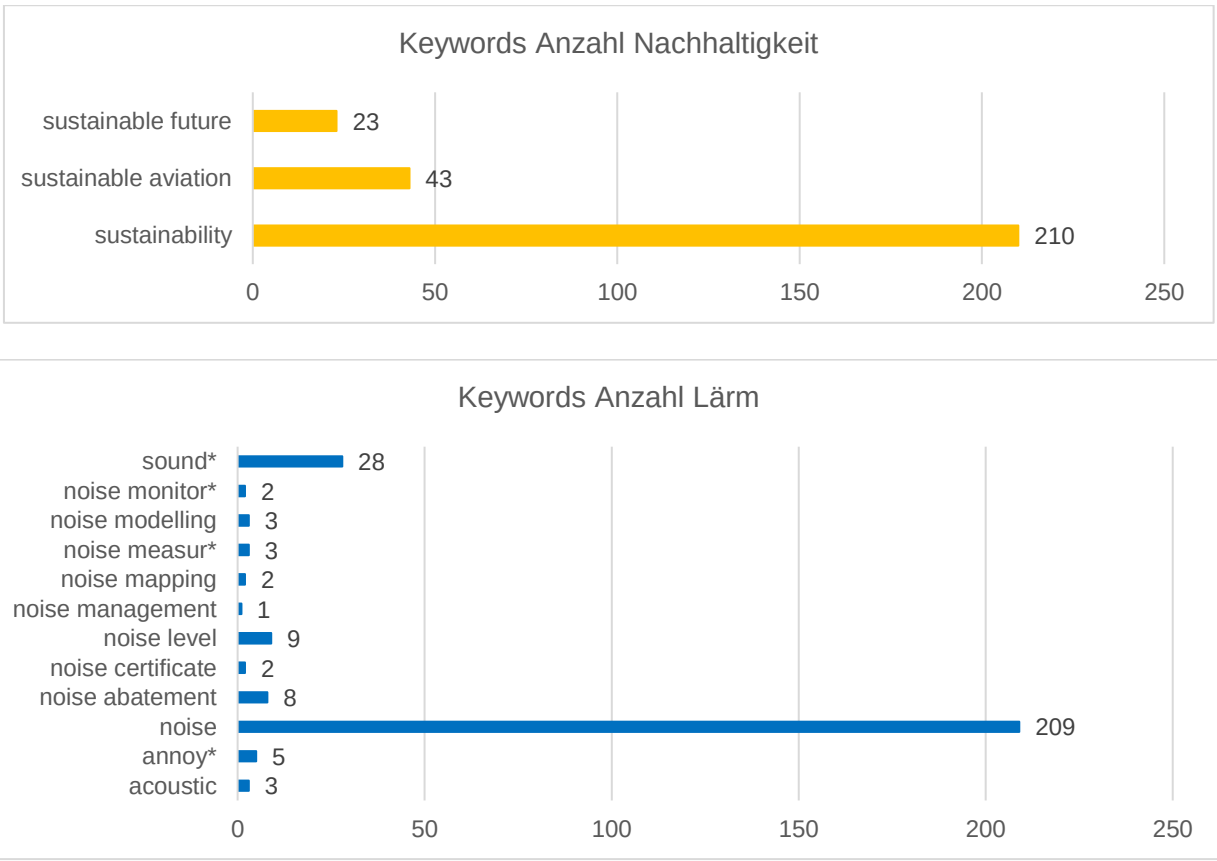


Abbildung 24: Häufigkeiten der Schlüsselbegriffe pro Kategorie

Quelle: Eigene Darstellung

Die Auswertung der letzten Kategorie „Reports“ wurde in den bisherigen Darstellungen bewusst weggelassen, da es sich hier nicht um eine Keyword Suche im eigentlichen Sinne handelt, sondern lediglich um die Identifikation, ob derartige Berichte genannt beziehungsweise von den ANSPs verfasst werden. Die Auswertung kann hier also nicht als Zählung der Vorkommnisse der Begriffe verstanden werden, sondern vielmehr als zutreffend oder nicht zutreffend. Die Auswertung dieser Kategorie hat ergeben, dass acht der insgesamt 32 analysierten europäischen ANSPs Umweltberichte verfassen und auch diese auf deren Webseiten erwähnen oder zum Download zur Verfügung stellten. Diese ANSPs sind untenstehend aufgeführt.

Tabelle 5: ANSPs mit Umweltberichten

ANSP	Homepage	Land
Avinor Flysikring	www.avinor.no	Norwegen
Croatia Control	www.crocontrol.hr	Kroatien
ENAV	www.enav.it	Italien
NATS	www.nats.aero	United Kingdom
NAVIAIR	www.naviair.dk	Dänemark
PANSA	www.pansa.pl	Polen
Skyguide	www.skyguide.ch	Schweiz
ISAVIA	www.isavia.is	Island

Die Umweltberichte dieser ANSPs wurden anschließend von den Webseiten heruntergeladen und nach relevanten Informationen durchsucht. Wurden im Sinne der vorliegenden Forschung relevante Aspekte entdeckt, wurden diese anschließend in die Datenbasis des jeweiligen ANSPs aufgenommen.

Im finalen Schritt der Keyword-Analyse wurde versucht die eigentlich relevanteste Information aus diesen Ergebnissen zu extrahieren. In diesem Schritt wurde analysiert, inwiefern die Thematik der Umwelt im gesamten extrahierten Inhalt der ANSPs eine Rolle spielt. Hierfür wurde die gesamte Anzahl an extrahierten Paragraphen in das Verhältnis mit jenen Paragraphen gesetzt, welche einen Bezug zum in Frage stehenden Thema aufweisen. Als Basis hierfür wurden die Ergebnisse der Keyword-Analyse verwendet. Diese Information ist deshalb relevant, da durch sie ein Eindruck gewonnen werden kann, ob mit relevanten Ergebnissen in der Datenbasis gerechnet werden kann. Ist die gesamte Thematik der Nachhaltigkeit und des Umweltschutzes im Verhältnis zum gesamten Inhalt relativ gering repräsentiert, kann davon ausgegangen werden, dass die Menge an relevanten Informationen innerhalb der Datenbasis überschaubar ist. Dahingegen kann bei einem hohen Verhältnis darauf geschlossen werden, dass die in Frage stehende Thematik durchaus auf den Webseiten der ANSPs behandelt wird und somit auch zahlreiche relevante Informationen gefunden werden können. Umso stärker die Thematik des Umweltschutzes in der Datenbasis repräsentiert ist, mit umso mehr Informationen und Ergebnissen kann logischerweise gerechnet werden. Das Ergebnis dieser Analyse kann in Abbildung 25 eingesehen werden.

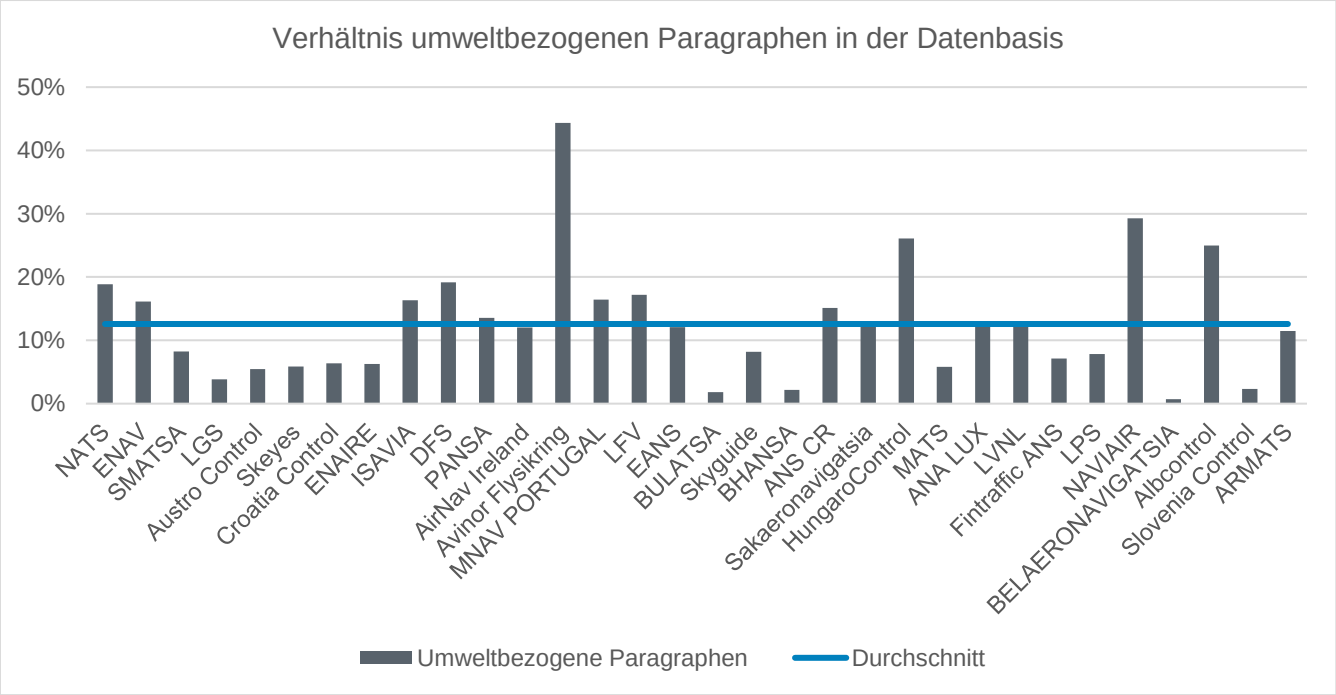


Abbildung 25: Anteil umweltbezogene Paragraphen

Quelle: Eigene Darstellung

Wie aus Abbildung 25 zu entnehmen ist, ist das Thema der Umwelt beziehungsweise der verwendeten Keywords unterschiedlich stark oder schwach in der Datenbasis der unterschiedlichen ANSPs vertreten. Nichtsdestotrotz widmen die ANSPs im Durchschnitt ca. 13,0% der extrahierten Inhalte der Thematik der Umwelt. Wie des Weiteren gesehen werden kann, ist diese Zahl bei einigen ANSPs

weitaus höher. Mit einem Ergebnis von durchschnittlich knapp 13,0% kann durchaus festgehalten werden, dass die relevante Thematik von den ANSPs auf deren Webseiten behandelt wird. Dies erhöht wiederum die Zuversicht, die relevanten Techniken und Systeme in den Textdaten zu identifizieren, da der Thematik offensichtlich Aufmerksamkeit geschenkt wird. Dies stellt also eine gute Ausgangslage für den nächsten Schritt dar. Im nächsten Schritt geht es nun um die eigentliche Identifikation der angewendeten Systeme, Techniken und Praktiken, welche die europäischen ANSPs zur Reduzierung der negativen Auswirkungen des Luftverkehrs auf die Umwelt anwenden.

6.2. Identifizierung der Praktiken und Systeme

Der finale Schritt der Text-Mining-Analyse verfolgt das Ziel der Identifizierung der eigentlichen Systeme, Techniken oder Praktiken, welche von den analysierten ANSPs derzeit angewendet werden, um die negativen Auswirkungen des Luftverkehrs auf die Umwelt und das Klima zu reduzieren oder zu vermeiden. Auch in diesem Schritt bilden die zuvor extrahierten Textinformationen der ANSPs den Grundstein der Analyse. Des Weiteren wurde sich auch in diesem Schritt mittels der Programmiersprache R geholfen. Nichtsdestotrotz ist für eine derartige Analyse aber auch viel manuelle Investigation nötig. Dies liegt vor allem daran, dass es nicht von Beginn an klar definiert ist, wonach eigentlich gesucht wird. Eine stichprobenartige manuelle Investigation der Webseiten hat ebenfalls ergeben, dass viele ANSPs ähnliche oder gleiche Techniken unterschiedlich benennen. Zur Identifikation dieser wird also manuelle Analyse gefragt sein. Die gesammelte Methodik für diesen Schritt wird im folgenden Unterkapitel vorgestellt.

6.2.1. Methodik

Die Methodik dieses finalen Schrittes der Text Mining Analyse enthält die folgenden Schritte. Im ersten Schritt geht es darum, die relevanten Paragraphen aus den Textdaten der verschiedenen ANSPs zu extrahieren. Besondere Herausforderung besteht hierbei darin, alle relevanten Informationen zu extrahieren und gleichzeitig so viel wie möglich irrelevante Daten auszuschließen um den Aufwand einer Analyse möglichst gering zu halten. Nach der Extrahierung dieser Paragraphen erfolgt eine detaillierte Analyse dieser, welche die angewandten Techniken, Systeme und Praktiken der ANSPs zum Vorschein bringen sollte. Im ersten Schritt wurde sich, wie zuvor bereits erwähnt, mit einem in der Programmierumgebung R verfassten Code Unterstützung verschafft.

Zu Beginn wird im unten ersichtlichen Code der Text aus einer Datei gelesen und anschließend in Kleinbuchstaben umgewandelt. Die Segmentierung des Textes erfolgt in Sätzen, wobei nicht nur der Punkt, sondern auch das Ausrufezeichen und das Fragezeichen als Satzenden betrachtet werden.

Die eigentliche Funktion dieser Code-Abschnitte besteht darin, Sätze nach vordefinierten Schlüsselwörtern zu durchsuchen und den Kontext um die gefundenen Sätze zu extrahieren. Hierbei wird für jedes identifizierte Schlüsselwort der vorherige, aktuelle und nachfolgende Satz gespeichert. Dieser Kontext wird strukturiert in einer Liste abgelegt.

Die Ausgabe erfolgt durch das Durchlaufen der gefundenen Sätze und ihrer Kontexte. Dabei wird darauf geachtet, dass keine zusätzlichen Leerzeilen eingefügt werden, wenn kein vorheriger oder folgender Satz vorhanden ist. Die Ausgabe der identifizierten Inhalte erfolgt daher gruppiert nach den verwendeten Begriffen. Diese Funktionalität ermöglicht eine detaillierte Analyse von Texten im Hin-

blick auf spezifische Schlüsselwörter, wobei der Kontext für eine weiterführende Untersuchung bereitgestellt wird. Wie in Listing 9 dargestellt, kann der Code beispielsweise für die Extrahierung des Kontextes rund um die Thematiken der Schlüsselbegriffe Umwelt und Nachhaltigkeit verwendet werden.

Listing 9: Code Identifikation relevanter Techniken

```
# Einlesen der Textdateien
setwd(Pfad zum Arbeitsverzeichnis mit den Textdokumenten)
text <- readLines("Beispieldatei")

# Liste von Keywords erstellen
keywords <- c("umwelt", "nachhaltigkeit")
text <- paste(text, collapse = " ")

# Text vorverarbeiten - in Kleinbuchstaben konvertieren
text <- tolower(text)

# Text in Sätze aufteilen - Punkt, Ausrufezeichen und Fragezeichen als Trennzeichen
sentences <- unlist(strsplit(text, "[\\.\\?!] "))

# Initialisieren einer Liste um Sätze und Kontext darin zu speichern
matching_sentences <- list()

# Funktion zur Überprüfung ob ein Satz ein Keyord enthält - gegebenenfalls mit Wildcard
contains_keyword <- function(sentence, keyword) {
  keyword_pattern <- gsub("\\*", ".*", keyword)
  return(grepl(paste0("\\b", keyword_pattern, "\\b"), sentence, ignore.case = TRUE))
}

# Suche nach Sätzen mit Keywords und Extraktion des Kontextes darumherum
for (i in 1:length(sentences)) {
  sentence <- sentences[i]
  for (keyword in keywords) {
    if (contains_keyword(sentence, keyword)) {
      context <- list()

      # Kontext vor dem aktullen Keyword-Satz
      if (i > 1) {
        context$previous <- sentences[i - 1]
      }
```

```
      # Aktueller Keyword-Satz
      context$current <- sentence

      # Kontext nach dem aktullen Keyword-Satz
      if (i < length(sentences)) {
        context$next_sentence <- sentences[i + 1]
      }

      if (!exists(keyword, where = matching_sentences)) {
        matching_sentences[[keyword]] <- list()
      }

      matching_sentences[[keyword]][[sentence]] <- context
    }
  }
}

# Schleife über identifizierte Sätze und Keyword - Ausgabe
for (keyword in keywords) {
  if (exists(keyword, where = matching_sentences)) {
    cat("KONTEXT FÜR SCHLÜSSELWORT: ", keyword, "\n")
    for (sentence in names(matching_sentences[[keyword]])) {
      context <- matching_sentences[[keyword]][[sentence]]

      if (!is.null(context$previous)) {
        cat(context$previous, "\n")
      }

      if (!is.null(context$current)) {
        cat(context$current, "\n")
      }

      if (!is.null(context$next_sentence)) {
        cat(context$next_sentence, "\n")
      }

      cat("\n")  # Leere Zeile zwischen Einträgen
    }
    cat("\n")  # Leere Zeile zwischen Einträgen
  }
}
```

Wurden die Textdokumente der spezifischen ANSPs durchlaufen, ist dadurch wiederum ein neues, verkürztes Textdokument entstanden. Anschließend geht es nun darum, die relevanten Praktiken, Techniken und Systeme zu identifizieren, welche derzeit von den europäischen ANSPs angewendet werden. Hierfür ist zumindest zu Beginn dieses Prozesses manuelle Investigation nötig, da nicht von Beginn an klar ist, nach welchen Namen oder Ausdrücken eigentlich gesucht wird. Das Ergebnis des Codes in Listing 9 beziehungsweise die manuelle Investigation zur Identifikation der Techniken, ist in Abbildung 26 veranschaulicht. In Abbildung 26 sind beispielhaft zwei extrahierte Textabschnitte zu sehen. Der manuelle Prozess zur Identifikation relevanter Techniken, nachdem der oben präsentierte Code in Listing 9 die Textabschnitt extrahiert hat, wird veranschaulicht.

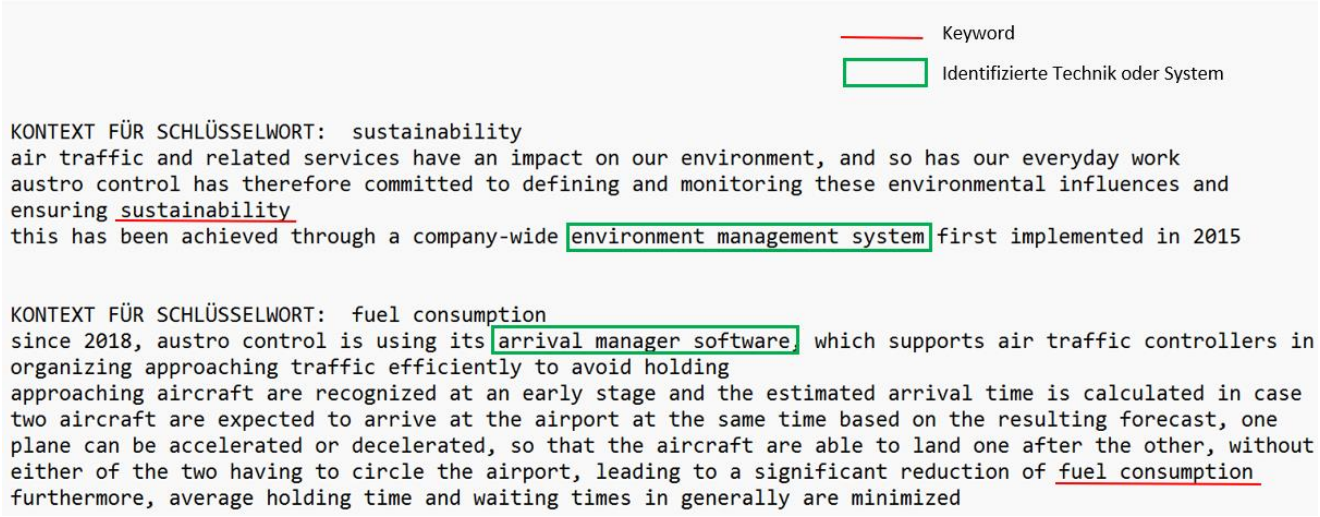


Abbildung 26: Identifikation relevanter Techniken und Systeme

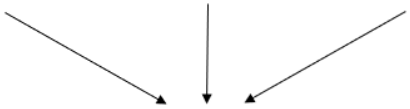
Quelle: Eigene Darstellung

Wie in Abbildung 26 veranschaulicht, wurden auf diese Art und Weise die relevanten Praktiken und Systeme der ANSPs identifiziert. Dieser Prozess wurde auf alle zur Verfügung stehenden Textdaten angewendet. Das Ergebnis hierbei zeigte eine Liste mit allen Praktiken und Systemen je ANSP. Um folgend eine finale Übersicht zu erhalten, wurden diese Praktiken und Systeme zu einer einzigen Liste über alle europäische ANSPs hinweg zusammengefasst. Diese finale Liste repräsentiert die zusammengefassten Techniken und Praktiken der ANSPs, zusammen mit deren Anwendungshäufigkeit. Die Häufigkeit wird bei der anschließenden Erstellung der finalen Übersicht, sowie bei der Erstellung des Fragebogens eine Rolle spielen.

6.2.2. Ergebnisse

Ein Auszug der im vorherigen Schritt erstellten Listen, ist in Abbildung 27 sichtbar. Folgend wird der soeben erklärte Prozess visuell präsentiert. Es sind einige identifizierte Techniken von drei verschiedenen Flugsicherungsorganisationen zu sehen. Diese individuellen Listen werden anschließend zu einer einzigen Liste zusammengefasst, in der auch die absolute Häufigkeit der Anwendung der Praktiken und Systeme festgehalten wird. Dieses Vorgehen wird natürlich auf alle zur Verfügung stehenden individuellen Listen angewendet.

NATS - UK	AUSTROCONTROL - AUSTRIA	DFS - GERMANY
Environmental Management System	Free Route Airspace	NIROS - Noise Impact Reduction and Optimisation System
Continuous Descent	Arrival Management Software	Noise Monitoring
Free Route Airspace	A-CDM Airport Collaborative Decision Making	Free Route Airspace
Efficient Route Design	Direct and Shorter Routes	Continous Climb
PBN - Performance Based Navigation	...	Intelligent Flight Routes
Intelligent Approach Support System		Precision Appraoch System
...		...



Techniken und Systeme	Anwendungshäufigkeit
Route Optimisation Procedures/Systems	24
Free Route Airspace	20
Continuous Descent	17
Environmental Management System (EMS)	18
Single European Sky	16
Arrival Management/Decision Software + GBAS	15
ISO 14001	13
Continuous Climb	9

Abbildung 27: Integration der Techniken in eine Liste

Quelle: Eigene Darstellung

Die dadurch erstellte Liste ist natürlich noch sehr umfangreich, vor allem unterschiedliche Bezeichnungen der Techniken und Systeme der ANSPs, haben dies zu verschulden. Um eine übersichtliche, und auch für den Fragebogen verwendbare Liste zu generieren, wurden diese Praktiken und Systeme nach ihrem Anwendungsbereich und den verfolgten Zielen zusammengefasst. Dies ist im Rahmen einer zweiten Absprache mit Herrn Emir Ganić erfolgt. Die finalen Systeme und Techniken in Originalbezeichnung sowie die übergeordnete Kategorie, können Tabelle 6 entnommen werden.

Tabelle 6: Finale Liste der aktuell verwendeten Techniken

MERGED TECHNIQUES
Operational Efficiency
Route Optimisation Procedures Free Route Airspace Continuous Descent/Climb Operations Curved Approaches
Environmental Management
Environmental Management System (EMS)
Decision Support Software and Air Traffic Planning
Arrival Management/Decision Support Software Departure Management/Decision Support Software
Environmental Monitoring and Modelling
Noise Monitoring/Mapping/Modelling Noise Impact Reduction and Optimisation System

Es wurde ebenfalls besonders darauf geachtet, einen ganzheitlichen Überblick der angewandten Techniken und Systeme zu erhalten, weswegen auch nur als relevant empfundene Praktiken in die finale Liste aufgenommen wurden. Für die Erstellung der zusammengeführten Liste wurde einerseits die Häufigkeit der Anwendung der spezifischen Techniken und Systeme berücksichtigt, sowie auch die Experteneinsichten von Herrn Ganić. Dabei ging es vor allem um besonders effektive oder auch neuartige Methoden und Praktiken. Die finale Liste stellt nun die Ausgansbasis für das weitere Vorgehen dar. Im nächsten Schritt geht es darum, eine Kontrolle dieser Ergebnisse von Repräsentativen der ANSPs einzuholen, um auf die Richtigkeit der identifizierten Liste schließen zu können. Darüber hinaus wird dies ebenfalls eine Grundlage darstellen, Aussagen über den Web-Scraping Prozess abzuleiten. Um diesen angestrebten Vergleich einzuholen, wurde sich für die Methodik eines Fragebogens entschieden, welcher im Anschluss an Repräsentanten der europäischen ANSPs gesendet wird.

Um die Web-Scraping- und Text-Mining-Ergebnisse sowie die Resultate des Fragebogens kohärent darzustellen und mit der Dokumentensprache in Einklang zu bringen, wird nachfolgend Tabelle 7 präsentiert. Diese Tabelle zeigt denselben Inhalt wie Tabelle 6, jedoch in deutscher Übersetzung, insofern es für die konkreten Techniken eine passende deutsche Übersetzung gibt. Für die Durchführung der Umfrage wurden die Bezeichnungen in englischer Sprache verwendet. Die Analyse und folgende Ergebnispräsentationen erfolgen jedoch mit den deutschen Übersetzungen, um eine einheitliche Sprache im Dokument sicherzustellen.

Tabelle 7: Deutsche Übersetzung der finalen Liste

ZUSAMMENGEFÜHRTE TECHNIKEN
Operationale Effizienz
Routenoptimierungsverfahren Free Route Airspace Kontinuierliche Steig-/Sinkflugverfahren Kurvenanflugverfahren
Umweltmanagement
Umweltmanagementsystem
Entscheidungshilfesoftware und automatisierte Luftverkehrsplanung
Ankunftsmanagement-/Entscheidungsunterstützungssoftware Abflugmanagement-/Entscheidungsunterstützungssoftware
Umweltüberwachung und -modellierung
Lärmüberwachung/Lärmmodellierung/Lärmkartierung Noise Impact Reduction and Optimisation System (NIROS)

7. Umfrage unter Air Navigation Service Provider

Vor der Erstellung der Fragebogens wurden die eigentlichen Ziele beziehungsweise Fragen definiert, welche mittels des Fragebogens beantwortet werden sollten. Augenmerk liegt darauf, den tatsächlichen aktuellen Stand im Sinne von derzeit verwendeten Techniken und Systemen, zur Reduzierung der negativen Umweltauswirkungen des Luftverkehrs, der europäischen ANSPs abzuleiten. Darüber hinaus sollten auch die Web-Scraping-Ergebnisse beziehungsweise Text-Mining-Ergebnisse in den Fragebogen integriert werden, um eine anschließende Vergleichbarkeit der Resultate aus Text-Mining und dem Fragebogen zu gewährleisten. Darüber hinaus soll der Fragebogen einen möglichst komprimierten Umfang ausweisen, um die Barrieren für die Teilnahme so gering wie möglich zu halten.

7.1. Entwicklung der Umfrage

Um die definierten Ziele des Fragebogens zu erreichen, wurde der Fragebogen gemäß Abbildung 28 in drei Teile strukturiert. Anschließend an Abbildung 28 folgt eine ausführlichere Beschreibung der drei Sektionen sowie deren Ziele und Absichten.

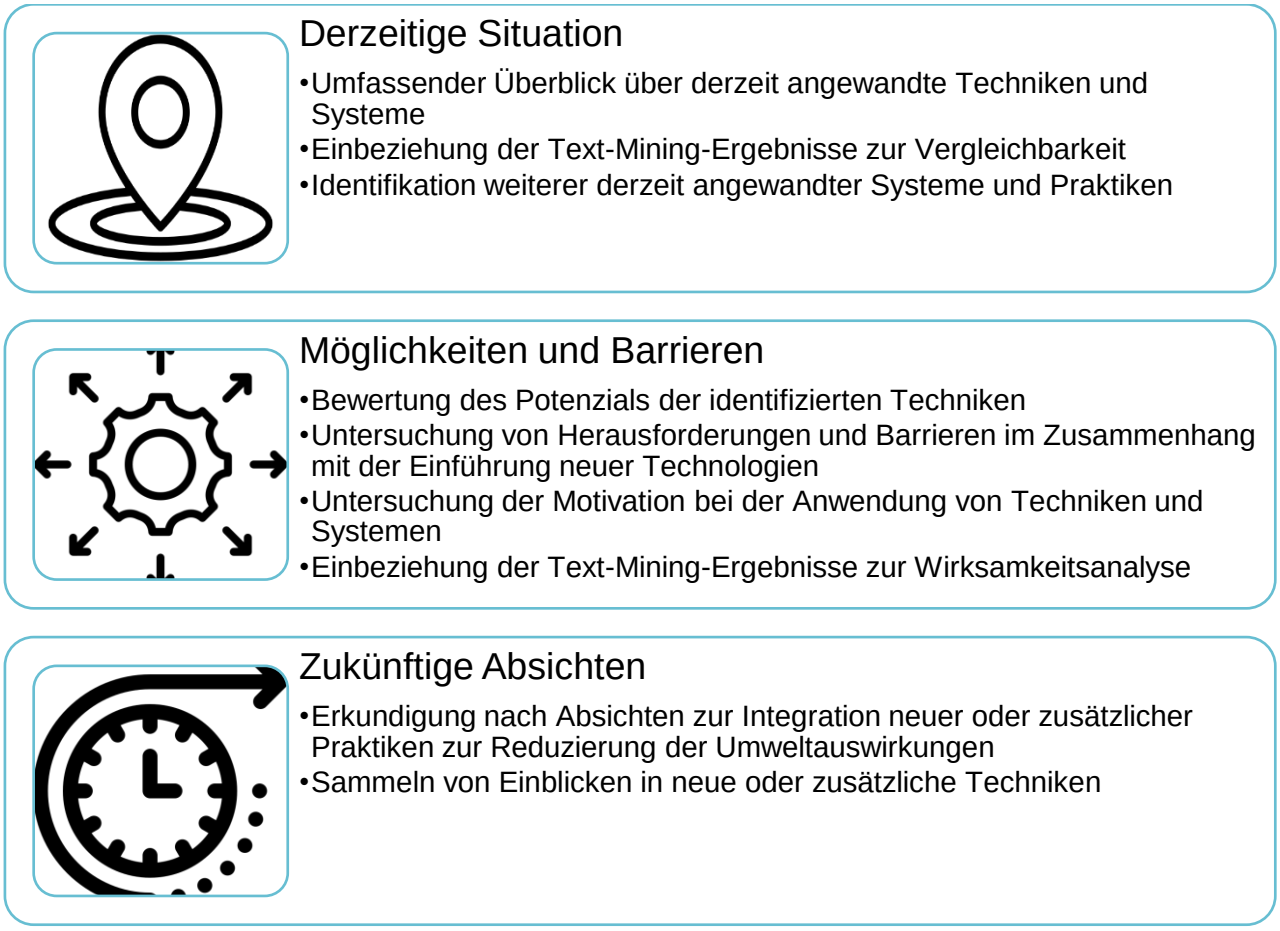


Abbildung 28: Aufbau und Sektionen des Fragebogens

Quelle: Eigene Darstellung

Derzeitige Situation: Diese erste Sektion im Fragebogen verfolgt fokussiert jenes Ziel, Einblick in die derzeitige Situation der ANSPs zu erlangen. Dies bietet eine hervorragende Möglichkeit, die mit-

tels Text Mining identifizierte Liste miteinzubinden, da auch diese einen aktuellen Stand der Techniken repräsentiert. Die mittels Text Mining identifizierte Liste wurde in dieser Sektion in den Fragebogen eingebunden, und es kann von den Teilnehmerinnen und Teilnehmern angegeben werden, ob die genannten Techniken und Systeme verwendet werden oder nicht. Natürlich bietet diese Sektion auch die Möglichkeit weitere verwendete Techniken und Systeme, welche derzeit in der Organisation Anwendung finden, zu nennen und zu beschreiben. Darüber hinaus wird in dieser Sektion auch über die Motivation zur Anwendung oder Implementierung der Systeme und Techniken gefragt. Somit ist sichergestellt, dass diese Sektion den aktuellen Stand im Sinne von angewandten Praktiken zur Reduzierung der negativen Umwelteinflüsse durch den Luftverkehr zum Vorschein bringt. Zusammengefasst beinhaltet dies die aktuell verwendeten Techniken und Systeme, über die Text-Mining-Liste hinaus, sowie die Motivation für deren Anwendung. Durch die Einbindung der Text-Mining-Ergebnisse wird die gewünschte Vergleichbarkeit sichergestellt.

Möglichkeiten und Barrieren: Diese Sektion verfolgt zwei Ziele. Einerseits die Ermittlung der Barrieren welchen ANSPs gegenüberstehen bei der Implementierung von neuen Techniken und Systemen, sowie auch die Ableitung der Wirksamkeit der Techniken und Systeme in der Text-Mining-Liste. Dies wurde erreicht indem einerseits nach den Barrieren gefragt wird, mit einigen bereitgestellten Antwortmöglichkeiten, sowie auch natürlich der Möglichkeit eigene Antworten zu definieren. Zur Erreichung des zweiten Zieles wurde wiederum die Text-Mining-Liste in die Sektion integriert. Diesmal können Teilnehmerinnen und Teilnehmer die wahrgenommene Wirksamkeit der Techniken und Systeme auf einer Skala von sehr hoch bis sehr niedrig beurteilen.

Zukünftige Absichten: Diese Sektion verfolgt das Ziel, Einblicke in die zukünftigen Absichten der ANSPs zu erlangen. Es dreht sich hierbei also um die Frage, ob die ANSPs innerhalb der nächsten fünf Jahre planen, neue oder zusätzliche Praktiken zum Umwelt- und Klimaschutz im Unternehmen zu integrieren. Dadurch kann einerseits auf vielversprechende oder zukünftige Techniken aufmerksam gemacht werden, sowie auch abgeschätzt werden, ob die ANSPs eine hohe Intention aufzeigen bezüglich weiterer Techniken und Systeme für den Umweltschutz.

Insgesamt ist es gelungen, diese Sektionen und die damit hergehenden Ziele mittels nur neun Fragen in den Fragebogen zu integrieren. Besonderes Augenmerk wurde dabei auf kurze und prägnante Antworten gelegt. In Summe wird mittels diesen drei Sektionen der gesamtheitliche aktuelle Stand der europäischen ANSPs abgeleitet. Einerseits stellt jede Sektion für sich interessante Einblicke zur Verfügung, darüber hinaus wird aber auch die Analyse der Zusammenhänge zwischen den Sektionen neue Einblicke in derzeitige Situation ermöglichen. Die spätere Möglichkeit zum Vergleich mit den Text-Mining-Ergebnissen wurde sichergestellt, indem diese direkt in den Fragebogen integriert wurden.

7.2. Durchführung der Umfrage

Nachdem das Design der zu erstellenden Umfrage feststeht, wurde diese im Anschluss mithilfe des Online-Tools LimeSurvey erstellt (LimeSurvey, o. J.). Bei LimeSurvey handelt es sich um eine Open-Source-Software zur Erstellung und zur Verwaltung von Umfragen und Fragebögen. Mithilfe dieser Software können Benutzerinnen und Benutzer benutzerdefinierte Umfragen erstellen, und diese anschließend über verschiedenen Kanäle wie das Web oder E-Mail verteilen (LimeSurvey, o.J.). Die

Software bietet darüber hinaus auch noch eine Vielzahl an weiteren Funktionen, darunter die Möglichkeit, komplexe Fragebögen mit verschiedenen Fragetypen zu erstellen, die Integration von Bildern und Multimedia-Inhalten, die Unterstützung für mehrsprachige Umfragen, sowie auch die Anpassung von Layout und Design (LimeSurvey, o. J.).

LimeSurvey mit dessen vielen Funktionalitäten stellt damit eine perfekte Softwarelösung dar, um den zuvor designten Fragebogen zu erstellen und auszusenden. Die Umfrage wurde entweder mittels E-Mail an die ANSPs gesendet oder durch ein Kontaktformular, das direkt auf den Webseiten der ANSPs eingebunden war. Vor dem Aussenden mussten dann nur noch die mitgeschickten Einladungstexte zur Teilnahme verfasst werden. Hierbei handelt es sich einerseits um den Text für die erstmalige Aussendung, sowie um den Text für eine Erinnerungsmail, falls nach einiger Zeit noch keine Antwort erhalten wurde. Nach der Fertigstellung dieser Texte wurde mit dem Aussenden der Umfrage and die europäischen ANSPs begonnen.

7.3. Ergebnisse

Die Ergebnisse des Fragebogens werden zunächst innerhalb der drei definierten Sektionen der Umfrage einzeln analysiert. Im Anschluss daran, werden mögliche Zusammenhänge zwischen den Sektionen aufgezeigt. Von den insgesamt 32 angefragten europäischen ANSPs aufgelistet in Tabelle 2 haben 15 ANSPs innerhalb des Befragungszeitraums vollständige Antworten geliefert. Dies entspricht einer Rücklaufquote von knapp 47 %.

7.3.1. Derzeitiger Stand

Um eine kurze und übersichtliche Schreibweise in der Analyse des Fragebogens zu gewährleisten, werden untenstehend nochmals die vier Kategorien der Techniken und Systeme aufgelistet. Diesen werden Nummern zugewiesen, um eine übersichtliche Beschriftung in Abbildungen zu gewährleisten. Dadurch ist es auf den ersten Blick zu erkennen, welche spezifischen Methoden oder Praktiken, welcher übergeordneten Kategorie angehören.

- Kategorie 1: Operationale Effizienz
- Kategorie 2: Umweltmanagement
- Kategorie 3: Entscheidungshilfesoftware und automatisierte Luftverkehrsplanung
- Kategorie 4: Umweltüberwachung und -modellierung

Das Hauptziel der ersten Sektion im Fragebogen ist es, die aktuell verwendeten Praktiken und Systeme der ANSPs zu identifizieren, welche zur Reduzierung der negativen Umweltauswirkungen des Luftverkehrs angewendet werden. Die Anwendungsraten der zuvor durch Text Mining identifizierten Techniken und Systeme sind in Abbildung 29 visuell präsentiert. Auf den ersten Blick ist erkennbar, dass vor allem die Techniken in der ersten Kategorie (1), jener der Operationalen Effizienz, besonders hohe Anwendungsraten zugutekommen. Ausgenommen davon ist die Methode des sogenannten Curved Approachs beziehungsweise des Kurvenanflugverfahrens. Im Vergleich mit den restlichen Anwendungsraten innerhalb dieser Kategorie weist das Kurvenanflugverfahren mit knapp 33% Zustimmung eine sehr niedrige Anwendung in der Praxis auf. Die Techniken und Systeme der Kategorien Umweltmanagement sowie Entscheidungshilfesoftware und automatisierte Luftverkehrsplanung weisen eine Anwendungsrate von mindestens 60% auf. Auffallend ist, dass die beiden Techni-

ken innerhalb der Kategorie Umweltüberwachung- und -modellierung den letzten und drittletzten Platz im Sinne der Anwendungsraten in der Praxis zugesprochen bekommen. Ebenso steht es um die Ausprägung der Unwissenheit, sprich das die Teilnehmerinnen und Teilnehmer des Fragebogens nicht wussten, ob diese Techniken derzeit in ihrer Luftsicherungsorganisation Anwendung finden oder nicht. Die beiden Techniken aus dieser Kategorie fokussieren sich auf die Reduzierung der Lärmentwicklung durch den Flugverkehr.

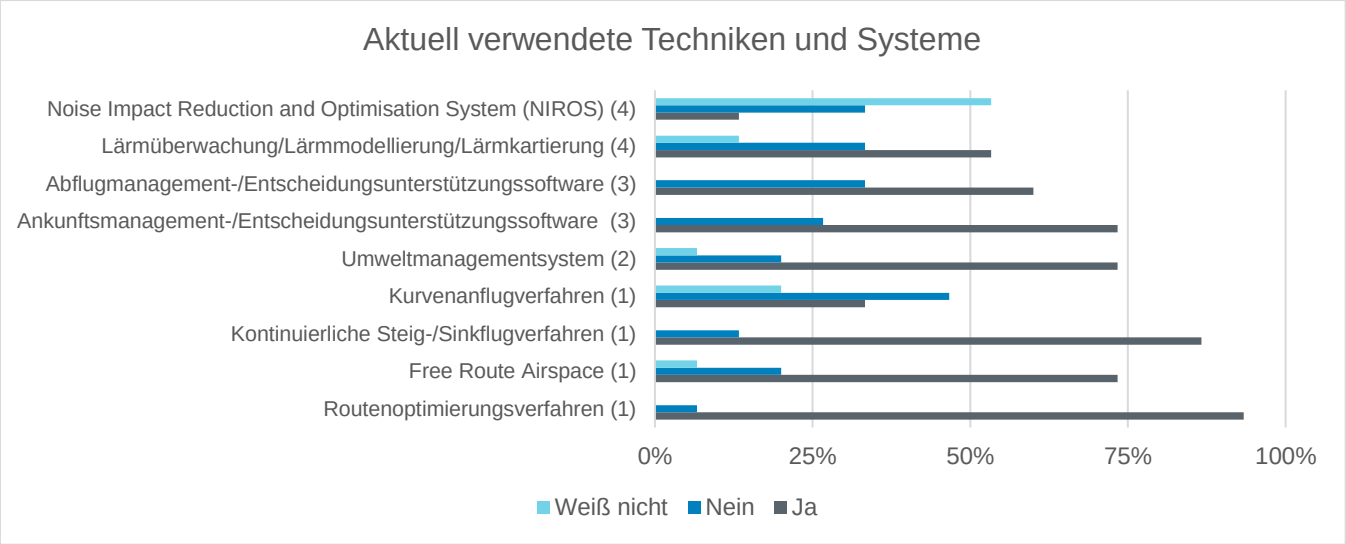


Abbildung 29: Anwendungsraten aktuell verwendeter Techniken und Systeme
Quelle: Eigene Darstellung

Um den derzeitigen Stand zusammenfassend auf der Ebene der Kategorien einsehen zu können, ist dies in Abbildung 30 abgebildet. Diese Abbildung zeigt die Anwendungsraten auf Kategorie-Ebene. Im Durchschnitt weisen die ersten drei Kategorien mit einer Zustimmung von über 66% eine sehr hohe Nutzungsrate auf. Die Kategorie des Umweltmanagement liegt hierbei mit ca. 73% vor der Operationalen Effizienz und der Entscheidungshilfesoftware und automatisierten Luftverkehrsplanung. Auch auf der Ebene der Kategorien fällt die jene der Umweltüberwachung und -modellierung auf. Mit knapp 32% Anwendungsrate belegt diese klar den letzten Platz. Auf dieser Ebene wird auch das Ausmaß der Unwissenheit mit ebenfalls knapp 32% Zuspruch verdeutlicht.

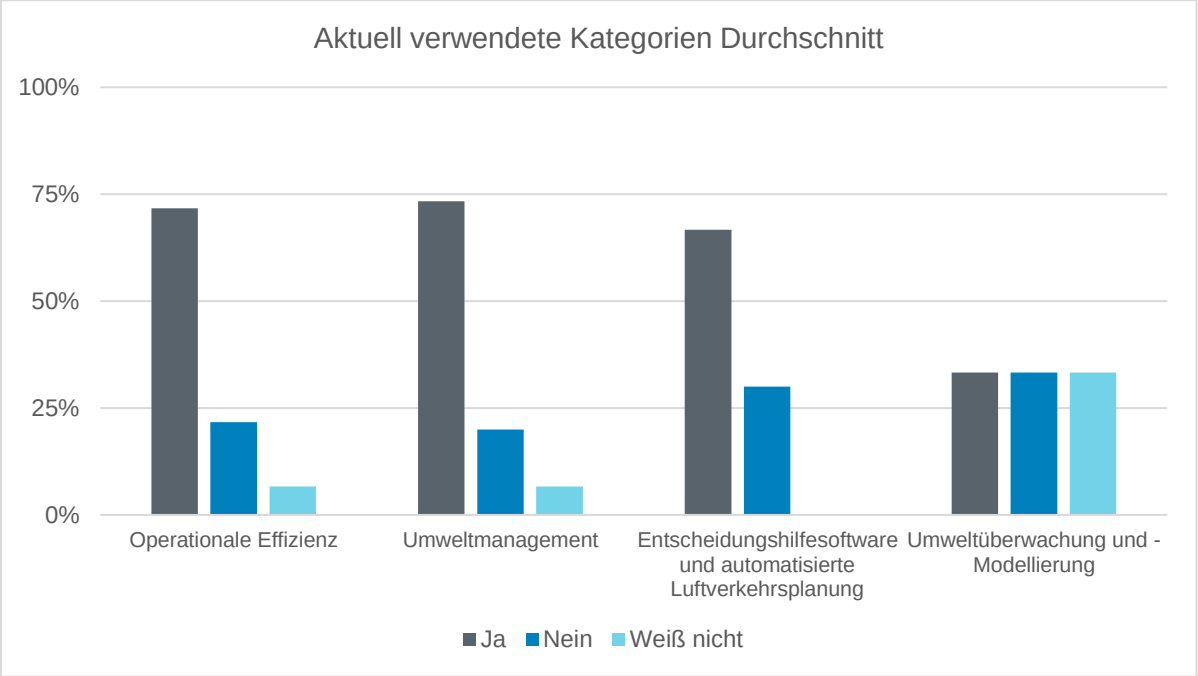


Abbildung 30: Anwendungsraten aktuell verwendeter Kategorien

Quelle: Eigene Darstellung

Ein weiteres Ziel dieser Sektion im Fragebogen war es, weitere aktuell angewandte Techniken und Systeme der ANSPs zum Vorschein zu bringen. Als Antwort wurden zahlreiche unterschiedliche Systeme und Praktiken genannt. Allerdings handelt es sich hierbei um ein sehr breites Spektrum an Praktiken mit unterschiedlichen Zielen. Daher ist es nur schwer möglich ein Muster hierbei zu erkennen. Nichtsdestotrotz können ziemlich eindeutig zwei Kategorien oder zwei verfolgte Ziele der zusätzlich derzeit angewandten Praktiken abgeleitet werden. Dabei ist einerseits die Rede von neuen Software- und Planungssystemen sowie von nachhaltige Initiativen. Beispielhafte Nennungen in dieser Sektion sind folgend zu lesen.

- Verkehrskollisionsvermeidungssystem (TCAS) – Software und Planungssystem
- Erforschung der Nutzung alternativer Kraftstoffquellen – nachhaltige Initiative
- Erneuerbare Energien – nachhaltige Initiative
- Wettervorhersagemodelle – Software und Planungssystem
- Kollaborative Entscheidungsfindungssysteme (A.CDM) – Software und Planungssystem
- Leit- und Steuersysteme – Software und Planungssystem

Der letzte Bereich innerhalb dieser ersten Sektion des Fragebogens zielte darauf ab, die Motivation der ANSPs zum Vorschein zu bringen, warum sie sich überhaupt mit dem Thema Umweltauswirkungen des Luftverkehrs und deren Bekämpfung auseinandersetzen. In Abbildung 31 ist die Auswertung dieser Frage zu beobachten. Folgende vier Ausprägungen haben hierbei die Mehrheit erreicht. Die Steigerung der Betriebseffizienz im Luftverkehr, die Reduzierung des Kraftstoffverbrauchs, sowie die Steigerung der Luftqualität befinden sich mit ähnlichen Zustimmungsraten im Spitzenfeld der Motivationsfaktoren. Etwas weiter zurück ist mit ca. 17% der Motivationsfaktor der Reduzierung der Lärmemissionen zu finden. Dies spiegelt sich auch in den aktuellen Anwendungsraten der Methoden und Praktiken von zuvor wieder.

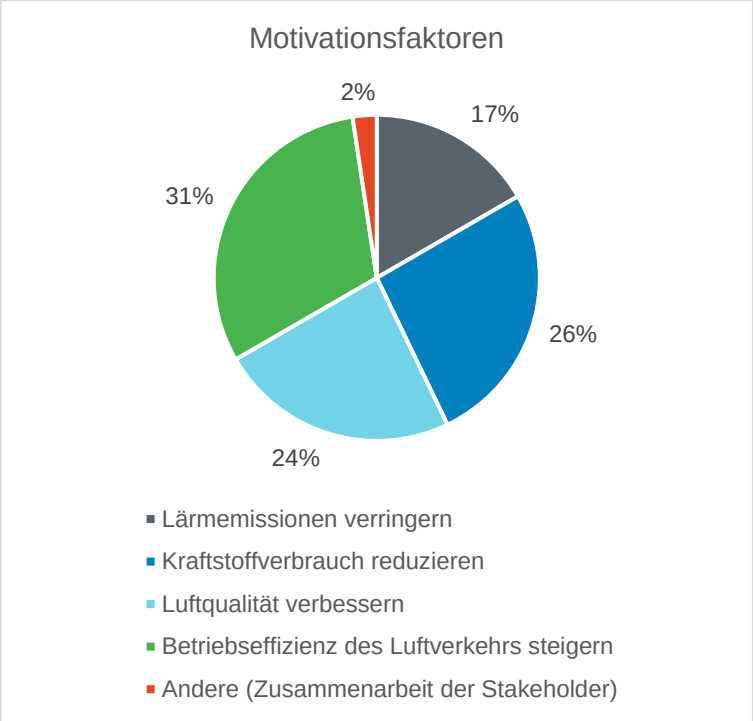


Abbildung 31: Motivationsfaktoren

Quelle: Eigene Darstellung

7.3.2. Potential und Herausforderungen

Die zweite Sektion im Fragebogen behandelte einerseits die Thematik des Potentials der gegebenen Techniken und Systeme sowie die Herausforderungen und Barrieren bei der Implementierung und Planung neuer Praktiken. In dieser Sektion wurden die teilnehmenden ANSPs zuerst gebeten, die durch Text Mining identifizierten Techniken, Praktiken und Systeme nach deren eingeschätzter Wirksamkeit zur Reduzierung der negativen Umweltauswirkungen des Luftverkehrs in Bezug auf Luftqualität, Kraftstoffverbrauch sowie Lärmentwicklung zu beurteilen. Die Beurteilung basierte dabei auf einer Skala von sehr hoch bis sehr niedrig. Anschließend an diese Einschätzung wurden mittels einer folgenden Frage die Herausforderungen und Barrieren identifiziert, welchen die ANSPs bei der Einführung neuer Technologien und Systemen gegenüberstehen.

Folgend wird zunächst die Wirksamkeit der Techniken und Systeme analysiert. Das Potential wird dabei wiederum wie zuvor zuerst auf Ebene der spezifischen Techniken und Systeme dargestellt. In Abbildung 32 sind die einzelnen Techniken und Systeme mit deren beurteiltem Potential dargestellt. Aus dem Diagramm abzulesen ist, dass laut den teilnehmenden ANSPs die Praktik der kontinuierlichen Steig- und Sinkflugverfahren das höchste Potential zur Reduzierung der spezifizierten negativen Umwelteinflüsse aufweist. Gefolgt von den Entscheidungshilfesystemen bei An- und Abflug. Die einzigen Praktiken mit einer Angaben für sehr geringes Potential sind Routenoptimierungsverfahren sowie auch das Noise Impact Reduction and Optimisation System (NIROS) mit jeweils knapp 7%. Auch in Abbildung 32 sticht das Kurvenanflugverfahren ins Auge. Dieses weißt entsprechend dessen geringer Nutzung aus vorheriger Analyse auch geringes Potential auf, und wurde von den teilnehmenden ANSPs mit dem höchsten Wert für niedriges Potential eingestuft. Den höchsten Grad an Unwissenheit zeigt auch in dieser Abbildung das Noise Impact Reduction and Optimisation System (NIROS) mit ca. 43%.

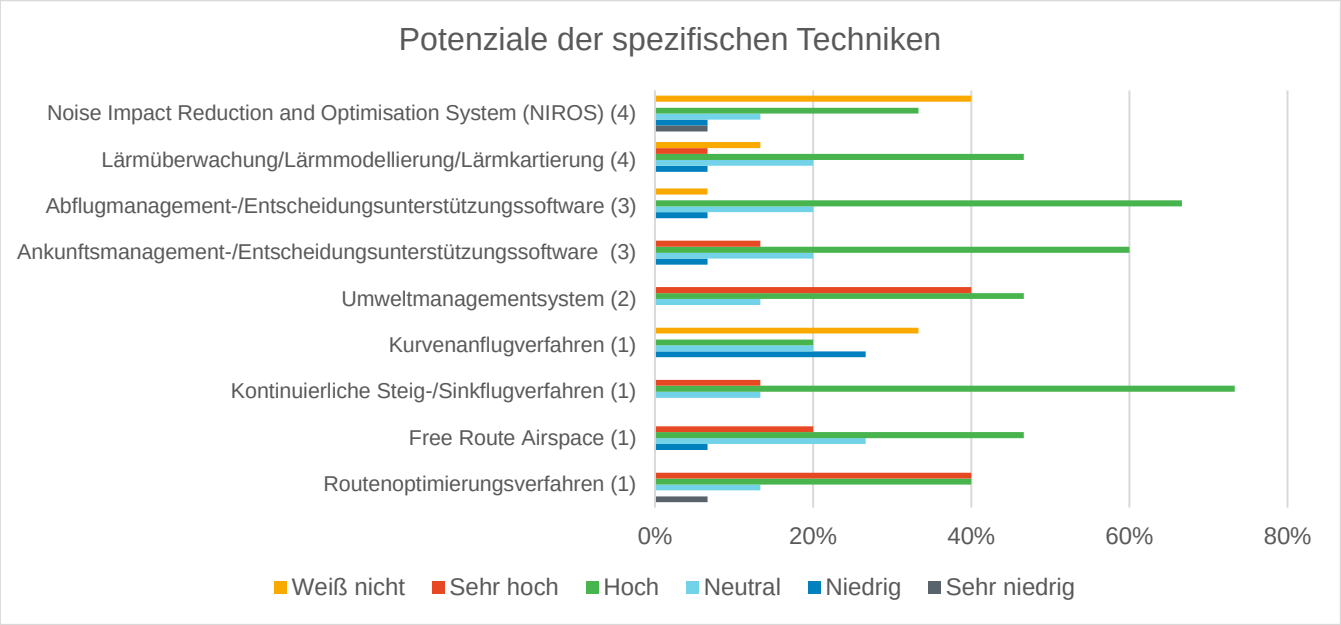


Abbildung 32: Potentiale der spezifischen Techniken und Systeme

Quelle: Eigene Darstellung

Um eine bessere Übersicht und Interpretierbarkeit des Potentials der spezifischen Systeme und Methoden zu ermöglichen, ist dies untenstehend in Abbildung 33 dargestellt. Allerdings wurden die Merkmalsausprägungen für das eingeschätzte Potential zusammengefasst. Abbildung 33 zeigt nun die spezifischen Techniken und Systeme gruppiert nach Unwissenheit, hohem oder sehr hohem-, neutralem-, oder niedrigen und sehr niedrigem Potential. Die Techniken des Kontinuierlichen Sink- und Steigflugverfahren weisen auch in dieser zusammengefassten Grafik das höchste Potential auf, gleichauf die Umweltmanagementsysteme. Ebenfalls ersichtlich ist, dass die Praktiken und Systeme der vierten Kategorie am schlechtesten abschneiden. Nichtsdestotrotz denken knapp 40% der Befragten, dass die Methodiken der Lärmüberwachung, -modellierung, und -kartierung ein hohes bis sehr hohes Potential aufweisen. Wiederum auffällig ist, dass das Noise Impact Reduction and Optimisation System (NIROS) mit über 40% eine sehr hohe Unwissenheitsrate aufweist. Auch die Kurvenanflugverfahren zeigen ein eher niedriges Potential um zur Reduzierung der negativen Umweltauswirkungen des Luftverkehrs in Bezug auf Luftqualität, Kraftstoffverbrauch sowie Lärmentwicklung beizutragen.

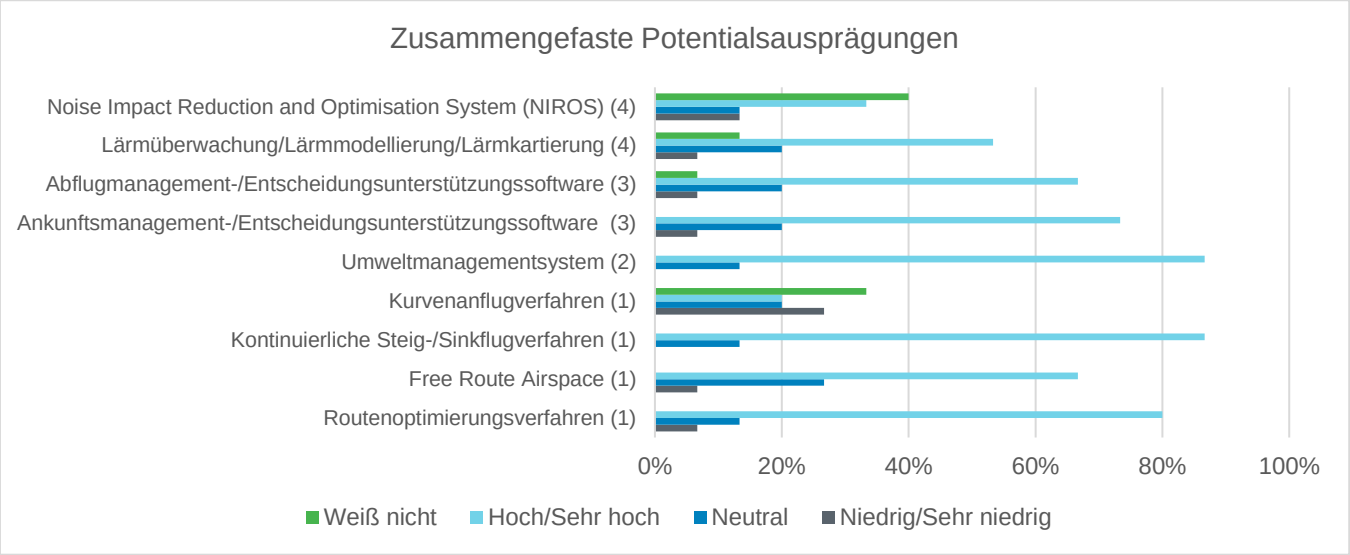


Abbildung 33: Zusammengefasste Potentiale der spezifischen Techniken

Quelle: Eigene Darstellung

Um auch beim Potential wiederum den Vergleich auf der Ebene der übergeordneten Kategorien anstellen zu können, ist folgend Abbildung 34 ersichtlich. Besonders die Techniken innerhalb der Kategorie Entscheidungshilfesysteme und automatisierte Luftverkehrsplanung werden mit einem hohen Potential eingestuft, wobei das Umweltmanagement mit Abstand die höchste Ausprägung für den Wert sehr hoch annimmt. Auch im Kategorienvergleich ist die Unterrepräsentation der vierten Kategorie auffällig. Gut zu sehen in Abbildung 34 ist der große Anteil an Unwissenheit über die darin enthaltenen Praktiken und Systeme. Dies lässt auf einen Bereich mit Potential schließen um die Bekanntheit zu steigern, da ca. 40% der teilnehmenden ANSPs, welchen diese Methoden bekannt waren, sie mit hohem oder sehr hohem Potential eingestuft haben.

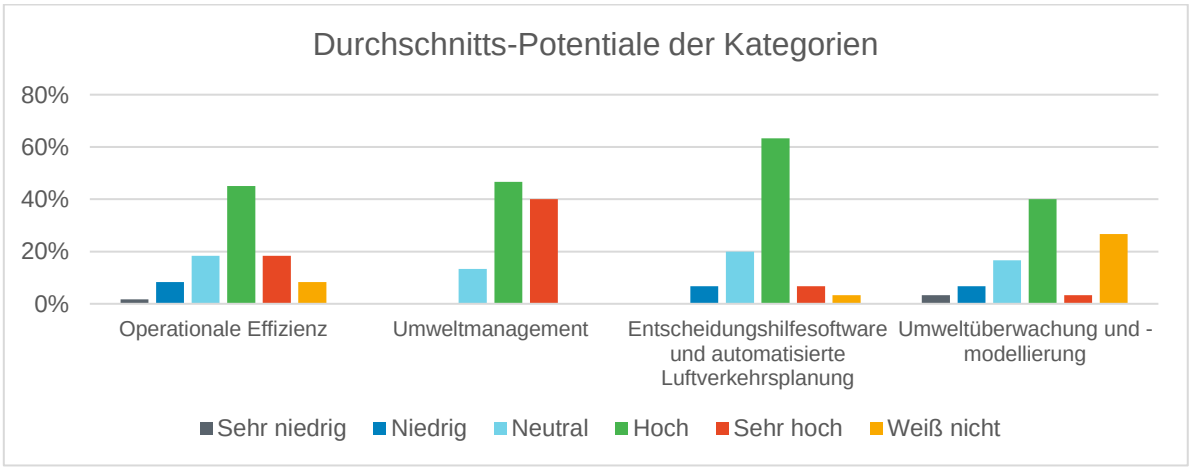


Abbildung 34: Durchschnittspotentiale pro Kategorie

Quelle: Eigene Darstellung

Der zweite Teil dieses Abschnitts im Fragebogen behandelte das Thema der Barrieren bei der Implementierung neuer Technologien oder Praktiken. Das Ergebnis dieses Abschnitts ist in Abbildung 35 zusammengefasst. Hier ist auf den ersten Blick ersichtlich, dass die Finanzierung, die Inkompatibilität mit dem derzeitigen System sowie diverse Regulierungen die größten Barrieren und Heraus-

forderungen darstellen. Auch die Koordination mit anderen ANSPs stellt mit 47% Zuspruch eine bedeutende Barriere bei der Implementierung neuer Methoden und Praktiken dar.

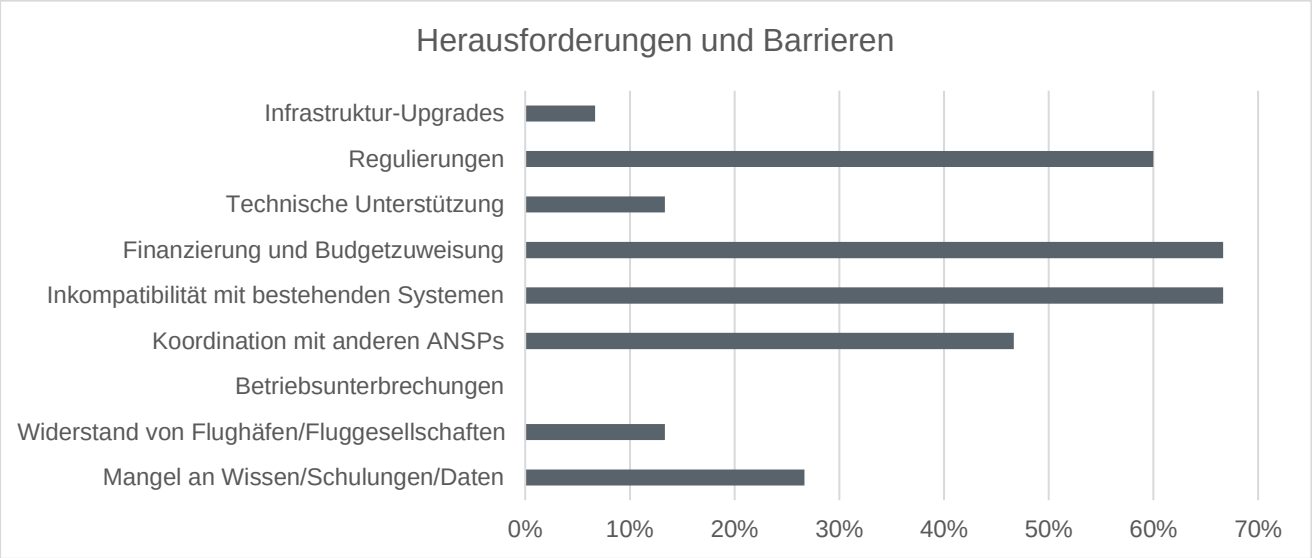


Abbildung 35: Herausforderungen und Barrieren

Quelle: Eigene Darstellung

7.3.3. Zukünftige Absichten

Die Sektion der zukünftigen Absichten zeigt ein ähnliches Bild wie zuvor jenes der weiteren oder zusätzlich angewandten Systeme und Praktiken, im Sinne von zukünftig geplanten Implementierungen. Jedoch kann festgehalten werden, dass der Tatendrang der europäischen ANSPs in einen grüneren und nachhaltigeren Luftverkehr hoch ist. 67% der befragten Organisationen haben angegeben, neue oder zusätzliche Praktiken und Systeme zur Reduzierung der Umweltauswirkungen in ihre Luftverkehrsmanagementaktivitäten in den nächsten 5 Jahren integrieren zu wollen, wie Abbildung 36 dargestellt.

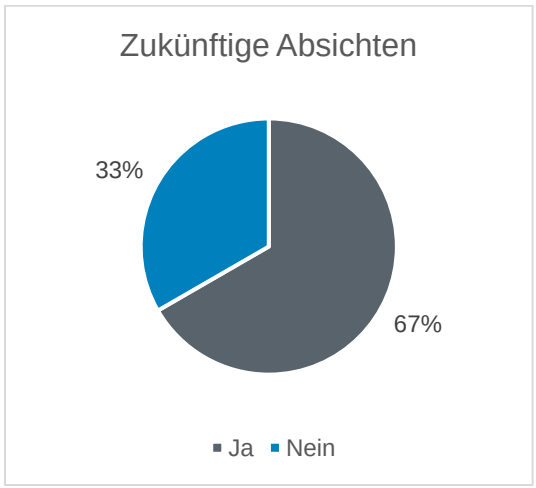


Abbildung 36: Zukünftige Absichten

Quelle: Eigene Darstellung

Für zukünftige Absichten stehen laut den teilnehmenden ANSPs vor allem neue Softwarelösungen und technologische Innovationen im Vordergrund. Die Teilnehmer des Fragebogens nennen konkre-

te Beispiele für neue Softwarelösungen und technologische Innovationen, wie die Integration von Machine-Learning-basierten Trajektorie-Vorhersagesystemen, effizientere Flugplanungssoftware und den Einsatz von KI-Algorithmen zur dynamischen Luftraumplanung. Softwarelösungen spielen laut den Ergebnissen des Fragebogens scheinbar eine Schlüsselrolle für zukünftige Methoden bei der Optimierung von Flugrouten, bei der Reduzierung des Kraftstoffverbrauchs und der Minimierung von Lärmemissionen, was letztendlich zu einer nachhaltigeren Luftfahrt beiträgt.

7.3.4. Zusammenhängende Analyse

In einem finalen Schritt der Analyse der Umfrageergebnisse werden einige Zusammenhänge zwischen den drei Sektoren des Fragebogens analysiert. Ein zu beleuchtender Aspekt ist dabei der Zusammenhang zwischen den Anwendungsraten der Methoden sowie dem eingeschätzten Potential. Auszugehen wäre von der Gegebenheit, dass Praktiken und Systeme mit hohem und sehr hohem Potential häufiger Anwendung finden als jene, mit neutralem, niedrigem oder sehr niedrigem Potential. Um diesen Zusammenhang visuell einschätzen zu können werden die Anwendungsraten mit eingeschätztem hohem bis sehr hohem Potential gegenübergestellt, wie in Abbildung 37 zu sehen ist.

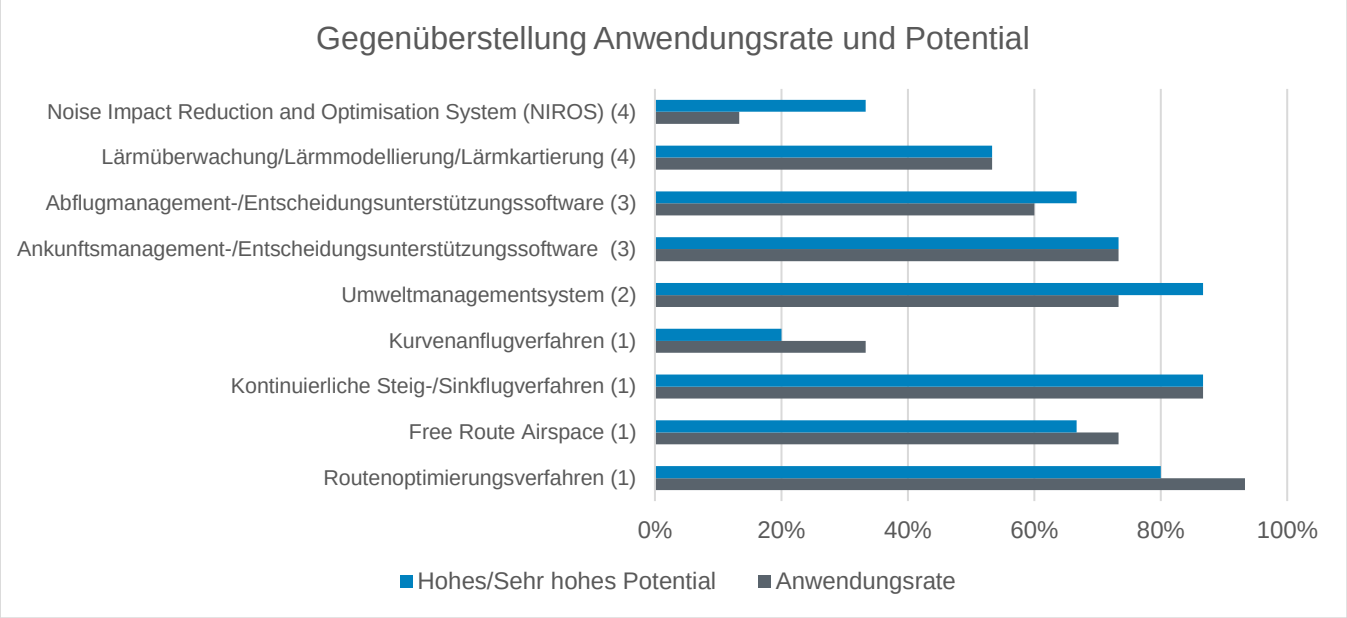


Abbildung 37: Gegenüberstellung Anwendungsraten mit hohem Potential

Quelle: Eigene Darstellung

Wie in Abbildung 37 ablesbar, herrscht tatsächlich ein starker Zusammenhang zwischen Methoden und Praktiken mit hohem bis sehr hohem Potential und deren Anwendungshäufigkeit in der Praxis. Dieses Ergebnis ist bedeutsam, da es ein Hinweis darauf ist, dass die angewandten Techniken nicht nur theoretisches Potenzial aufweisen, sondern auch tatsächlich in der Praxis von den ANSPs darauf basierend fokussiert werden. Diese konsistente Übereinstimmung zwischen potenziellen Möglichkeiten und tatsächlicher Anwendung veranschaulicht die Bedeutsamkeit der nachhaltigen Luftverkehrsindustrie für die europäischen ANSPs. Der Fokus liegt daher auf der Implementierung und Anwendung effektiver Methoden.

Ein weiterer interessanter Vergleich zwischen den Sektionen des Fragebogens ist das Potential mit den Herausforderungen und Barrieren. Vor allem auf Ebene der Kategorien bekommen die Systeme

zur Entscheidungshilfen bei An- und Abflug den weitaus höchsten Zuspruch für ein hohes Potential. Allerdings belegen diese Systeme auf Kategorie-Ebene im Sinne der Anwendungshäufigkeit nur den dritten Platz. Ein möglicher Grund dafür könnte die Herausforderung der Inkompatibilität mit den derzeitigen Systemen sein, welche jene Herausforderung mit dem höchsten Zuspruch ist. Dies könnte hierbei die erklärende Barriere darstellen, die dazu führt, dass die Kategorie mit dem höchsten Potential nicht die höchste Anwendungsrate widerspiegelt.

7.3.5. Fazit Umfrage

Zusammenfassend kann über die Ergebnisse der Umfrage folgendes festgehalten werden. Die Analyse der Fragebogenergebnisse offenbart Erkenntnisse über die Nutzung und Bewertung von Technologien und Systemen durch europäische Flugsicherungsorganisationen. Die zu bewertenden Techniken und Systeme sind in die genannten vier Hauptkategorien untergliedert: operationale Effizienz, Umweltmanagement, Entscheidungshilfesoftware sowie Luftverkehrsplanung und Umweltüberwachung und -modellierung. Es zeigt sich, dass Techniken und Systeme aus den ersten drei Kategorien entsprechend ihres eingeschätzten Potentials in der Praxis Anwendung finden. Im Gegensatz dazu weist die Kategorie der Umweltüberwachung und -modellierung sowohl geringere Anwendungs- als auch höhere Unwissenheitsraten auf, dies trotz einem im Vergleich hoch eingeschätzten Potential.

Die Motivation der ANSPs, nachhaltige Praktiken zu implementieren, ist vor allem durch das Streben nach Steigerung der operationalen Effizienz im Luftverkehr, Reduzierung des Kraftstoffverbrauchs und Verbesserung der Luftqualität geprägt. Die Lärmentwicklung hingegen wird weiter unten auf der Prioritätenliste angeführt. Diese Priorisierung spiegelt sich wie zuvor genannt in den aktuell verwendeten Technologien und Systemen wieder.

Bezüglich der Wirksamkeit der zu bewertenden Technologien und Praktiken zur Reduzierung der Umweltauswirkungen stufen die ANSPs insbesondere kontinuierliche Steig- und Sinkflugverfahren, als auch das Umweltmanagementsystem als hoch bis sehr hoch ein. Die Analyse identifiziert jedoch auch bedeutende Herausforderungen, wie die Finanzierung, Systeminkompatibilitäten und regulatorische Rahmenbedingungen, welche die Implementierung neuer Methoden behindern können.

Zusätzlich angewandte, über den zu bewertenden Systemen und Methoden hinaus, sowie auch zukünftigen Absichten der ANSPs weisen auf eine starke Tendenz hin, in den kommenden fünf Jahren innovative Praktiken und Systeme zu integrieren, um die Umweltauswirkungen weiter zu minimieren. Hierbei stehen vor allem fortschrittliche Softwarelösungen und technologische Innovationen im Vordergrund.

Zusammenhängende Analysen zwischen den diversen Sektionen der Umfrage haben zwei weitere Einsichten zum Vorschein gebracht. Zunächst ist, wie eventuell schon vorher annehmbar, eine starker Zusammenhang zwischen Anwendungsraten und eingeschätztem Potential zu sehen. Dies bedeutet, dass die ANSPs vermehrt Techniken und Systeme einsetzen, welche auch ein hohes bis sehr hohes Potential zur Reduzierung der negativen Umweltauswirkungen aufweisen. Darüber hinaus bekommt die Kategorie der Entscheidungshilfesoftware und Luftverkehrsplanung das größte Potential zur Reduzierung der Umwelteffekte zugesprochen. Im Sinne der Anwendungsraten belegt diese Kategorie allerdings nur den dritten Platz. Hier könnte ein Zusammenhang zwischen den An-

wendungsraten dieser Kategorie sowie den identifizierten Herausforderungen und Barrieren bestehen. Die Herausforderung der Inkompatibilität mit dem derzeitigen System hat im Fragebogen den ersten Platz belegt. Dies könnte eine Erklärung sein, warum die Systeme innerhalb der Kategorie Entscheidungshilfesoftware und Luftverkehrsplanung das höchste Potential aufweisen, allerdings bei den Anwendungsraten hinter dem Umweltmangement und der operationalen Effizienz liegen.

8. Vergleich Text-Mining-Resultate mit Umfrageergebnissen

Um in diesem Abschnitt die erwarteten sowie die tatsächlichen Ergebnisse und deren Zusammenhänge zu diskutieren, werden die erwarteten Ergebnisse, resultierend aus dem Web-Scraping und Text-Mining-Prozess, mit den tatsächlichen Resultaten aus dem Fragebogen verglichen. Dies wird wie folgt auf zwei unterschiedlichen Ebenen durchgeführt. Die erste Ebene bezieht sich auf eine generelle Analyse, welche den gesamten aktuellen Stand aus beiden Auswertungen vergleicht. Die Basis hierfür stellen die durch Text Mining identifizierten Methoden und Systeme dar. Die zweite Ebene beinhaltet den Vergleich der Ergebnisse auf individueller Organisationebene pro ANSP.

Basierend auf den individuellen Methodenlisten der ANSPs lässt sich ablesen, welche ANSPs welche spezifischen Techniken und Praktiken derzeit anwenden. Dies wurde also bereits im Vorhinein durch das Text Mining bestimmt, und im Anschluss mit den tatsächlichen Ergebnissen aus dem Fragebogen verglichen. Die erwarteten und die tatsächlichen Anwendungsraten der identifizierten Methoden und Praktiken sind in Abbildung 38 dargestellt. Auf den ersten Blick ist zu erkennen, dass die Anwendungsraten laut Text Mining für alle Methoden und Techniken niedriger sind, als diese per Fragebogen bestätigt wurden. Die mittlere Abweichung zwischen den erwarteten und tatsächlichen Resultaten beträgt ca. 18 Prozentpunkte.

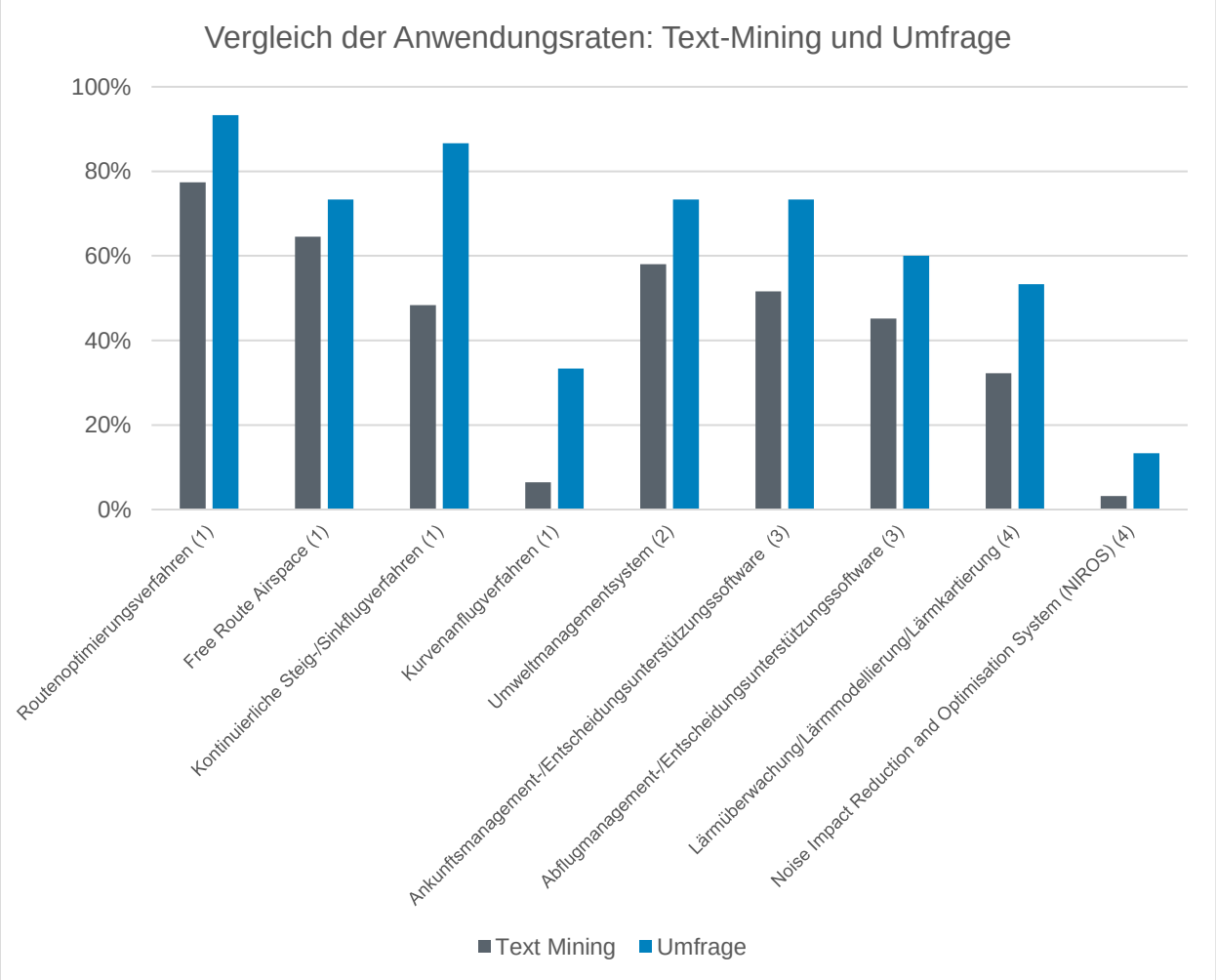


Abbildung 38: Vergleich der Anwendungsraten: Text-Mining und Umfrage

Quelle: Eigene Darstellung

Natürlich stellt sich nun als erstes die Frage, aus welchen Gründen diese Abweichungen in den Anwendungsraten vorgekommen sind. Hierfür werden drei Faktoren diskutiert, welche als mögliche Gründe genannt werden. Diese sind der manuelle Technik-Identifizierungsprozess, die Teilnehmerorganisationen an der Umfrage, sowie die Webseiteninformationen und -struktur. Es wird folgend zusammenfassend erläutert, aus welchen Gründen diese Faktoren eine Rolle spielen könnten.

Technik-Identifizierungsprozess: Wie bei der Durchführung des Text Minings beschrieben, wurden die identifizierten Techniken und Systeme, basierend auf der Keyword-Analyse, manuell aus den Textdaten extrahiert. Diese manuelle Identifikation aus den Textdaten birgt das Risiko, dass Techniken oder Systeme übersehen werden könnten. Dies kann einerseits durch Unaufmerksamkeit beim Identifizierungsprozess passiert sein, oder weil bestimmte Schlüsselwörter auf den Webseiten nicht verwendet wurden, um schlussendlich alle Techniken und Systeme zu identifizieren. Ob dies eine entscheidende Rolle gespielt hat ist in der generellen Analyse kaum bestimmbar. Nichtsdestotrotz sollte es hierfür Indizien in der folgenden individuellen Analyse geben.

Teilnehmerorganisationen der Umfrage: Es wurde festgestellt, dass jene ANSPs, die laut spezifischer Text-Mining-Ergebnisse im Durchschnitt viele Techniken und Systeme zur Reduzierung von Umweltschäden einsetzen, den Fragebogen ausgefüllt haben. Diese Schlussfolgerung basiert darauf, dass bei diesen ANSPs im Web-Scraping und Text-Mining-Prozess eine höhere Anzahl an identifizierten Techniken und Systemen vorliegt im Vergleich zum Durchschnitt aller ANSPs. Dies lässt natürlich darauf schließen, dass die Anwendungsraten in der Umfrage höher sind als jene des Text Minings, da hierbei alle ANSPs in Betracht gezogen worden sind. Diese Aussage kann natürlich nur für jene ANSPs bestätigt werden, welche auch den Namen bei der Auswertung des Fragebogens angegeben haben. Für jene kann aber auf jeden Fall festgehalten werden, dass diese im Schnitt laut den Text-Mining-Resultaten mehr Praktiken anwenden, als der Durchschnitt über alle analysierten ANSPs hinweg. Dies ist wiederum ein Indiz, dass die Studie vermehrt von jenen ANSPs beantwortet wurde, welche in der Thematik der Nachhaltigkeit und des grünen Luftverkehrs stärker engagiert sind. Dies stellt damit eine mögliche Erklärung für die höheren Anwendungsraten in den Umfrageergebnissen dar.

Webseiteninformationen und -struktur: Eine dritte mögliche Begründung für die gezeigten Diskrepanzen ist jene der Webseiteninformationen sowie der Webseitenstrukturen. Hier liegt das potentielle Risiko speziell im Web-Scraping-Prozess beziehungsweise in der Vorbereitung dessen. Hierbei könnten die Diskrepanzen aufgrund von zwei Faktoren aufgetreten sein. Ersterer ist, dass die relevanten Informationen auf der Webseite in Tags gespeichert sind, welche im Prozess nicht extrahiert wurden. Dies könnte die Folge einer unsorgfältigen Webseitenanalyse darstellen. Der zweite Faktor ist, dass die gesuchten Informationen gar nicht auf der Webseite des spezifischen ANSPs genannt wurden. Somit kann diese folgend auch nicht identifiziert werden und auch nicht in die Statistiken einfließen. Beide Faktoren haben also einen direkten Einfluss auf das Ergebnis und sind mögliche Gründe für die Unterschiede in den Anwendungsraten. Auch dieser Punkt lässt sich in der folgenden Individualanalyse besser erklären.

Um die drei eben identifizierten und erklärten Faktoren weiter zu untersuchen, wird nun die spezifische Analyse auf ANSP-Ebene angestellt. Dies kann für alle ANSPs angewendet werden, welche den Organisationsnamen bei der Beantwortung der Umfrage angegeben haben. Wie bereits bekannt,

besteht der Fragebogen aus drei Sektionen, welche darauf abzielen aktuell verwendete Technologien und Systeme zu identifizieren. Dies beinhaltete einerseits die durch Text Mining identifizierte Liste, zusätzlich angewendete Praktiken sowie zukünftige Pläne zur Integration neuer Methoden. Die Antworten auf diese Sektionen können nun mit den Text-Mining-Ergebnissen verglichen werden. Es wird also verglichen, wie viele von den im Fragebogen genannten Techniken und Systemen in jeder der drei Sektionen zuvor durch Text Mining identifiziert wurden. Eine Übersicht über einige ANSPs, welche die gängigsten Abweichungsgründe aufwiesen, sind in Tabelle 8 abzulesen. Eine Analyse über alle namentlich genannten ANSPs hinweg, zeigt eine mittlere Technik-Identifizierungsrate von über 80%.

Tabelle 8: Individuelle Technik-Identifizierungsraten

Land ANSP	Technik-Identifizierungsrate	Abweichungsgründe und Kommentare
United Kingdom	100,0%	■ Alle Techniken und Systeme identifiziert ■ möglicher Sprachenvorteil
Deutschland	80,0%	■ Fehlende Webseitenreferenzen ■ Abweichende Formulierungen
Spanien	66,67%	■ Fehlende Webseitenreferenzen
Albanien	80,0%	■ Fehlende Webseitenreferenzen
Tschechien	83,0%	■ Fehlende Webseitenreferenzen
Niederlande	72,73%	■ Fehlende Webseitenreferenzen ■ Abweichende Formulierungen

Wie in Tabelle 8 zu erkennen ist, war die größte Herausforderung beziehungsweise der Hauptgrund für nicht identifizierte Techniken und Systeme, das nicht Vorhandensein von Webseitenreferenzen. Einige wenige Techniken und Systeme waren darüber hinaus betroffen von der Verwendung abweichender Formulierungen, was zu fehlender Identifikation im Text Mining führte.

Um die Diskrepanzen in den Anwendungsraten festzustellen wurden drei mögliche Faktoren analysiert. Dabei handelte es sich wie bereits erwähnt um den manuellen Technik-Identifizierungsprozess, die Teilnehmerorganisationen der Umfrage, sowie die Webseiteninformationen und -struktur. Eine generelle Situationsanalyse sowie auch eine individuelle Analyse der spezifischen ANSPs hat folgendes ergeben. Die potentiellen Gründe für die Abweichungen in den erwarteten und tatsächlichen Ergebnissen lassen sich hauptsächlich durch die Teilnehmerorganisationen sowie das nicht Vorhandensein von Referenzen auf den Webseiten erklären. Die hohe Anzahl an umwelttechnisch engagierten ANSPs in den Teilnehmerorganisationen des Fragebogens ließen die Anwendungsraten der Umfrage in die Höhe steigen. Darüber hinaus hat die individuelle Analyse die fehlenden Webseiteninformationen als größte Herausforderung dargestellt. Ohne die entsprechenden Informationen auf den Webseiten, können diese im Rahmen des Web-Scrapings und Text Minings auch nicht identifiziert werden, und werden dadurch auch nicht in den erwarteten Anwendungsraten repräsentiert. Abschließend kann aber festgehalten werden, dass unter der Rücksichtnahme auf diese zwei Herausforderungen, die mittlere Abweichung mit nur ca. 18 Prozentpunkten überschaubar ist. Dies zeigt

wiederum das der angewandte Web-Scraping und Text-Mining-Prozess durchaus in der Lage war, den aktuellen Stand darzustellen. Dies natürlich alles unter Berücksichtigung der Limitationen, welche folgend erwähnt werden.

Unter den Hauptlimitationen können die Sprache, abweichende Formulierungen sowie auch die Webseitenstruktur genannt werden. Letzteres wurde zuvor bereits näher erklärt. Die abweichenden Formulierungen stellen dahingehend eine Limitation dar, da sie im Endeffekt dasselbe Resultat wie fehlende Referenzen und Inhalte auf der Webseite mit sich bringen. Wenn bestimmte Techniken oder Systeme von den unterschiedlichen ANSPs im Kontext unterschiedlicher Schlüsselwörter beschrieben werden, so ist es auch nicht möglich diese bei der Durchführung des angewandten Prozesses zu identifizieren. Hierfür müsste in erster Linie sorgfältig geklärt werden, unter welchen Namen die ANSPs bestimmte Systeme beschreiben und anwenden. Dies stellt natürlich einen großen zeitlichen Aufwand dar, welcher im Rahmen dieser Arbeit nicht tragbar ist. Eine weitere Limitation stellen die zahlreichen unterschiedlichen Sprachen im europäischen Raum dar. Durch das Verwenden der englischen Inhalte auf den Webseiten, war es des Öfteren zu sehen, dass die Anzahl an Paragraphen und Inhalten schrumpfte im Vergleich zu der Heimatsprache des entsprechenden Landes. Ein Indiz dafür ist das bereits in Tabelle 7 gezeigte ANSP aus Großbritannien. Hierbei musste die Webseite nicht in der englischen Version angezeigt werden, sondern in der Heimatsprache, was dazu führte, dass alle Techniken und Systeme im Text-Mining-Prozess identifiziert wurden. Hier könnte man sich für zukünftige Forschungen überlegen, die Inhalte in der Heimatsprache zu extrahieren und anschließend einen Übersetzer anzuwenden. Dies könnte den Informationsverlust aufgrund unterschiedlicher Anzeigesprachen der Webseiten verhindern. Die Performance beziehungsweise die Fähigkeit des Web-Scrapings und Text-Minings-Prozesses, den aktuellen Stand darzustellen, könnte dies dadurch weiter erhöhen.

9. Schluss

Im Prozess der vorliegenden Masterarbeit wurde das Web-Scraping und Text Mining dafür angewendet, um den aktuellen Stand darzustellen, welche Methoden, Initiativen oder Systeme europäische ANSPs derzeit anwenden, um den negativen Einfluss des Luftverkehrs auf die Umwelt zu reduzieren. Hier waren besonders Methoden, Systeme und Praktiken im Fokus, welche auf die Faktoren der Luftqualität, Lärmentwicklung sowie den Treibstoffverbrauch eingehen. Aufbauend auf diesen Ergebnissen wurde unter Rücksichtnahme der Expertenmeinung von Herrn Emir Ganić, PhD, der erwartete aktuelle Stand abgeleitet. Durch die Integration der mittels Text Mining generierten erwarteten Ergebnisse in einen Fragebogen, der anschließend an Vertreterinnen und Vertreter europäischer ANSPs versendet wurde, war es möglich, den tatsächlichen aktuellen Stand zu hinterfragen. Die Ergebnisse des Fragebogens haben eindeutig eine Unterrepräsentation der Techniken und Möglichkeiten im Kontext der Lärmvermeidung aufgezeigt. Dies obwohl diese Techniken und Systeme laut den Teilnehmerinnen und Teilnehmern ein durchaus hohes Potential zur Lärmvermeidung aufweisen. Nichtsdestotrotz wird diese Unterrepräsentation sowohl in den Anwendungsraten dargelegt, als auch durch die Motivationsfaktoren bestätigt. Auch bei der Motivation landet die Vermeidung der Lärmentwicklung auf dem letzten Platz. Die Entscheidungshilfesysteme belegen auf Ebene der Kategorien den vorletzten Platz in den Anwendungsraten bei sehr hoch eingeschätztem Potential. Eine mögliche Erklärung hierfür könnten die Implementierungsschwierigkeiten sein, welche die ANSPs dabei betreffen. Die größte Herausforderung der ANSPs bei der Implementierung von Techniken und Systemen zur Reduzierung der negativen Umweltauswirkungen ist die Inkompatibilität mit dem aktuellen System. Die restlichen beiden Kategorien, jene des Umweltmanagements, sowie der operationalen Effizienz zeigen sowohl hohe Anwendungsraten als auch ein hohes Potential zur Reduzierung der negativen Einflüsse. Weitere derzeit angewendete sowie auch zukünftige Intentionen der europäischen ANSPs bieten eine breit gefächerte Auflistung von Methoden und Systemen. Zusammenfassen lassen sich diese übergreifend unter den Kategorien neuartiger Software als auch nachhaltiger Initiativen. Auf Ebene der spezifischen Techniken und Systeme konnten hierbei keine klaren Muster abgeleitet werden.

Die erwarteten Ergebnisse, welche durch Text Mining und Web-Scraping identifiziert wurden, haben für alle Methoden und Techniken eine niedrigere Anwendungsrate prognostiziert, als durch die Auswertung des Fragebogens bestätigt. Im Schnitt liegt die Abweichung in den Anwendungsraten bei ca. 18 Prozentpunkten. Eine angestellte Analyse auf Ebene der spezifischen ANSPs hat herausgestellt, dass es zumeist durch fehlende Informationen auf den Webseiten zu diesen Abweichungen kommt. Ein weiterer Grund sind unterschiedliche Benennungen von Systemen oder Techniken der unterschiedlichen ANSPs. Für weitere Diskrepanzen in den Anwendungsraten verantwortlich ist, dass ANSPs, welche vermehrt im Umweltschutz engagiert sind und dadurch auch mehr Techniken und Systeme dafür anwenden, an der Umfrage teilgenommen haben. Dies lässt natürlich die praktischen Anwendungsraten in die Höhe klettern. Dies berücksichtigt kann allerdings sehr wohl festgehalten werden, dass der Prozess des Web-Scrapings und Text Minings sehr gut in der Lage war, die aktuelle Situation im Kontext derzeit angewandter Techniken und Systeme zu prognostizieren. Die Identifikationsrate aller Techniken und Systeme der spezifischen ANSPs liegt bei über 80%. Auch die Tatsache, dass bei der Auswertung des Fragebogens keine starken oder klaren Muster in den Sektionen der zusätzlich angewandten oder zukünftig anzuwendenden Techniken und Systemen identifi-

ziert wurden, ist ein Indiz dafür, dass die durch Text Mining erstellte Liste mit den aktuell angewandten Techniken, den tatsächlichen Stand widerspiegelt.

Nichtsdestotrotz unterliegt die vorliegende Analyse einiger Limitationen. Die stärkste davon ist wohlgemerkt jene, dass alle Webseiten in deren englischen Versionen verwendet wurden. Dies hat teilweise für einen Informationsverlust auf den Webseiten gesorgt, und stellt damit eine namhafte Limitation dar, welche darüber hinaus auch für die Diskrepanzen in den Anwendungsraten zur Verantwortung gezogen werden kann. Ein Informationsverlust stellt immer eine große Herausforderung für Text-Mining-Analysen dar. Diese Limitation birgt allerdings auch die Möglichkeit für zukünftige Forschungen. Das Scrapen der Webseiten in deren Heimatsprache sowie ein anschließendes Übersetzen der Inhalte, könnte durchaus die Performance beziehungsweise die Möglichkeit den aktuellen Stand zu prognostizieren, verbessern. Dadurch könnte sichergestellt werden, dass alle zur Verfügung stehenden Daten auf den Webseiten in die Analysen einfließen und somit das Risiko des Informationsverlust umgangen werden könnte.

10. Literaturverzeichnis

Allahyari, M., Pouriyeh, S., Assefi, M., Safaei, S., Trippe, E. D., Gutierrez, J. B., & Kochut, K. (2017). *A Brief Survey of Text Mining: Classification, Clustering and Extraction Techniques* (arXiv:1707.02919). arXiv. http://arxiv.org/abs/1707.02919

Atmosfair. (o. J.). *Klimawirkung des Flugverkehrs*. Klimawirkung des Flugverkehrs. Abgerufen 20. November 2023, von https://www.atmosfair.de/de/fliegen_und_klima/flugverkehr_und_klima/klimawirkung_flugverkehr/

Austro Control. (2019). *3 Bereiche der Flugsicherung*. https://www.austrocontrol.at/flugsicherung/fluglotsen/3_bereiche

Austro Control. (2019a). *Austro Control Umwelt-Pionier*. https://www.austrocontrol.at/unternehmen/profil/umwelt

BDL. (2023). *Klimaschutz im Luftverkehr*. Klimaschutz im Luftverkehr. https://www.bdl.aero/de/themen-positionen/nachhaltigkeit/klimaschutz/

BUND für Naturschutz und Umwelt Deutschland. (o. J.). *Luftverkehr: Klimaschädlich und hoch subventioniert*. Abgerufen 20. November 2023, von https://www.bund.net/themen/mobilitaet/infrastruktur/luftverkehr/

CANSO. (o. J.). *Guidlines on Airport-Collaborative Decision Making (A-CDM) Key Performance Measures*. Abgerufen 21. November 2023, von https://www.icao.int/SAM/Documents/2019-06901-SAMIG23/CANSO%20Guidelines%20on%20Airport-Collaborative%20Decision%20Making%20(A-CDM)%20Key%20Performance%20Measures.pdf

DFS. (2023). *Wie funktioniert Flugsicherung?* https://www.dfs.de/homepage/de/flugsicherung/betrieb/

DFS. (2023a). *Kontinuierlicher Sinkflug*. https://www.dfs.de/homepage/de/umwelt/fluglaerm/kontinuierlicher-sinkflug/

Diouf, R., Sarr, E. N., Sall, O., Birregah, B., Bousso, M., & Mbaye, S. N. (2019). Web Scraping: State-of-the-Art and Areas of Application. *2019 IEEE International Conference on Big Data (Big Data)*, 6040–6042. https://doi.org/10.1109/BigData47090.2019.9005594

EASA. (2022). *Europäischer Luftfahrt- und Umweltbericht 2022*. https://www.easa.europa.eu/eco/sites/default/files/2022-09/EnvironmentalReport_EASA_summary_DE_06.pdf

Eurocontrol. (2020, November 6). *Continuous climb and descent operations (CCO / CDO)*. https://www.eurocontrol.int/concept/continuous-climb-and-descent-operations

Eurocontrol. (2023, November 20). *Aviation sustainability*. https://www.eurocontrol.int/aviation-sustainability

Europäisches Parlament. (2019, Dezember 5). *CO_2-Emissionen des Luft- und Schiffsverkehrs: Zahlen und Fakten*. https://www.europarl.europa.eu/news/de/headlines/society/20191129STO67756/co2-emissionen-des-luft-und-schiffsverkehrs-zahlen-und-fakten-infografik

Hippner, H., & Rentzmann, R. (2006). Text Mining. *Informatik-Spektrum*, *29*(4), 287–290. https://doi.org/10.1007/s00287-006-0091-y

Holubliev, V., & Simishko, V. (2021). Web Scraping and Text Mining of Ukranian News Articles About Ecology. *11th International Conference on Advanced Computer Information Technologies (ACIT)*, 672–675. https://doi.org/10.1109/ACIT52158.2021.9548450

Jarošová, M., & Pajdlhauser, M. (2022). Aviation and Climate Change. *Transportation Research Procedia*, *65*, 216–221. https://doi.org/10.1016/j.trpro.2022.11.025

Khder, M. (2021). Web Scraping or Web Crawling: State of Art, Techniques, Approaches and Application. *International Journal of Advances in Soft Computing and Its Applications*, *13*(3), 145–168. https://doi.org/10.15849/IJASCA.211128.11

Lee, J., & Lee, M.-J. (2018). Measuring Contribution of Spatial Information to Environmental Research Using Text Mining Techniques. *IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium*, 5289–5291. https://doi.org/10.1109/IGARSS.2018.8517444

LimeSurvey. (o. J.). *LimeSurvey—Free Online Survey Tool*. Abgerufen 13. April 2024, von https://www.limesurvey.org/de

Mine, D., & Mine, C.-R. (2021). Web Scraping in the Statistics and Data Science Curriculum: Challenges and Opportunities. *JOURNAL OF STATISTICS AND DATA SCIENCE EDUCATION*, *1*(29), 112–122.

Modapothala, J. R., & Issac, B. (2009). Study of economic, environmental and social factors in sustainability reports using text mining and Bayesian analysis. *2009 IEEE Symposium on Industrial Electronics & Applications*, 209–214. https://doi.org/10.1109/ISIEA.2009.5356467

RDocumentation. (o. J.). *Rvest package*. Abgerufen 20. Dezember 2023, von https://www.rdocumentation.org/packages/rvest/versions/0.3.6

Rybka, J. (o. J.). *HTML-Grundgerüst: Aufbau einer HTML-Seite*. Abgerufen 27. November 2023, von https://blog.hubspot.de/website/html-grundgeruest

Skyguide. (2023). *Kerngeschäft*. https://www.skyguide.ch/de/unternehmen/, https://www.skyguide.ch/de/unternehmen/kerngeschaft

Talib, R., Kashif, M., Ayesha, S., & Fatima, F. (2016). Text Mining: Techniques, Applications and Issues. *International Journal of Advanced Computer Science and Applications*, *7*(11). https://doi.org/10.14569/IJACSA.2016.071153

Wickham, H. (o. J.). *SelectorGadget*. Abgerufen 27. November 2023, von https://rvest.tidyverse.org/articles/selectorgadget.html

Wuttke, L. (2022). *Text Mining: Definition, Methoden und Anwendung*. https://datasolut.com/wiki/text-mining/

Zhao, B. (2017). *Web Scraping* (S. 1–3). https://doi.org/10.1007/978-3-319-32001-4_483-1

11. Anhang

11.1. Fragebogen

Section 1 - Current Situation

1 Please state your company's name:

_____

2 Please indicate for all the listed practices and systems below whether your organisation is currently using them for environmental reasons (Yes), not using them (No), or if you do now know about the usage of this particular practice in your company (Don't know).

Practice or System	Yes	No	Don't know
Operational Efficiency			
Route Optimisation Procedures	☐	☐	☐
Free Route Airspace	☐	☐	☐
Continuous Descent/Climb Operations	☐	☐	☐
Curved Approaches	☐	☐	☐
Regulatory and Management Frameworks	☐	☐	☐
Environmental Management System (EMS)	☐	☐	☐
Decision Support Software and Air Traffic Planning	☐	☐	☐
Arrival Management/Decision Support Software	☐	☐	☐
Departure Management/Decision Support Software	☐	☐	☐
Environmental Monitoring and Modelling	☐	☐	☐
Noise Monitoring/Mapping/Modelling	☐	☐	☐
Noise Impact Reduction and Optimisation System	☐	☐	☐

3 Are there any other systems, techniques, or practices currently used in your company aimed at reducing the environmental impact of air traffic in the focused areas(air quality, fuel consumption, or noise emissions) that have not been mentioned before?

□ YES □ NO

IF YES

3.1 Please provide details on the additional systems, techniques, or practices currently used in your company. Include names, descriptions, functionalities, environmental goals or any other relevant information.

_____

4 Are you currently using any other decision support systems or automated air traffic planning tools to enhance the environmental impact reduction efforts in any of the focused areas (air pollution, fuel consumption, noise emissions) that have not been already mentioned?

□ YES □ NO

IF YES

4.1 Please provide details on the decision support systems or automated air traffic planning tools that your organisation is currently using. Include names, descriptions, functionality, environmental goals, or any other relevant information.

_____

IF question 2 at least one YES and/or question 3 YES and/or question 4 YES – THEN:

5 In your organisations pursuit of reducing the environmental impact of air traffic operations, what specific goals or desired outcomes are you actively focusing on? Please select all that apply:

□ Reduce Noise Emissions □ Reduce Fuel Consumption

□ Enhance Air Quality □ Enhance Operational Efficiency of Air Traffic

□ None Specific

□ Other(please specify): _____

Section 2 - Possibilities

6 Please assess the potential of the following practices, techniques and systems to reduce the environmental impact of air traffic operations in terms of air quality, fuel consumption and noise emissions. Rate the perceived usefulness or promise of each named system, technique or practice from very high to very low.

PRACTICE OR SYSTEM	Very High	High	Neutral	Low	Very Low	Don't Know
Operational Efficiency	☐	☐	☐	☐	☐	☐
Route Optimisation Procedures	☐	☐	☐	☐	☐	☐
Free Route Airspace	☐	☐	☐	☐	☐	☐
Continuous Descent/Climb Operations	☐	☐	☐	☐	☐	☐
Curved Approaches	☐	☐	☐	☐	☐	☐
Regulatory and Management Frameworks	☐	☐	☐	☐	☐	☐
Environmental Management System (EMS)	☐	☐	☐	☐	☐	☐
Decision Support Software and Air Traffic Planning	☐	☐	☐	☐	☐	☐
Arrival Management/Decision Support Software	☐	☐	☐	☐	☐	☐
Departure Management/Decision Support Software	☐	☐	☐	☐	☐	☐
Environmental Monitoring and Modelling	☐	☐	☐	☐	☐	☐
Noise Monitoring/Mapping/Modelling	☐	☐	☐	☐	☐	☐
Noise Impact Reduction and Optimisation System	☐	☐	☐	☐	☐	☐

7 What challenges or barriers do you foresee in adopting new technologies like decision support systems and automated air traffic planning tools for environmental impact reduction?

□ Lack of Knowledge/Training or Data
□ Airport/Airline Opposition
□ Operational Disruption
□ Coordination with other ANSPs
□ Incompatibility with Current Systems
□ Funding or Budget Allocation
□ Technical Support
□ Regulatory Frameworks
□ Infrastructure Upgrades
□ None
□ Others, please explain _____

Section 3 – Future Intensions

8 Do you have intentions to integrate new or additional environmental impact reduction practices or systems into your air traffic management activities within the next 5 years?

□ YES □ NO

IF YES

8.1 Please provide details on the new environmental impact reduction practices your company plans to integrate in the near future. Include names, descriptions, functionality, environmental goals, or any other relevant information.

11.2. Link zum Online-Repository

Folgender Link ermöglicht den Zugang zum Online-Repository. Unter diesem Link sind die verwendeten Codestücke, weitere Ergebnisse sowie deutsche Übersetzungen der Umfrage und deren Ergebnisse zu finden.

LINK: https://figshare.com/s/a74c1d661eb0aec6d239