

# **Masterarbeit**

**Florian Hartart**

**Algorithmische Erkennung von Gefahrenlagen  
durch Interpretation urbaner Daten**

**Florian Hartart**

Algorithmische Erkennung von Gefahrenlagen  
durch Interpretation urbaner Daten

Masterarbeit eingereicht im Rahmen der Masterprüfung

im Studiengang Next Media  
am Department Informatik  
der Fakultät Technik und Informatik  
der Hochschule für Angewandte Wissenschaften Hamburg

Betreuender Prüfer: Prof. Dr. Kai von Luck  
Zweitgutachter: Prof. Dr. Marc Schütte

Abgegeben am 4.9.2018

**Florian Hartart**

**Thema der Masterarbeit**

Algorithmische Erkennung von Gefahrenlagen durch Interpretation urbaner Daten

**Stichworte**

Anomalieerkennung, Bevölkerungsschutz, Big Data, Digitalisierung, Feuerwehr, Gefahrenabwehr, Gefahrenlagen, Hamburg, Künstliche Neuronale Netzwerke, Maschinelles Lernen, Open Data, Polizei, Sensorik-Leitstelle, Smart City, Soziale Medien, urbane Daten

**Kurzzusammenfassung**

Die vorliegende Masterthesis thematisiert die Erkennung von Gefahrenlagen im städtischen Raum durch algorithmische Auswertung urbaner Daten aus sozialen Netzwerken, Sensornetzwerken und Open Data-Portalen. Am Beispiel der Stadt Hamburg wird gezeigt, dass urbane Daten zwar vorhanden sind, ihre potentiellen Möglichkeiten und Inhalte jedoch zum jetzigen Zeitpunkt nicht in Echtzeit repräsentiert werden können. Wesentliches Merkmal der Datenanalyse ist die Integration des Knowledge Discovery in Databases (KDD-)Prozesses in die Netzwerke einer Stadt. Zur Auswertung sind verschiedene Methoden der künstlichen Intelligenz und des maschinellen Lernens zielführend; es werden verschiedene Auswertungsalgorithmen vorgestellt. Zur Einschätzung der Realisierbarkeit einer Echtzeiterfassung wird das Text Mining als eine Spezifikation des Data Mining fokussiert. Die Interaktion mit den gewonnenen Informationen erfolgt in spezialisierten Sensorik-Leitstellen. Die gewonnenen Erkenntnisse bilden für Wissenschaft und Praxis einen theoretischen Mehrwert und regen weitere Forschung an.

**Florian Hartart**

**Title of the paper**

Algorithmic recognition of dangerous situations through the interpretation of urban data

**Keywords**

Anomaly detection, artificial neural networks, big data, civil protection, digitalization, awareness system, emergency, fire department, Hamburg, hazard control, law enforcement agency, machine learning, open data, sensor control center, smart city, social media, urban data

**Abstract**

This master's thesis addresses the recognition of dangerous situations in urban areas through the algorithmic evaluation of urban data collected from social networks, sensor networks and Open Data portals. As an example, the city of Hamburg shows that whilst there is indeed urban data available for the city, at present potential opportunities and contents cannot be represented in real time. An essential feature of the data analysis is the integration of the Knowledge Discovery in Databases (KDD) process into the networks of a city. Different methods of artificial intelligence and machine learning are conducive to the evaluation; various evaluation algorithms are presented. For the assessment of the feasibility of a real-time connection, text mining is focused on as a specification of data mining. The interaction with the information obtained is done in specialised sensor control centres. The knowledge obtained constitutes a theoretical added value for science and practice and encourages further research.

# Inhaltsverzeichnis

<b>1</b>	<b>Einleitung .....</b>	<b>7</b>
1.1	Aufbau der Arbeit.....	12
<b>2</b>	<b>Smart Cities: Urbane Daten für die Gefahrenabwehr .....</b>	<b>16</b>
2.1	Gefahrenabwehr als Aufgabe von Smart Cities .....	17
2.2	Urbane Daten sind Big Data .....	20
2.3	Urbane Daten repräsentieren Städte in Echtzeit.....	25
2.4	Sensorik-Leitstellen als digitale Lagezentren .....	26
<b>3</b>	<b>Daten: Repräsentanten der realen Welt.....</b>	<b>28</b>
3.1	Datentypen.....	28
3.1.1	Strukturierte Daten .....	30
3.1.2	Unstrukturierte Daten.....	30
3.1.3	Semistrukturierte Daten .....	31
3.2	Datenökonomie am Beispiel der Stadt Hamburg .....	31
<b>4</b>	<b>Sensorik-Leitstellen: Anforderungen und Konzeption .....</b>	<b>40</b>
4.1	Grundsätzliche Anforderungen .....	41
4.2	Technische Konzeption .....	42
4.2.1	Sensing layer – Die Suche nach Informationen.....	43
4.2.2	Data layer – Die Speicherung von Daten.....	45
4.2.3	Logical layer – Die Anwendung von Algorithmen .....	45
4.2.4	Knowledge Discovery in Databases (KDD-Prozess).....	47
4.2.5	Presentation layer – Die Visualisierung von urbanen Daten .....	49

<b>5</b>	<b>Algorithmen: Methoden zur Auswertung .....</b>	<b>51</b>
5.1	Künstliche neuronale Netzwerke .....	53
5.2	Deep learning .....	57
5.3	Beispielhafte künstliche neuronale Netzwerke .....	58
5.4	Maschinelles Lernen .....	59
5.5	Data Mining .....	60
5.6	Text Mining .....	62
5.6.1	Auswahl der Textdaten .....	64
5.6.2	Vorbereitung der Textdaten .....	64
5.6.3	Transformation der Textdaten .....	69
5.6.4	Anwendungsbeispiel .....	72
5.7	Web Mining .....	76
<b>6</b>	<b>Visualisierung: Technische Möglichkeiten .....</b>	<b>78</b>
6.1	Interaktiver Gefahrenabwehr-Leitstand .....	78
6.2	Digitales Stadtmodell .....	82
<b>7</b>	<b>Diskussion .....</b>	<b>86</b>
<b>8</b>	<b>Fazit .....</b>	<b>95</b>
	<b>Abkürzungsverzeichnis .....</b>	<b>99</b>
	<b>Abbildungsverzeichnis .....</b>	<b>100</b>
	<b>Tabellenverzeichnis .....</b>	<b>102</b>
	<b>Literaturverzeichnis .....</b>	<b>103</b>

# 1 Einleitung

Die Digitalisierung betrifft unterschiedlichste Lebensbereiche, so auch die Gefahrenabwehr. Zunehmend werden moderne Informationstechnologien zur Erkennung und Abwehr von Gefahren eingesetzt. Eine Kernaussage der 64. Jahresfachtagung der Vereinigung zur Förderung des Deutschen Brandschutzes e. V. (vfdb) ist, dass die Digitalisierung ein Treibstoff zur Verbesserung der zivilen Sicherheit sein kann [1].

Durch den Einsatz von mobilen Endgeräten (Smartphones) werden die Prozesse der Notfallversorgung optimiert. So können freiwillige Ersthelfer<sup>1</sup> bei einem präklinischen Herz-Kreislaufstillstand zur Hilfe gerufen werden, um frühzeitige Reanimationsmaßnahmen durchzuführen. Erste Systeme dieser Art wurden in einigen deutschen Landkreisen und Städten eingeführt [2]. (Smartphone-) Applikationen werden zur Warnung der Bevölkerung vor Gefahrenlagen eingesetzt. Eine bekannte Applikation (App) ist die Warn-App „NINA“ des Bundesamtes für Bevölkerungsschutz und Katastrophenhilfe [3].

---

<sup>1</sup> Im Folgenden wird aus Gründen der besseren Lesbarkeit ausschließlich die männliche Form benutzt. Es sind stets männliche und weibliche Personen gleichermaßen gemeint.

Soziale Netzwerke sind mittlerweile ein probates Mittel zur Kommunikation in Katastrophenszenarien geworden. Neben der Kommunikation werden beispielsweise Hilfsersuchen und freiwillige Helfer koordiniert. Das Forschungsprojekt „Emergent“ untersucht die positiven und negativen Einflüsse von sozialen Netzwerken in Katastrophenszenarien [4].

Eine weitere Möglichkeit ist die Erkennung von Gefahrenlagen durch eine gezielte Auswertung der sozialen Netzwerke. Die New Yorker Gefahrenabwehrbehörde (City's Office of Emergency Management) wertet verschiedene soziale bzw. digitale Medien zur Abwehr von Gefahren aus [5]. Zudem setzt die New Yorker Polizei ein sogenanntes „Domain Awareness System“ zur Erkennung von Gefahrenlagen ein. Im Wesentlichen besteht dieses System aus unterschiedlichen Sensoren. Über Mikrofone (Audio-Sensoren) können beispielsweise Waffenschüsse oder über chemische Sensoren Gefahrgutunfälle detektiert werden [6]. Im Grunde basieren die vorgenannten Systeme auf einer Erkennung von Anomalien in großen Datenbeständen, die ihrerseits zur Gefahrenabwehr beitragen.

Moderne Städte entwickeln sich heutzutage hin zu Smart Cities [7,8]. Zahlreiche Aspekte des realen Lebens werden zunehmend mit Daten abgebildet und von technischen Systemen erfasst. Beispiele sind die Umwandlung von beruflichen Verbindungen durch soziale Geschäftsnetzwerke oder von Echtzeitinformationen aus sozialen Netzwerken in Daten. Cukier und Mayer-Schoenberger [S. 29] [9] beschreiben die zunehmende Datenerfassung mit dem Begriff „Datafizierung“. Zukünftige städtische Entwicklungen basieren auf der Datafizierung des realen Lebens oder werden von diesen Daten vorangetrieben [7,8].



Wesentlich sind sogenannte offene Daten, die von den Bürgern frei verwendet werden dürfen. Die vielfältigen Daten einer Stadt können als urbane Daten beschrieben werden. Urbane Daten werden von technischen Sensoren, sozialen Sensoren und der öffentlichen Verwaltung bereitgestellt. Technische Messeinrichtungen und Bürger können als Sensoren ihre Umwelt aufzeichnen. Technische Sensoren sind vorher programmierte Messeinheiten und erfassen die Umwelt als Messwerte. Bürger geben ihre subjektive Wahrnehmung der Umwelt als Beiträge in sozialen Netzwerken wieder [10–12].

Mit urbanen Daten ist es möglich, Städte in Echtzeit zu erfassen. Mit dem Begriff „Real-Time City“ werden in der Literatur methodische und technische Ansätze zur Erfassung, Auswertung und Visualisierung urbaner Daten beschrieben [13]. Vielversprechende Methoden zur Auswertung sind künstliche neuronale Netzwerke [14–16]. Die Stadt Rio de Janeiro hat ein System entwickelt, mit dem Datensätze von dreißig unterschiedlichen Fachbehörden mit den Daten von unterschiedlichen Akteuren und Sensoren einer Stadt kombiniert werden [17].

Für Behörden und Organisationen der Gefahrenabwehr bietet sich im Umgang mit großen Datenbeständen die Möglichkeit zur Erkennung von Gefahrenlagen [12]. Gefahrenlagen sind Abweichungen (Anomalien) von den bekannten Datenströmen [14]. Algorithmen sind in der Lage, die Datenströme zu erfassen, einen Grundwert zu ermitteln und Anomalien zu erkennen. Bisher mussten die Leitstellen der Gefahrenabwehrbehörden (v. a. Feuerwehr und Polizei) Informationen zur aktuellen Lage einer Gefahrenlage von meldenden Bürgern, Einsatzkräften oder Dritten anfordern. Die Auswertung urbaner Daten ermöglicht die frühzeitige Identifikation von Gefahrenlagen; mit der (u.a. bildlichen) Darstellung von Informationen wird eine

Verbesserung der Situationsaufmerksamkeit und indem Einsatzmöglichkeiten erreicht.

Erste Arbeiten zeigen, dass mit Beiträgen im Microblog-Dienst Twitter Erdbebenszenarien erkannt werden können [18].

Die Stadt Hamburg hat als erstes Bundesland mit dem 2012 eingeführten Hamburgischen Transparenzgesetz eine Vorreiterrolle im Bereich der offenen Daten eingenommen. Die Behörden stellen den Bürgern abfrageunabhängig städtische Daten online zur Verfügung. Zur Bereitstellung der Daten wird von der Stadt Hamburg seit Februar 2013 ein eigenes Transparenzportal betrieben. Zu den angebotenen hamburgischen Datenbeständen zählen amtliche Statistiken, Gutachten, Studien und Geodaten [19]. Anfang 2015 wurde ein Strategiepapier „Digitale Stadt“ als Wegweiser zu einer digitalen sowie smarten Stadt veröffentlicht. Als ein konkretes Beispiel für die Digitalisierung wird die Daseinsvorsorge im Hinblick auf die öffentliche Infrastruktur und deren Transportmittel beschrieben [20].

Die Gefahrenabwehr, als Teil der öffentlichen Daseinsvorsorge, ist bisher nicht explizit genannt. Auch die bisherigen Projekte (Digitale Verwaltung, intelligente Verkehrssysteme, smartPORT, Hamburg Open Online University, Intelligente Bildungsnetze, Smarte Geodaten, Smart Energy, eCulture) legen bisher keinen direkten Fokus auf die Gefahrenabwehr [20].

Mit den ersten gesetzlichen Grundlagen hin zu einer digitalen Stadt und der damit verbundenen offenen Datenkultur bietet Hamburg in besonderer Weise

---

Möglichkeiten an, um prototypisch urbane Daten zur Erkennung von Gefahrenlagen auszuwerten und so die Sicherheit der Bevölkerung zu erhöhen.

Abschließend kann festgehalten werden, dass Daten ein wertvoller Rohstoff des 21. Jahrhunderts sind und die Erzeugung von neuem Wissen und Entwicklung von innovativen Dienstleistungen vorantreiben. Die wesentliche Herausforderung im Hinblick auf urbane Daten besteht darin, die zur jeweiligen Fragestellung passende Erhebung, Auswertung und Analyse sowie Interpretation und Visualisierung zu generieren. Die zukünftige Vermittlung entsprechender Ergebnisse an die Gefahrenabwehrbehörden unterstützt die Aufrechterhaltung der öffentlichen Sicherheit und Ordnung.

Die vorliegende Masterthesis zeigt neue Wege zur Erkennung von Gefahrenlagen im städtischen Raum auf. Sie will Prozesse zur Nutzung und Bearbeitung urbaner Daten anregen und darüber hinaus die Einbindung der Daten der Bürger einer Stadt thematisieren.

Eine smarte Anwendung der Gefahrenabwehr (smartGAb)- hier als Begriff vom Autor explizit eingeführt - hat das Potential die bisherigen Bestrebungen der Freien und Hansestadt Hamburg richtungsweisend zu erweitern und zum Bevölkerungsschutz beizutragen.

## 1.1 Aufbau der Arbeit

Eine smarte Anwendung der Gefahrenabwehr basiert auf der Hypothese, dass Gefahrenlagen im urbanen Raum durch die algorithmische Auswertung von Daten erkannt werden können. Im Kern dieser Arbeit werden dazu *urbane Daten* als grundlegende Wissensbasis eingeführt und mögliche Algorithmen der künstlichen Intelligenz zur automatisierten Auswertung beschrieben. Zur Operationalisierung der Daten werden das *Konzept einer Sensorik-Leitstelle* und technische Mindestanforderungen erörtert. Gleichzeitig wird weitere Forschung angeregt.

Gegliedert ist die vorliegende Arbeit in **acht Kapitel**, die im folgenden Abschnitt beschrieben werden.

Zum Einstieg in die Thematik werden im **zweiten Kapitel** Smart Cities als urbane Systeme vorgestellt, die auf technischen Entwicklungen der Informations- und Kommunikationstechnologie begründet sind. Ein Grundverständnis über die stetig wachsenden Netzwerke zwischen Bürgern und der öffentlichen Verwaltung ist notwendig, um den Mechanismus von partizipativen Entscheidungen und der kollaborativen Weiterentwicklung von städtischen Prozessen verstehen zu können. Zudem erklärt sich mit der zunehmenden Vernetzung ein immenser Datenbedarf. Hierzu werden *urbane Daten* als Spezifikation von Big Data eingeführt. Urbane Daten können zur Repräsentation der Echtzeitstadt eingesetzt werden. Als ein Anwendungsszenario wird das Erkennen von Gefahrenlagen, im Sinne von Anomalien in Datenströmen, durch sogenannte Sensorik-Leitstellen beschrieben.

Im **dritten Kapitel** werden unterschiedliche Quellen urbaner Daten thematisiert. Zum allgemeinen Verständnis werden Daten als eine Option zur Erfassung von physikalischen sowie nicht-physikalischen Entitäten beschrieben. Unterschieden werden strukturierte sowie semi- und unstrukturierte Daten. Eine Grundkenntnis über die unterschiedlichen Datentypen ist erforderlich, da jeder Datentyp eigene Anforderungen an die technischen Konzepte zur Auswertung stellt. Am Beispiel der Stadt Hamburg werden anschließend ausgewählte Datensätze aufgeführt, um die Realisierbarkeit einer Echtzeitauswertung evaluieren zu können.

Im **vierten Kapitel** werden die wesentlichen Anforderungen an das **‘Konzept einer Sensorik-Leitstelle’** im Hinblick auf die Mensch-Computer-Interaktion beschrieben. Hierzu wird für den technischen Aufbau ein vierschichtiges Softwarearchitekturmodell vorgestellt. Das Kernstück des Systems ist die logische Schicht. Als ein wesentlicher Handlungsablauf wird der Knowledge Discovery in Databases (KDD)-Prozess vorgestellt. Eine Besonderheit urbaner Daten ist, dass Wissen erst in der Interaktion mit den Daten entsteht und ein iteratives Durchlaufen des KDD-Prozesses erfordert.

Das **fünfte Kapitel** beschreibt den multimethodischen Ansatz zur Auswertung urbaner Daten. Im Wesentlichen werden verschiedene Techniken künstlicher Intelligenz thematisiert. Das Fokussieren auf die automatische Erkennung von Anomalien in Datenströmen -die als Repräsentanten von Gefahrenlagen gelten- erfordert die Konzentration auf künstliche neuronale Netzwerke und Konzepte des maschinellen Lernens. Ausgehend von dem Ansatz, dass Mitteilungen der Bevölkerung in sozialen Medien ein erster Hinweis auf eine Gefahrenlage sein können, wird zur Auswertung von Textdaten das Text Mining vertieft.

Die aus den urbanen Daten gewonnenen Informationen müssen visualisiert werden, damit sie durch Vertreter der Gefahrenabwehrbehörden beurteilt werden können. Im **sechsten Kapitel** wird das '*Konzept einer Sensorik-Leitstelle*' im Sinne einer multimedialen Plattform für die Gefahrenabwehr beschrieben. Die Darstellung von urbanen Daten erfordert geeignetere Visualisierungstechniken als die meist verwendeten Geoinformationssysteme. Digitale Stadtmodelle, wie das City Scope des MIT Media Labs, werden ebenso wie ein interaktiver Leitstand in den methodischen Kontext urbaner Datennutzung gestellt.

Im **siebten Kapitel** werden die Erkenntnisse der vorliegenden Arbeit diskutiert. Zu Beginn wird die Charakterisierung von urbanen Daten thematisiert. Aufgegriffen wird unter anderem die Unterscheidung von urbanen Daten in eine soziale und eine konventionelle Komponente. Zudem wird das Konzept der „fünf V“ als ein mögliches Modell zur Beschreibung von urbanen Daten diskutiert.

Als Ergebnis dieser Masterthesis werden richtungsweisende Charaktermerkmale urbaner Daten zur wissenschaftlichen und praktischen Verwendung im Bereich der Gefahrenabwehr vorgestellt. Anknüpfend an die gewonnenen Erkenntnisse wird die Datenökonomie der Stadt Hamburg mit den wesentlichen Merkmalen urbaner Daten

- Verfügbarkeit,
- Relevanz und
- Aktualisierungsintervall

verglichen, weil mit der „Verfügbarkeit“ die wesentlichen Erklärungsmodelle von Big Data (wie das fünf V-Modell) erweitert und mit der „Relevanz“ sowie dem „Aktualisierungsintervall“ zusammengefasst werden können.

---

Zum Ende der Diskussion werden lexikonbasierte Methoden und Verfahren des maschinellen Lernens als Verfahren zur Bewertung der Semantik und Aussagekraft von Textdaten herangezogen.

Zum Abschluss der Masterthesis werden mit dem Fazit im **achten Kapitel** die wesentlichen Aussagen und Ergebnisse zusammengefasst und offene Forschungsaspekte aufgezeigt.

## **2 Smart Cities: Urbane Daten für die Gefahrenabwehr**

Die zunehmende Verwendung von modernen Informations- und Kommunikationstechnologien hat die Weiterentwicklung von Städten hin zu Smart Cities maßgeblich beeinflusst. Die Bildung von digitalen Netzwerken und der zunehmende Datenaustausch bedingen die Entwicklung von Smart Cities. Qualitativ hochwertige Daten sind ein treibender Faktor bei der Weiterentwicklung urbaner Räume. Von intelligenten Verkehrsleitsystemen, über intelligente Mülltonnen bis hin zu Verwaltungsdienstleistungen, tragen Daten zu Innovationen der städtischen Aspekte bei.

In Kapitel 2.1 wird zunächst die Gefahrenabwehr als ein Aspekt einer Smart City thematisiert. Zur Verdeutlichung der Datenvielfalt in Städten werden in Kapitel 2.2 urbane Daten im Kontext der Big Data beschrieben. Kapitel 2.3 zeigt auf, dass mit urbanen Daten Städte in Echtzeit erfasst werden können. Abschließend wird in Kapitel 2.4 am Beispiel von Sensorik-Leitstellen die Möglichkeit zur Erkennung von Gefahrenlagen aufgezeigt.



## **2.1 Gefahrenabwehr als Aufgabe von Smart Cities**

Städte sind große künstliche Gebilde, die von Menschen geschaffen wurden, um die Lebensqualität einer Vielzahl von Menschen zu verbessern. Diese urbanen Systeme bauen neben ökonomischen und ökologischen Entwicklungen, sozialen und menschlichen Interaktionen auf technologische Konzepte [8].

Insbesondere die Omnipräsenz von technischen Endgeräten und die Verwendung von modernen Informations- und Kommunikationstechnologien sowie das veränderte Medienverhalten der Bürger haben maßgeblich zu einer Digitalisierung der Städte beigetragen.

Ein wesentlicher Effekt ist die Bildung von hochkomplexen Netzen, die den Austausch zwischen den unterschiedlichen Stakeholdern einer Stadt fördern. Stakeholder sind unterschiedliche Anspruchsgruppen einer Stadt. Hierzu zählen u. a. die Bürger, Vertreter der Politik, Fachbehörden, Betreiber Kritischer Infrastrukturen, Polizei, Feuerwehr, Rettungsdienstorganisationen. Einerseits wird auf diese Weise die partizipative Beteiligung der Bürger an den Prozessen der Stadt, andererseits der Wissensaustausch unter den Gruppen gefördert [21].

Städte mit einer hochgradigen digitalen Vernetzung zwischen Bürgern und staatlichen Akteuren werden als Smart Cities bezeichnet. Kitchin [13] beschreibt, dass der Begriff „Smart City“ im Wesentlichen durch zwei Merkmale charakterisiert wird:

Einerseits können Smart Cities mit dem ständigen Vorhandensein von Computern (Ubiquitous Computing) und digitalen Sensoren (z. B. kabellose Kommunikationsnetze, Kameras, digitalüberwachte Infrastrukturen) beschrieben werden. Die Erfassung der städtischen Prozesse erfolgt nahezu in Echtzeit und wird durch die zunehmende Popularität von mobilen Endgeräten (z. B. Smartphones) technisch erleichtert [13].

Andererseits wird eine Smart City durch die Bürger der Stadt geprägt. Basierend auf smarten Bürgern werden partizipativ Entscheidungen in der Politik getroffen oder kollaborativ die Prozesse der Stadt weiterentwickelt [13]. Ebenso können Städte mit smarten Bürger auf Disruptionen smart reagieren [22,23]. Zu den Möglichkeiten der Stakeholder einer Smart City kann auch das Reagieren auf Gefahrenlagen gezählt werden.

Die Grundgedanken einer Smart City -die Verwendung einer Netzwerk-Infrastruktur und die Kollaboration mit den Bürgern- wirken sich nahezu in allen Lebensbereichen einer Stadt aus. Giffinger et al. [S. 10-12] [22] beschreiben sechs Dimensionen einer Smart City, die zahlreiche urbane Lebensbereiche umfassen. Neben den Menschen (Smart People) werden als weitere Dimensionen Smart Economy (als Merkmal der Wettbewerbsfähigkeit), Smart Governance (als Merkmal der Partizipation), Smart Mobility (als Merkmal für die Mobilität), Smart Environment (als Merkmal der Umwelt) und Smart Living (als Merkmal für die Lebensqualität) beschrieben [23]. Abbildung 1 stellt die sechs Dimensionen und dazugehörige Handlungsfelder einer Smart City dar.

<b>Smart Economy (Competitiveness)</b> <ul style="list-style-type: none"> <li>• Innovative spirit</li> <li>• Entrepreneurship</li> <li>• Economic image &amp; trademarks</li> <li>• Productivity</li> <li>• Flexibility of labour market</li> <li>• International embeddedness</li> <li>• <i>Ability to transform</i></li> </ul>	<b>Smart People (Social &amp; Human Capital)</b> <ul style="list-style-type: none"> <li>• Level of qualification</li> <li>• Affinity to life long learning</li> <li>• Social and ethnic plurality</li> <li>• Flexibility</li> <li>• Creativity</li> <li>• Cosmopolitanism/Open-mindedness</li> <li>• Participation in public life</li> </ul>	<b>Smart Governance (Participation)</b> <ul style="list-style-type: none"> <li>• Participation in decision-making</li> <li>• Public and social services</li> <li>• Transparent governance</li> <li>• <i>Political strategies &amp; perspectives</i></li> </ul>
<b>Smart Mobility (Transport and ICT)</b> <ul style="list-style-type: none"> <li>• Local accessibility</li> <li>• (Inter-)national accessibility</li> <li>• Availability of ICT-infrastructure</li> <li>• Sustainable, innovative and safe transport system</li> </ul>	<b>Smart Environment (Natural resources)</b> <ul style="list-style-type: none"> <li>• Attractivity of natural conditions</li> <li>• Pollution</li> <li>• Environmental protection</li> <li>• Sustainable resource management</li> </ul>	<b>Smart Living (Quality of life)</b> <ul style="list-style-type: none"> <li>• Cultural facilities</li> <li>• Health conditions</li> <li>• Individual safety</li> <li>• Housing quality</li> <li>• Education facilities</li> <li>• Touristic attractivity</li> <li>• Social cohesion</li> </ul>

Abbildung 1: Sechs Dimensionen und dazugehörige Handlungsfelder einer Smart City nach Giffinger et al. [22]

Mit der Dimension Smart Living wird die Sicherheit der Bürger (individual safety) als ein Handlungsfeld einer Smart City beschrieben. Die Gewährleistung der öffentlichen Sicherheit und Ordnung ist eine staatliche Aufgabe. Entsprechende Behörden und Organisationen leisten im staatlichen Auftrag Sicherheitsaufgaben der polizeilichen und nicht-polizeilichen Gefahrenabwehr. Sie können im Hinblick auf die Gefahrenabwehr zu unmittelbaren Profiteuren der Nutzung urbaner Daten werden.

Das folgende Kapitel 2.2 greift die beschriebenen sechs Dimensionen auf und verdeutlicht darüber hinaus das qualitative und quantitative Potential urbaner Daten.

## **2.2 Urbane Daten sind Big Data**

Gemeinsamer Treiber der technologiebasierten Veränderungen in Smart Cities sind urbane Daten. Generiert werden die digitalen Daten in den sechs Dimensionen einer Smart City (Abbildung 1) [8]. Die unterschiedlichen Handlungsfelder der sechs Dimensionen lassen die Aussage zu, dass smarte Daten in vernetzten urbanen Zusammenhängen mit verschiedener Qualität und Quantität erzeugt werden.

Erfasst werden die Daten über ein verteiltes Sensornetzwerk, bestehend aus sozialen und technischen Sensoren. Soziale Sensoren sind die Bürger, die beispielsweise über Smartphones, Fitness-Armbänder oder Pulsmessuhren fortlaufend Vitaldaten (Quantified Self) erfassen oder über soziale Medien über Geschehnisse in der Stadt (implizites Wissen) berichten [8,10]. Goodchild [S. 25-27] [10] beschreibt den hohen Informationsgehalt der Bürger als kollektive Intelligenz, welche über sechs Billionen Bürger weltweit verteilt ist. Sagl et al. [S. 60-62] [24] veranschaulicht dieses Wissen mit einer Auswertung des Nutzungsverhaltens von Touristen auf der Fotoplattform „Flickr“. Basierend auf den hochgeladenen Fotodaten konnten interessante Orte einer Stadt identifiziert werden [24]. In Abbildung 2 werden die für Touristen interessanten Orte am Beispiel New York als Hotspots abgebildet.

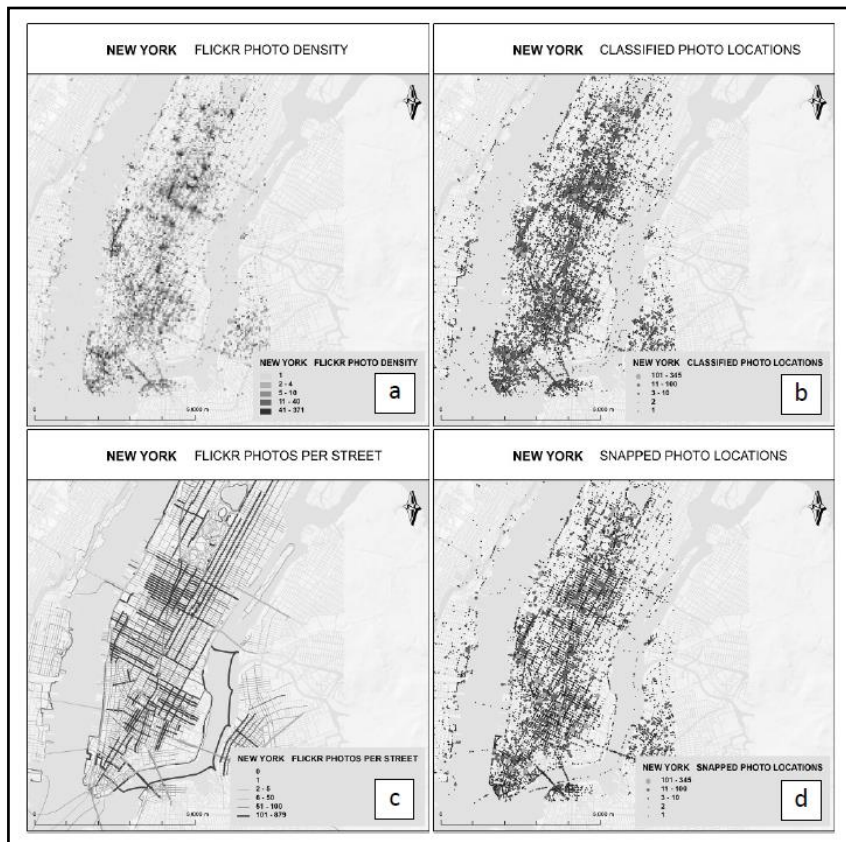


Abbildung 2: Nutzungsverhalten der Fotoplattform "Flickr" durch Touristen [24]

Zudem verdeutlicht Abbildung 2, dass die „Augen“ der sozialen Sensoren über die gesamte Stadt verteilt sind. Aufgrund dieser Feststellung wird eine Realität greifbar, in der neben Sehenswürdigkeiten ebenso Gefahrenlagen erfasst werden. Xu et al. [12] beschreiben das Potential der Bürger zur Erkennung von Gefahrenlagen als eine Möglichkeit des Crowdsourcing<sup>2</sup>. Weiterhin unterteilen sie die sozialen Sensoren je nach der zu beobachtenden Gefahrenlage. Wird beispielsweise durch einen Bürger ein Feuer erfasst, handelt es sich um einen Feuer-Sensor. Ausgehend von diesem

<sup>2</sup> Crowdsourcing bezeichnet die Auslagerung von Aufgaben an eine Gruppe von Internetnutzern.

Konzept können soziale Sensoren als spezialisierte Sensoren („Everyone as a Sensor“) betrachtet werden [12].

Nicht nur soziale, sondern auch technische Netzwerke zeichnen eine Smart City aus [25]. Jin et al. [S. 113-114] [26] beschreiben die Etablierung eines technischen Sensornetzwerkes mit Technologien des `Internet of Things` (IoT)<sup>3</sup>. Mit den Sensoren können die Lärmbelästigung, die Luft- und Wasserqualität, das Nutzungsverhalten des öffentlichen Nahverkehrs oder die Energienutzung erfasst werden [26]. Data Science Communities (wie das Projekt „[luftdaten.info](http://luftdaten.info)“) veranschaulichen, dass die IoT-Technologien aktiv durch die Bürger eingesetzt werden [27]. Jin et al. [S. 113] [26] nennen als mögliches Einsatzgebiet für technische Sensoren die Förderung der Sicherheit in Städten (city safety). Nicht nur die Bürger setzen Sensoren zur Erfassung von Umweltparametern ein. Die öffentliche Verwaltung hat in Städten ebenfalls umfassende Sensornetzwerke installiert. Beispielsweise werden die Luftqualität in Städten<sup>4</sup> oder die Strahlenbelastung<sup>5</sup> (Gamma-Ortsdosisleistung) in Deutschland erfasst [28,29].

---

<sup>3</sup> Mit dem Sammelbegriff Internet of Things (zu Deutsch Internet der Dinge) werden Technologien beschrieben, die eine Vernetzung von physischen Gegenständen durch Informations- und Kommunikationstechniken ermöglicht.

<sup>4</sup> <http://luft.hamburg.de/>

<sup>5</sup> <https://odlinfo.bfs.de/DE/index.html>

Sensordaten sind nicht die einzigen Daten mit denen die Prozesse einer Stadt erfasst werden können. Weitere Datenquellen sind staatliche Institutionen, die offene Daten zur weiteren Nutzung anbieten. Offene Daten (Open Data) umfassen sämtliche Datenbestände, die den Bürgern frei zugänglich sind und wiederverwendet werden dürfen [30]. Im Sinne der offenen Verwaltungsdaten (Open Government Data) sind alle Daten und Informationen inbegriffen, die nicht als geheim eingestuft werden und dessen Veröffentlichung im Interesse der Öffentlichkeit ist [31]. Hierunter fallen Statistiken, Geodaten, Umweltdaten oder Verkehrsinformationen [32]. Abbildung 3 veranschaulicht die Zusammensetzung der urbanen Daten.

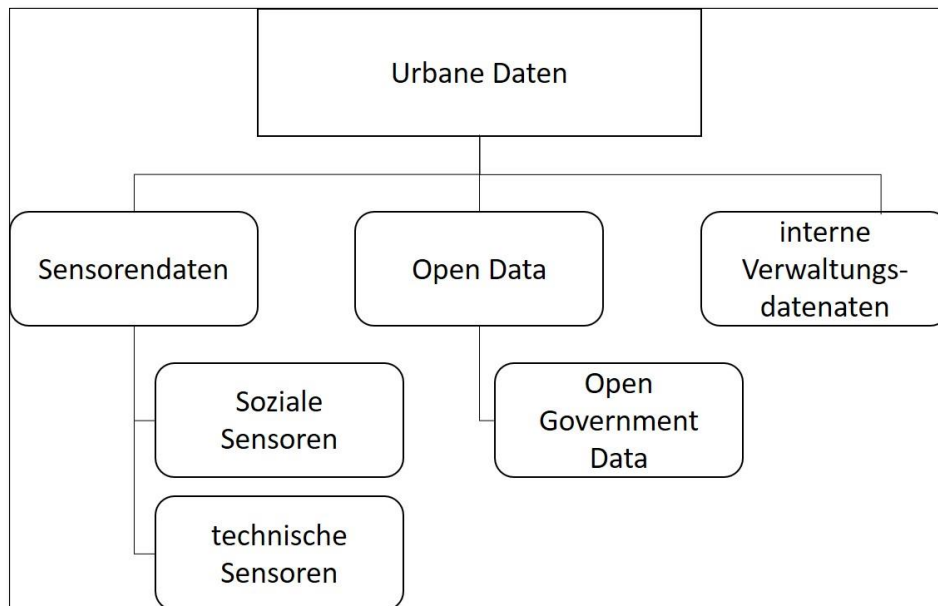


Abbildung 3: Urbane Daten - Datenquellen in Smart Cities

Urbane Daten sind vielfältige, aus unterschiedlichen Quellen zusammengesetzte Datenbestände, die in kurzen Intervallen erzeugt werden. Auf diese Weise entstehen quantitativ und qualitativ bemerkenswerte Datensätze für nahezu alle städtischen

Lebensbereiche wie Konsum (Digitalisierung des Einzelhandels, Amazon), Verkehr (Mobilverbände, Uber), ökonomische Leistungsfähigkeit (geobasierte Kreditvergabe, PayPal) oder Videodaten (Überwachungskameras, Facebook, Google Streetview). Laney [33] hat zur Beschreibung dieser Datenbestände den Begriff Big Data geprägt.

Big Data werden als high-volume- (Menge), high-velocity- (Geschwindigkeit), high-variety- (Vielfältigkeit) Informationen bezeichnet [33]. Neben den drei grundlegenden „V’s“ sind zusätzlich zwei weitere Charakteristika zu nennen. Mit der Veracity, als eine Beschreibung der Wahrhaftigkeit und Glaubwürdigkeit von Daten, wird die Qualität der gesammelten Daten bewertet. Mit der zunehmenden Menge an Daten besteht die Gefahr, dass die Datenqualität und die Aussagekraft gemindert werden. Die Aussagekraft der Daten wird mit dem fünften Charakteristikum der Value ergänzt und erfasst den Umstand, dass der Informationsgehalt verglichen mit der Datenmenge eher gering ist [34].

Die Datenbestände werden in immer kürzeren Zeitabständen generiert und verdoppeln sich in weniger als zwei Jahren. Hinzukommt, dass die Updatezyklen zunehmend kürzer werden. Das Daten nahezu in Echtzeit generiert werden und die Prozesse einer Stadt abbilden können wird in Kapitel 2.3 vertieft. Mit konventionellen Methoden und Algorithmen können die Datenmengen nicht ausgewertet werden [34]. Es sind neue Konventionen und Fähigkeiten notwendig, um mit technischen Rahmenbedingungen (vgl. Kapitel 4) und Strategien (vgl. Kapitel 5) ein angepasstes Datenmanagement zu etablieren [35].



### **2.3 Urbane Daten repräsentieren Städte in Echtzeit**

Immer mehr Städte setzen urbane Daten nutzbringend ein. Beispielsweise werden Verkehrsdaten zur Optimierung der Ampelschaltungen eingesetzt oder die Polizei setzt Überwachungskameras zur Beobachtung von öffentlichen Räumen ein. Nach Kitchin [S. 5] [13] sind es Insellösungen und keine integrierten Ansätze mit der Einbindung verschiedener Fachbehörden sowie der Bevölkerung.

Ein hervorzuhebendes Beispiel für den behördenübergreifenden Einsatz von urbanen Daten ist das „Centro De Operacoes Prefeitura Do Rio“<sup>6</sup> der Stadt Rio de Janeiro in Brasilien. In einer Kooperation zwischen der Regierung und der Firma IBM, wurde ein System entwickelt, mit dem Daten von dreißig Behörden durch aktive Informationen von Mitarbeitern und der Bevölkerung ergänzt werden. Die Verarbeitung der Daten erfolgt einerseits automatisiert mit Algorithmen, andererseits durch Analysten. Basierend auf diesen Informationen können zukünftige Entwicklungen in der Stadt vorausgesagt oder potentielle Risiken, beispielsweise Hochwassersituationen, analysiert werden [13].

Ein weiteres Praxisbeispiel ist die Kooperation zwischen der New Yorker Polizei und der Firma Microsoft. In New York wurde ein sogenanntes „Domain Awareness System“ zur Auswertung von sicherheitsrelevanten Daten entwickelt. Mit diesem System können beispielsweise Videoüberwachungskameras, Notrufmeldungen, bereits erfasste Einsatzberichte und weitere Daten in Echtzeit ausgewertet werden [36]. Die Echtzeitinformationen werden u. a. genutzt, um Waffenschüsse zu detektieren und unmittelbar Einsatzkräfte an diesen Ort zu entsenden [6].

---

<sup>6</sup> Weiterführende Informationen: <https://www.nytimes.com/2012/03/04/business/ibm-takes-smarter-cities-concept-to-rio-de-janeiro.html>

Diese zwei Praxisbeispiele verdeutlichen, dass die Echtzeitauswertung von urbanen Daten zur Darstellung urbaner Prozesse kein utopischer Ansatz ist. Vielmehr entstehen weitere Möglichkeiten der Nutzung digitaler Daten im Zusammenwirken mit der Bevölkerung und entsprechender Technologien. Als eine Möglichkeit wird im folgenden Kapitel 2.4 eine Sensorik-Leitstelle als digitales Lagezentrum für die Gefahrenabwehr beschrieben.

## **2.4 Sensorik-Leitstellen als digitale Lagezentren**

Gefahrenlagen sind hochkomplexe und unvorhersehbare Situationen, die die vielfältigen Prozesse einer Stadt (z. B. Infrastruktur, soziales Miteinander, Wirtschaft) negativ beeinflussen und die Bevölkerung gefährden. In kürzester Zeit müssen von den Verantwortlichen der Gefahrenabwehr weitreichende Entscheidungen getroffen werden, um die öffentliche Sicherheit und Ordnung wiederherzustellen. Bevor die Organisationen der Gefahrenabwehr tätig werden können, muss eine Gefahrenlage erkannt werden. Die Sensoren einer Stadt können zur Erkennung von Gefahrenlagen hilfreiche Daten liefern.

Lu und Brelsford [37] haben die Daten von Twitter-Nutzern vor und nach dem Tsunami im Jahr 2011 in Japan ausgewertet. Mit ihrer Auswertung konnten sie zeigen, dass die Nutzung von sozialen Medien während Gefahrenlagen zunimmt [37]. Dies lässt die Möglichkeit zu, dass Daten aus sozialen Medien zum frühzeitigen Erkennen von Gefahrenlagen eingesetzt werden können. Bestärkt wird diese Annahme durch eine Untersuchung von de Longueville et al. [S. 75-79] [18], die zeigen konnten, dass mit Twitter-Meldungen ein Waldbrand in Frankreich räumlich erfasst werden konnte. Cassa et al. [38] haben die Twitter-Beiträge zum Anschlag auf

den Boston Marathon 2013 ausgewertet und konnten zeigen, dass mit Beiträgen in den sozialen Netzwerken Notfallorte in kürzester Zeit identifiziert werden konnten.

Ausgehend von den Praxisbeispielen in Rio de Janeiro (Brasilien) und New York (USA) wird in dieser Arbeit ein konzeptioneller Systementwurf einer Sensorik-Leitstelle für Behörden und Organisationen der Gefahrenabwehr zur Erkennung von Gefahrenlagen auf Grundlage einer algorithmischen Auswertung von urbanen Daten beschrieben (vgl. Kapitel 5 und 6). Wesentliches Systemmerkmal ist die Integration des Knowledge Discovery in Databases (KDD-)Prozesses in die Netzwerke einer Stadt (vgl. Kapitel 4.2.3). Zur Auswertung der urbanen Daten werden verschiedene Methoden der künstlichen Intelligenz eingesetzt, um frühzeitig Gefahrenlagen in der Stadt zu erkennen (vgl. Kapitel 5).

In Kapitel 3 werden die unterschiedlichen Datentypen und beispielhafte Datensätze ausführlich vorgestellt. Das Verständnis über unterschiedliche Daten bildet die entscheidende Grundlage für die Gestaltung technischer Konzepte zur Auswertung.

## 3 Daten: Repräsentanten der realen Welt

Auf dem Weg zu einer Echtzeit-Stadt müssen fortlaufend unterschiedliche Datenquellen erfasst werden. Mit den gewonnenen Daten können sowohl die physikalischen Elemente (z. B. Objekte, Personen, Fahrzeuge), als auch nicht-physikalische Elemente (z. B. Prozesse) beschrieben werden. Die Heterogenität der Datenquellen spiegelt sich in der Vielfältigkeit der Datentypen wieder.

Zum Verständnis der Datenvielfalt werden in Kapitel 3.1 die unterschiedlichen Datentypen beschrieben. Am Beispiel der Stadt Hamburg werden in Kapitel 3.2 exemplarische Datensätze aufgeführt.

### 3.1 Datentypen

Urbane Daten sind eine Ansammlung von Beobachtungen, Messungen oder statistischen Erhebungen, die Informationen mit einem räumlich-zeitlichen Bezug zur Stadt beinhalten (vgl. Kapitel 2.2) [39,40]. In der Abstraktion entstehen so Entitäten, mit denen die reale Welt abgebildet werden kann (vgl. Kapitel 2.3) [41].

Daten mit einem räumlichen Kontext sind Geodaten. Eine Unterscheidung erfolgt in Geobasis- und Geofachdaten. Geobasisdaten geben Auskunft zu den allgemeinen topographischen Gegebenheiten. Sie bilden die Referenzschicht für eine Visualisierung von räumlichen Daten. Standardisierte, flächendeckende und fortlaufend aktualisierte Geobasisdaten werden von staatlichen Institutionen<sup>7</sup> bereitgestellt. Ergänzt werden Geobasisdaten mit (Geo-)Fachdaten, die von einer fachlichen und räumlichen Komponente charakterisiert werden [42]. Letztendlich können Fachdaten mit urbanen Daten verglichen werden.

Wie in Kapitel 2.2 beschrieben, können Fachdaten über soziale und technische Sensoren, aus Open Data Beständen oder aus internen Quellen gewonnen werden. Die Daten liegen in strukturierter, semistrukturierter und unstrukturierter Form vor [39].

**Tabelle 1: Beispiele für strukturierte, semi- und unstrukturierte Daten [43]**

<b>Strukturierte Daten</b>	<b>Semistrukturierte Daten</b>	<b>Unstrukturierte Daten</b>
Daten aus Datenbanken	Social Media Schnittstellen (API)	Texte
Statistische Daten	Google/Bing Suchmaschinen Schnittstellen (API)	Bücher
Geodaten (Volunteered Geographic Information)	RSS Feeds	Worddokumente
Daten von Public Science Communities	Webseiten, Blogs, Online-Foren	Mediendateien (Bild-, Video-, Audiodateien)
Offene Verwaltungsdaten (Open Government Data)		Überwachungskameras
Offene Daten (Open Data)		Satelliten Bilddateien
		Deep/Dark Web Inhalte

<sup>7</sup> Beispielsweise sei hier das Geoportal des Bundes [geoportal.de](http://geoportal.de) genannt.

Die drei Strukturierungsgrade

- strukturierte,
- semi- und
- unstrukturierte Daten (siehe Tabelle 1),

werden in den folgenden drei Unterkapiteln (3.1.1 bis 3.1.3) näher beschrieben.

### **3.1.1 Strukturierte Daten**

Strukturierte Daten haben eine feste Struktur (z. B. mit einer festen Reihenfolge oder eindeutigen Attributen). In relationalen Datenbanken können diese Daten mit einer eindeutigen Bezeichnung dargestellt werden [39]. Insbesondere Daten aus Open Data-Portalen werden in einem für Tabellenkalkulationen gängigen Datenformat angeboten und können von Maschinen gelesen werden. Daten von technischen Sensoren werden ebenfalls in Datenbanken strukturiert. Beispielsweise werden durch das Open Data-Portal der Stadt Hamburg Daten des Luftmessnetzes in einem CSV-Dateiformat mit Kommaseparation (comma-separated values) angeboten [44]. Ein weiteres Anwendungsbeispiel ist die Darstellung von Geodaten in einem GIS. Beispielsweise können mehrere Informationen zur Bewertung eines Hochwasserrisikos (z. B. Einwohnerzahlen, Bevölkerungsdichten, Geländehöhen, Hochwasserrisikogebiete) für ein Bezugsgebiet in verschiedenen Kartenebenen übereinandergelegt werden [45].

### **3.1.2 Unstrukturierte Daten**

Mediale Daten wie Bilder, Videos, Audioaufnahmen und Texte werden als unstrukturierte Daten klassifiziert. Vor einer Bearbeitung durch Maschinen müssen die Datenbestände zunächst strukturiert werden [39]. Für eine Darstellung der Echtzeit-Stadt sind unstrukturierte Daten von besonderer Bedeutung. Mitteilungen der Bürger, als textueller oder visueller Beitrag, kommunizieren implizites Wissen.

Das Sprachrohr für die Bürger sind die sozialen Netzwerke wie Twitter oder Facebook. Xu et al. [12] haben mit ihrer Untersuchung gezeigt, dass aus Datenbeständen sozialer Netzwerke in Form von textuellen Beiträgen ein Brandgeschehen und mit Fotoaufnahmen eine Geiselnahme detektiert werden können. In einer weiteren Untersuchung beschreiben Xu et al. [46], dass Videoüberwachungssysteme in Kombination mit Beiträgen aus einem Microblog-Dienst eine Möglichkeit zu Früherkennung von Gefahrenlagen sind.

### **3.1.3 Semistrukturierte Daten**

Zu den semistrukturierten Daten zählen beispielsweise Texte von Webseiten, Blogs oder Online-Foren [39]. Im Grunde zählen Texte ebenso zu den unstrukturierten Daten. Durch die Nutzung der Hypertext-Auszeichnungssprache (Hypertext Markup Language (HTML)) werden diese Informationen strukturiert und als semistrukturiert beschrieben. Diese Form der Daten ist für den Einsatz in einer Sensorik-Leitstelle ebenfalls geeignet. Eine Arbeit von Xu et al. [47] stellt dar, dass durch die Auswertung von Webseiten (z. B. Nachrichtenseiten, Blogs und Foren) der zeitliche Verlauf von Notfallgeschehen erfasst wurde.

Mit der Kenntnis der unterschiedlichen Datentypen werden in Kapitel 3.2 am Beispiel der Stadt Hamburg exemplarische Datensätze beschrieben.

## **3.2 Datenökonomie am Beispiel der Stadt Hamburg**

Die Stadt Hamburg hat mit der Einführung des Hamburgischen Transparenzgesetzes als erstes Bundesland eine Vorreiterrolle im Bereich der offenen Daten eingenommen und die Voraussetzungen für eine offene Datenkultur geschaffen. Am Beispiel der Stadt Hamburg werden Datenquellen aufgezeigt, die eine Echtzeit-Stadt repräsentieren können. Unterschieden wird in Daten aus sozialen Sensoren,

technischen Sensoren und Open Data (vgl. Kapitel 2.2). Für die vorliegende Arbeit wird angenommen, dass interne Daten zum Schutz der öffentlichen Sicherheit und Ordnung einer Sicherheitsbeschränkung unterliegen. Eine Nachnutzung für Jedermann ist nicht vorgesehen, daher findet eine explizite Betrachtung interner Daten in dieser Arbeit nicht statt. Es werden offene Daten (Open Data) nach der Definition der Open Knowledge Foundation betrachtet und als geeignet eingestuft, weil die Nachnutzung sowie Kombination mit anderen Datensätzen zulässig ist [30].

Von den Behörden sowie Körperschaften, Anstalten, Stiftungen und Unternehmen der Freien und Hansestadt Hamburg werden Verwaltungsdaten veröffentlicht. Hierunter fallen Senatsbeschlüsse, Haushaltspläne, amtliche Statistiken, Gutachten und Studien, Geodaten, Ergebnisse von Messungen über schädliche Umwelteinwirkungen oder öffentliche Pläne. Ausgenommen sind personenbezogene Daten, die vor einer Veröffentlichung unkenntlich gemacht werden. Lediglich Informationen zum Schutz öffentlicher Belange, wie beispielsweise Informationen zur unmittelbaren Willensbildung oder vertrauliche Unterlagen zum Schutz der inneren Sicherheit, sind von der Pflicht zur Veröffentlichung ausgenommen [19].

Von der Stadt Hamburg wird seit Februar 2013 ein eigenes Transparenzportal betrieben. Bisher wurden über 1.700 Datensätze und über 47.000 Dokumente bereitgestellt. Die Strukturierung der Daten in 14 Kategorien ermöglicht dem Anwender eine einfache Suche (Abbildung 4). Mühlbacher et al. [S. 320] [48] haben die Datenbestände des Portals untersucht und stellten fest, dass über 90 Prozent der Datensätze der Kategorie Umwelt & Klima zugeordnet werden. Lediglich etwas über 2,6 % der Datensätze sind der Kategorie Geografie, Geologie & Geodaten zugeordnet, dichtgefolgt von der Kategorie Infrastruktur, Bauen & Wohnen (1,8 %) sowie



Transport & Verkehr (1,6 %). Knapp 44 % der Dokumente sind der Kategorie Infrastruktur, Bauen & Wohnen zugeordnet [48].

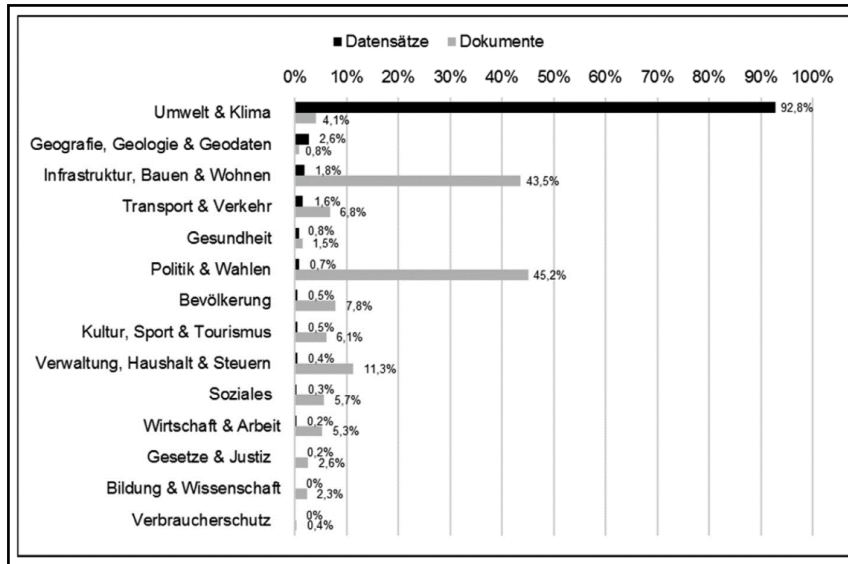


Abbildung 4: 14 Kategorien des Transparenzportals der Stadt Hamburg [48]

Die Datensätze werden in dreizehn und die Dokumente in neun unterschiedlichen Datenformaten veröffentlicht. Rund 90 Prozent der Datensätze werden ausschließlich im CSV-Format angeboten (Abbildung 5). Bei den Dokumenten sind PDF-Dateien und HTML-Datenformate die gängigen Datenformate. CSV-Formate beinhalten strukturierte Daten und sind maschinenlesbar. PDF-Dateien und HTML-Formate sind un- bzw. semistrukturierte Formate, deren maschinelle Verarbeitung eine entsprechende Vorbereitung erfordert (vgl. Kapitel 3.1).

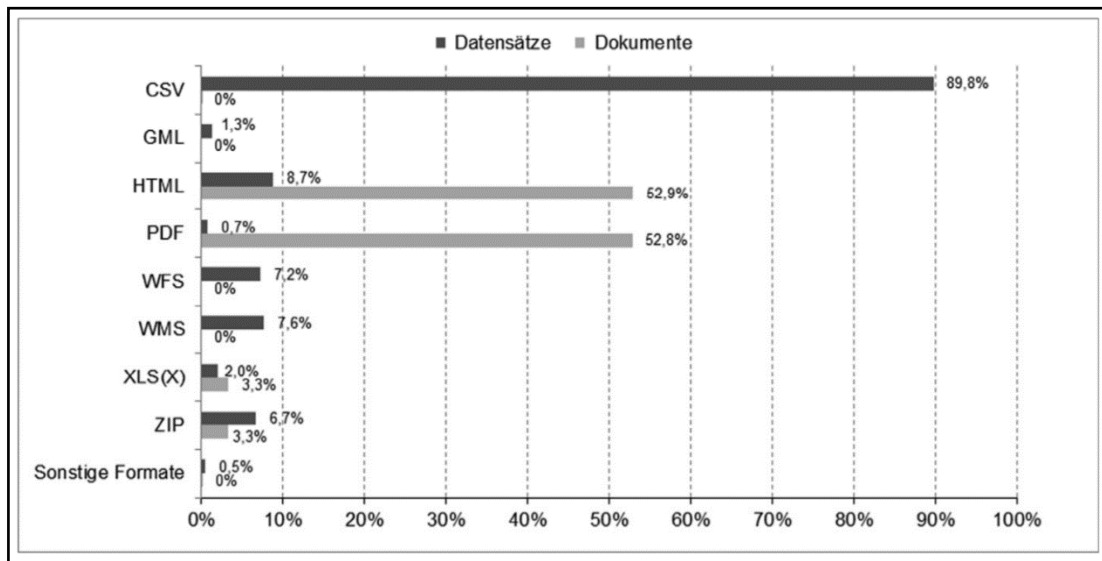


Abbildung 5: Datenformate im Transparenzportal der Stadt Hamburg [48]

Abbildung 6 veranschaulicht die häufigsten Suchanfragen der Nutzer des Hamburger Transparenzportals im Monat Mai 2018. Zu den favorisierten Suchanfragen zählen unter anderem Naturschutz, Kriminalität, Finanzen und Smart-City [49].

Die Zahlen lassen die Behauptung zu, dass die Bürger an einer Kollaboration mit der Stadt (Smart City) interessiert sind. In weiteren wissenschaftlichen Arbeiten bleibt die Hypothese zu prüfen, ob bei der Bevölkerung eine Bereitschaft zur Mitarbeit im Bereich der Gefahrenabwehr besteht.



Abbildung 6: Häufige Suchanfragen im Transparenzportal (Mai 2018) [49]

Murjahn und Tegtmeyer [S. 335] [50] beschreiben, dass Echtzeitdaten zur Straßenverkehrslage auf Autobahnen oder aktuelle Angaben zur Auslastung von Parkhäusern ebenfalls im besonderen Interesse der Nutzer von Transparenzportalen sind.

In Bezug auf eine algorithmische Auswertung von Datenbeständen zur Erkennung von Gefahrenlagen werden in Tabelle 2 beispielhafte Datensätze aus dem Transparenzportal der Stadt Hamburg und weiteren Quellen aufgeführt. Zur Auswahl der Daten wurden als Kriterien die Repräsentation von physikalischen sowie nicht-physikalischen Elementen und offene Daten berücksichtigt. Zudem sollten die Daten über ein möglichst kurzes Aktualisierungsintervall verfügen.

Tabelle 2: Beispielhafte Datenquellen

Nr.	Datensatz	Quelle	staatlich/ privat	Beschreibung	Anwendung (Relevanz)	Daten- format	Aktualisierungs- intervall	Link
1	3D-Stadtmodell	Transparenz-portal Hamburg	staatlich	Räumliche Bezugsgrundlage  Darstellung der Gebäude	Aufbau eines Geoinformationssystems  Repräsentation der Gebäudelandschaft	CITYGML	statische Grunddaten	<a href="http://suche.transparenz.hamburg.de/dataset/3d-stadtmodell-lod1-de-hamburg1?forceWeb=true">http://suche.transparenz.hamburg.de/dataset/3d-stadtmodell-lod1-de-hamburg1?forceWeb=true</a>  <a href="http://suche.transparenz.hamburg.de/dataset/3d-stadtmodell-lod2-de-hamburg2?forceWeb=true">http://suche.transparenz.hamburg.de/dataset/3d-stadtmodell-lod2-de-hamburg2?forceWeb=true</a>
2	AKTIS Digitales Basis Landschaftsmodell Hamburg	Transparenz-portal Hamburg	staatlich	Digitales Basis- Landschafts- modell	Geometrische und semantische Basisgrundlage für den Aufbau eines Geoinformationssystems  Verknüpfung mit raumbezogenen fachspezifischen Daten	GML	statische Grunddaten	<a href="http://suche.transparenz.hamburg.de/dataset/atkis-digitales-basis-landschaftsmodell-hamburg11?forceWeb=true">http://suche.transparenz.hamburg.de/dataset/atkis-digitales-basis-landschaftsmodell-hamburg11?forceWeb=true</a>
3	Digitales Höhenmodell	Transparenz-portal Hamburg	staatlich	Darstellung von Höhenlinien und Profile	Repräsentation der Stadtlandschaft  Bewertung von Hochwassergefahren	ASCII	statische Grunddaten	<a href="http://suche.transparenz.hamburg.de/dataset/digitales-hohenmodell-hamburg-dgm-104?forceWeb=true">http://suche.transparenz.hamburg.de/dataset/digitales-hohenmodell-hamburg-dgm-104?forceWeb=true</a>
4	Gefahrgutstraßen Hamburg	Transparenz-portal Hamburg	staatlich	Ausgewiesene Gefahrgutstraßen in Hamburg	Repräsentation der Stadtinfrastruktur  Identifikation von möglichen Gefahrenstellen	GML	statische Grunddaten	<a href="http://suche.transparenz.hamburg.de/dataset/gefahrgutstrassen-hamburg2?forceWeb=true">http://suche.transparenz.hamburg.de/dataset/gefahrgutstrassen-hamburg2?forceWeb=true</a>

Nr.	Datensatz	Quelle	staatlich/ privat	Beschreibung	Anwendung (Relevanz)	Daten- format	Aktualisierungs- intervall	Link
5	Störfallbetriebe Hamburg	Transparenz- portal Hamburg	staatlich	Standorte der Störfallbetriebe	Repräsentation der Stadtinfrastruktur  Identifikation von möglichen Gefahrenstellen	GML	statische Grunddaten	<a href="http://suche.transparenz.hamburg.de/dataset/storfallbetriebe-hamburg11?forceWeb=true">http://suche.transparenz.hamburg.de/dataset/storfallbetriebe-hamburg11?forceWeb=true</a>
6	Hamburger Luftmessnetz	Transparenz- portal Hamburg	staatlich	Erfassung von Schadstoff- komponenten Schwefeldioxid (SO <sub>2</sub> ), Stickstoff- monoxid (NO), Stickstoffdioxid (NO <sub>2</sub> ) und Staub (Feinstaub/PM10 : Partikel kleiner als 10 Mikrometer) Einige Stationen messen außerdem Kohlenmonoxid (CO)	Allgemeine Überwachung der Luftqualität	HTML CSV	Stunden-, Tages-, Monatswerte	<a href="http://suche.transparenz.hamburg.de/dataset/daten-des-hamburger-luftmessnetzes-halm1?forceWeb=true">http://suche.transparenz.hamburg.de/dataset/daten-des-hamburger-luftmessnetzes-halm1?forceWeb=true</a>  <a href="http://luft.hamburg.de/">http://luft.hamburg.de/</a>
7	Pegel an Binnengewässer	Transparenz- portal Hamburg	staatlich	Messdaten von ca. 400 Pegel- Messstellen Kontinuierliche Wasserstands- messung von Oberflächen-	Erkennung von Hochwassergefahren	CSV	5 Minuten Takt	<a href="https://www.wabiha.de/karte.html">https://www.wabiha.de/karte.html</a>

Nr.	Datensatz	Quelle	staatlich/ privat	Beschreibung	Anwendung (Relevanz)	Daten- format	Aktualisierungs- intervall	Link
				gewässern (ohne Elbe)				
8	Gamma-Ortsdosisleistung	Bundesamt für Strahlenschutz	staatlich	Messung der Gamma-Ortsdosisleistung (ODL) an etwa 1800 Stationen	Erkennung von höherer Strahlenbelastung	DAT JSON SVG	1-Stunden- und 24-Stunden-Mittelwert	<a href="https://odlinfo.bfs.de/DE/index.html">https://odlinfo.bfs.de/DE/index.html</a>
9	Pegelstandsmessung	Wasserstraßen- und Schifffahrtsverwaltung des Bundes	staatlich	Gewässer-kundliche Parameter von Binnen- und Küstenpegel	Erkennung von Hochwassergefahren	JSON	Live-Daten bis max. 30 Minuten	<a href="http://www.pegelonline.wsv.de/gast/start">http://www.pegelonline.wsv.de/gast/start</a>
10	Verkehrskameras	Transparenzportal Hamburg	staatlich	Auswahl von Verkehrskameras mit Standorten und Livebildern	Erkennung von Staulagen  Verbesserung der Situationsaufmerksamkeit	GML	nicht dauerhaft online	<a href="http://suche.transparenz.hamburg.de/dataset/verkehrskameras-hamburg12?forceWeb=true">http://suche.transparenz.hamburg.de/dataset/verkehrskameras-hamburg12?forceWeb=true</a>
11	Luftdaten	Luftdaten.info	privat	Messung der Feinstaubbelastung	Allgemeine Überwachung der Luftqualität  Erkennung von Überschreitung der Feinstaub-Grenzwerte	JSON	1-Minuten-Intervall	<a href="https://luftdaten.info/faq/#toggle-id-8">https://luftdaten.info/faq/#toggle-id-8</a>
12	Wetterdaten	Deutscher Wetterdienst	staatlich	Geodaten zu Modellvorhersagen, Radardaten, Bodenbeobachtungen,	Niederschlagsmessung (Bewertung von Hochwassergefahren)  Bewertung des Waldbrandrisikos	BUFR GRIB	Live-Daten	<a href="https://www.dwd.de/DE/leistungen/opendata/opendata.html">https://www.dwd.de/DE/leistungen/opendata/opendata.html</a>

Nr.	Datensatz	Quelle	staatlich/ privat	Beschreibung	Anwendung (Relevanz)	Daten- format	Aktualisierungs- intervall	Link
				Klimadaten Warndaten	Vorhersage von Unwettergefahren			
13	Bevölkerungs- zahlen	Statistisches Amt für Hamburg und Schleswig- Holstein	Staatlich	Bevölkerung am Ende des Monats in Schleswig- Holstein und Hamburg	Aufbau eines Geoinformationssystems  Beschreibung des Bezugsgebietes	CSV	Monatlich und jährlich	<a href="https://www.statistik-nord.de/zahlen-fakten/regionalstatistik-datenbanken-und-karten/monatszahlen/">https://www.statistik-nord.de/zahlen-fakten/regionalstatistik-datenbanken-und-karten/monatszahlen/</a>
14	Tweets	Twitter	privat	Erfassung von öffentlichen Beiträgen im Microblog-Dienst Twitter	Detektion von Gefahrenlagen  Situations- aufmerksamkeit	JSON	Live-Daten	<a href="https://developer.twitter.com/en/docs/tweets/filter-realtime/api-reference/post-statuses-filter.html">https://developer.twitter.com/en/docs/tweets/filter-realtime/api-reference/post-statuses-filter.html</a>
15	Beiträge	Flickr	Privat	Erfassung von öffentlichen Beiträgen der Fotocommunity Flickr	Detektion von Gefahrenlagen  Situations- aufmerksamkeit		Live-Daten (stündlich)	<a href="https://www.flickr.com/services/api/flickr.activity.userPhotos.html">https://www.flickr.com/services/api/flickr.activity.userPhotos.html</a>
16	Beiträge	Instagram	Privat	Erfassung von öffentlichen Beiträgen der Fotocommunity Flickr	Detektion von Gefahrenlagen  Situations- aufmerksamkeit		Live-Daten	<a href="https://www.instagram.com/developer/endpoints/media/">https://www.instagram.com/developer/endpoints/media/</a>

## **4 Sensorik-Leitstellen: Anforderungen und Konzeption**

Gefahrenlagen sind besonders in der Anfangsphase von einem Informationsdefizit geprägt. In der Abfolge einer Gefahreneinschätzung muss in kürzester Zeit ein ständig aktueller Informationspool aus verschiedenen Datenquellen zur Verfügung stehen. Daraus werden gefahrenrelevante Daten generiert, ausgewertet und beurteilt, mit dem Ziel passende Entscheidungen zur Eindämmung einer Gefahrenlage zu treffen. Bisher führen speziell ausgebildete Kräfte der Gefahrenabwehrbehörden den Prozess der Lagebeurteilung durch. Sensorik-Leitstellen können mit Algorithmen die Verarbeitung und Darstellung von Informationen bei der Lagebewältigung unterstützen. In Kapitel 4.1 werden zunächst die grundsätzlichen Anforderungen an eine Sensorik-Leitstelle basierend auf den Erkenntnissen der Mensch-Computer-Interaktion beschrieben. Anschließend wird in Kapitel 4.2 ein Konzept für den Aufbau einer Sensorik-Leitstelle in Form einer Schichtenarchitektur vorgestellt.



## 4.1 Grundsätzliche Anforderungen

Turoff et al. [51] und Carver & Turoff [52] haben die Mensch-Computer-Interaktion (Human Computer Interaction) in Notfallszenarien betrachtet und formulierten Anforderungen an ein System für die Gefahrenabwehr. Ausgehend von den beiden Arbeiten werden für eine Sensorik-Leitstelle folgende Anforderungen beschrieben [51,52]:

- **Einbindung verschiedener Datenquellen:** In einer Gefahrenlage sind Informationen (ob qualitativ oder quantitativ) zur Lagererkundung unerlässlich. Sowohl Daten der Bürger als auch technische Daten (z. B. aus Sensornetzwerken) sollen erfasst werden. Zudem sind Daten und Informationen anderer Fachbehörden oder Institutionen einzubeziehen. Zur Informationsgewinnung ist eine Verknüpfung der Informationen zielführend, um Zusammenhänge und Muster zu erkennen. Ebenso können Anomalien detektiert werden, die auf ein mögliches Notfallgeschehen hindeuten.
- **Echtzeiterfassung:** Eine Erfassung und Auswertung der Daten in Echtzeit ist ein wesentliches Merkmal. Bei älteren Datenbeständen muss die Zeitspanne bzw. der Zeitpunkt der Datenerfassung angegeben werden. Zur Überprüfung der Relevanz ist neben der zeitlichen auch eine räumliche Beschreibung erforderlich.
- **Multimethoden Ansatz zur Auswertung:** Komplexe Datenbestände erfordern unterschiedliche Methoden, Techniken und Algorithmen der Mathematik, der Statistik und des maschinellen Lernens (Künstliche Intelligenz) zur Auswertung. Die Algorithmen der künstlichen Intelligenz sollen nach einer Anlernphase selbständig und autark die Datenströme auswerten.

- **Einbeziehung von Fachexperten:** Das System für die Gefahrenabwehr ist durch Fachexperten anzulernen. Vergleichsebenen zur Auswertung können nur in Zusammenarbeit der Informatik und der Fachdisziplin erzeugt werden [53]. Ebenso sind die Interpretation der Informationen und das Treffen von Maßnahmen den Fachexperten vorbehalten.
- **Visuelle Darstellung:** Eine übersichtliche Visualisierung ist u. a. für das Treffen von Entscheidungen unerlässlich. Ein Visualisierungssystem für den Einsatz in der Gefahrenabwehr muss unter Zeitdruck und in komplexen Situationen die Zusammenarbeit verschiedener Fachbehörden ermöglichen.

Die Anforderungen an eine Sensorik-Leitstelle verdeutlichen, dass im Kernkonzept urbane Daten erfasst, algorithmisch ausgewertet und visualisiert werden. In Kapitel 4.2 werden technische Aspekte zum Aufbau einer Sensorik-Leitstelle beschrieben.

## 4.2 Technische Konzeption

Zur Realisierung einer Sensorik-Leitstelle wird eine aus vier Schichten bestehende Schichtenarchitektur vorgeschlagen (Abbildung 7).

Klassischer Weise bestehen Schichtenmodelle aus drei Schichten: Datenhaltung, Logik und Benutzeroberfläche [54]. Ausgehend von einer klassischen dreischichtigen Architektur besteht das vorgeschlagene System aus den drei Schichten

- presentation layer (Benutzeroberfläche) (vgl. Kapitel 4.2.5),
- logical layer (Logik) (vgl. Kapitel 4.2.3) und
- data layer (Datenhaltung) (vgl. Kapitel 4.2.2).

Zudem wird das System um eine vierte Schicht zur Datensuche (sensing layer) (vgl. Kapitel 4.2.1) ergänzt.

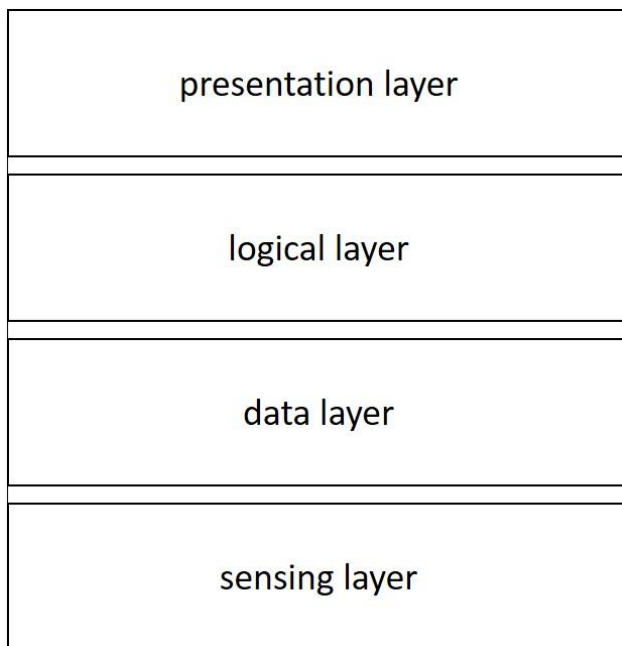


Abbildung 7: Schichtenarchitektur für eine Sensorik-Leitstelle

Der Vorteil einer mehrschichtigen Architektur besteht in der Austauschbarkeit von einzelnen Modulen. Beispielsweise könnte die Benutzeroberfläche verändert werden, ohne die darunter liegenden Schichten in der Funktionalität zu beeinflussen [54]. In den folgenden Kapiteln werden die einzelnen Schichten vorgestellt.

#### **4.2.1 Sensing layer – Die Suche nach Informationen**

Wie in Kapitel 2.2 beschrieben, geht es im Wesentlichen um die Erfassung von Daten der Bürger, der technischen Sensoren und internen Datenquellen. Die Datensammlung kann über die Nutzung von Programmierschnittstellen, die aktive Suche im Internet/auf Webseiten oder die aktive Eingabe erfolgen.

Programmierschnittstellen, sogenannte application programming interfaces (API), sind Schnittstellen in Softwaresystemen, die einen Datenaustausch zwischen zwei Programmen ermöglichen. Der Zugriff auf eine API kann sowohl öffentlich als auch nach einer Registrierung möglich sein. Von zahlreichen Webservices<sup>8</sup>, zum Beispiel den sozialen Medien wie Twitter, Facebook, Suchmaschinen, Open Data-Portalen oder technischen Sensornetzwerken, werden API angeboten. Die Verwendung ist vorteilhaft, da Daten in einem standardisierten Format (z. B. im Datenaustauschformat JSON) angeboten werden.

Technisch aufwändiger ist die aktive Datensuche auf Webseiten mit Webcrawlern oder durch Screen Scraping. Webcrawler sind spezielle Programme, die darauf ausgerichtet sind, die Webseiten des Internets zu durchsuchen. Die Crawler folgen dabei den Verlinkungen (Uniform Resource Locator (URL)), um selbstständig neue Webseiten untersuchen zu können. Auf diese Weise ist eine Speicherung und Kategorisierung der analysierten Webseiten möglich [43,55]. Screen Scraping beschreibt eine technische Methode zur vertiefenden Auswertung von Webseiten. Im Wesentlichen basiert dieses Konzept auf der Auswertung eines Quelltextes von Webseiten und der anschließenden Speicherung der enthaltenen Informationen. Kommerzielle Anbieter (beispielsweise [import.io](https://import.io)) ermöglichen die Umwandlung von Webseiten in Daten. Über eine API werden die Daten anschließend bereitgestellt [56,57].

---

<sup>8</sup> Die Webseite <https://apilist.fun/> stellt eine Auswahl von gängigen API zusammen.

Die Einbindung von internen Informationen ist von dem vorliegenden Datenformat abhängig. Maschinenlesbare Datenformate können sowohl manuell als auch automatisiert erfasst werden.

#### **4.2.2 Data layer – Die Speicherung von Daten**

Zur Speicherung der erfassten Daten sind Datenbanken erforderlich, die neben der Speicherung die Verwaltung und die beliebig häufige Manipulation ermöglichen sollen. Die Speicherung der Daten erfolgt in einem programmunabhängigen Datenformat auf einem persistenten Datenspeicher (z. B. Festplatten) [54]. Zur Speicherung und Analyse von Daten aus heterogenen Quellen haben sich Data Warehouse durchgesetzt. Diese Form einer Datenbank ist in der Lage für eine spätere Analyse diverse Datensätze zu einem konsistenten Datensatz zusammenzuführen [58].

#### **4.2.3 Logical layer – Die Anwendung von Algorithmen**

Die Suche nach Mustern und Zusammenhängen (Extraktion von Wissen) erfordert unterschiedliche Methoden, Techniken und Algorithmen der Mathematik, der Statistik und des maschinellen Lernens (Künstliche Intelligenz) (siehe Abbildung 8) [39]. Zudem verlangt eine Erfassung und Auswertung in Echtzeit automatisierte Prozesse [35].

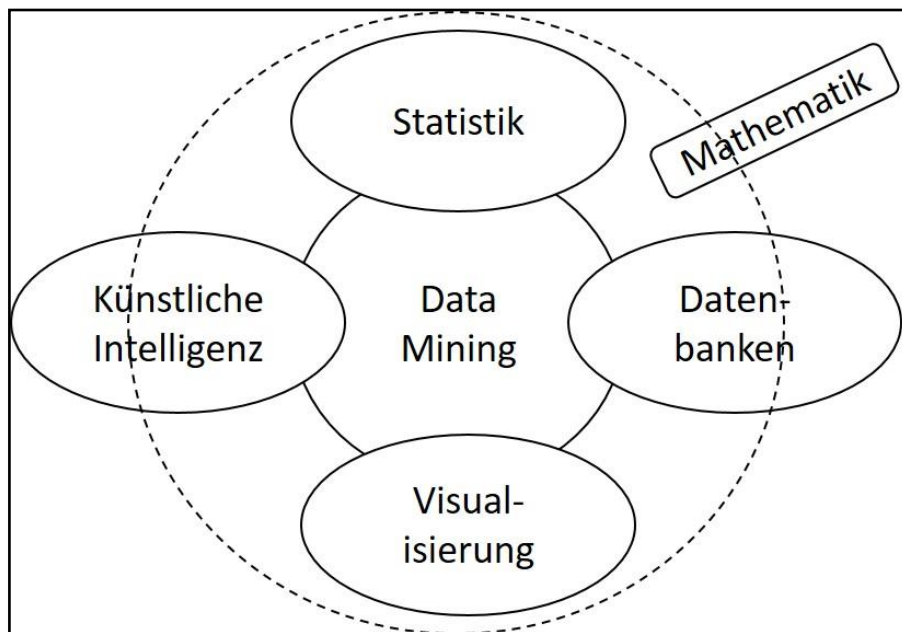


Abbildung 8: Data Mining - ein interdisziplinärer Ansatz nach Cleve und Lämmel [39]

Zwei Herausforderungen können im Umgang mit urbanen Daten identifiziert werden. Zum einen ist eine andere Art der Datenreduktion und Feature Extraktion notwendig, die ohne Hilfe von domänenspezifischen Fachexperten nicht zu leisten ist. Zum anderen fehlen bei komplexen Daten vorgegebene Vergleichsebenen für die Feature Vektoren. Entsprechende Vergleichsebenen müssen erst durch eine Kontextualisierung basierend auf passendem Fachwissen kollaborativ und explorativ geschaffen werden [53,59]. In diesem Zusammenspiel zwischen Data Science und externen Fachwissen (Domain-specific knowledge), können urbane Phänomene semantisch und adäquat repräsentiert werden [60].

Zur Informationssuche in urbanen Daten ist ein unstetiger, inkrementeller Prozess entlang des von Fayyad et al. [61] beschriebenen Auswertungsprozess „Knowledge Discovery in Databases“ Prozesses (vgl. Kapitel 4.2.4) erforderlich.

#### 4.2.4 Knowledge Discovery in Databases (KDD-Prozess)

Der KDD-Prozess ist ein iterativer Ansatz zur Suche nach Mustern und Zusammenhängen in großen Datenbeständen (Big Data). Der Ablauf einer Datenanalyse (siehe Abbildung 9) besteht aus mehreren Schritten [39]:

- **Selektion:** die Auswahl der geeigneten Daten
- **Datenvorbereitung:** die Datenbereinigung sowie die Behandlung von fehlenden oder widersprüchlichen Daten
- **Transformation:** die Umwandlung in maschinenlesbare Datenformate
- **Data Mining:** die eigentliche Suche nach Mustern
- **Interpretation und Evaluation:** die Auswertung der Ergebnisse

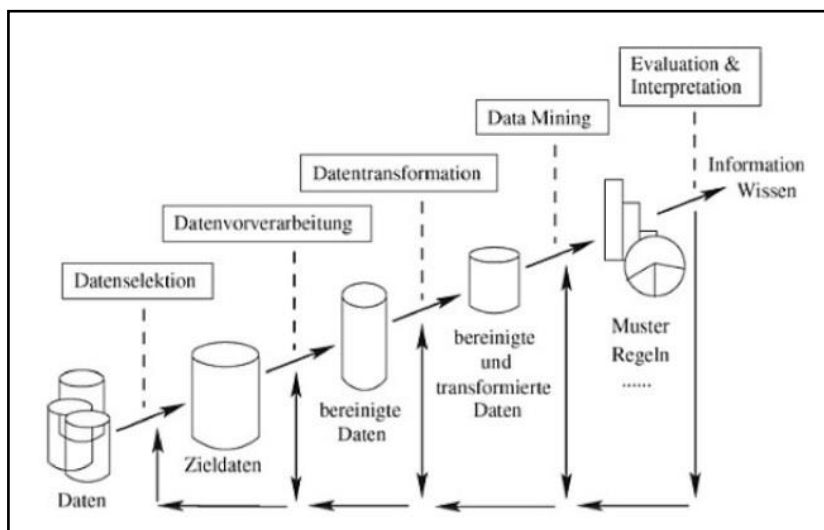


Abbildung 9: Ablauf einer Informationsgewinnung aus großen Datenbeständen [39]

Anders als von Fayyad et al. [61] vorgestellt, besteht bei urbanen Daten in einem ersten Schritt nicht die Möglichkeit domainspezifisches Wissen aufzubereiten, weil dies nicht vollständig vorliegt oder erst in der Interaktion mit der Datenanalyse entsteht. Zudem ist vorab nicht bekannt, nach welcher Anomalie gesucht wird.

Mit einer Sensorik-Leitstelle sollen Anomalien in den urbanen Daten erkannt werden. In einem ersten Schritt sind die regulären Datenströme zu erfassen, um im zweiten Schritt Anomalien als Abweichungen zu erkennen. Das Vorgehen basiert auf einem stetigen Feedback, um die Algorithmen der künstlichen Intelligenz anzupassen. Assendorp [S. 5-7] [14] hat im Rahmen seiner Masterarbeit eine angepasste Version des KDD-Prozesses zur Erkennung von Anomalien vorgestellt. Abbildung 10 zeigt den angepassten Prozess mit ständigen Feedbackschleifen.

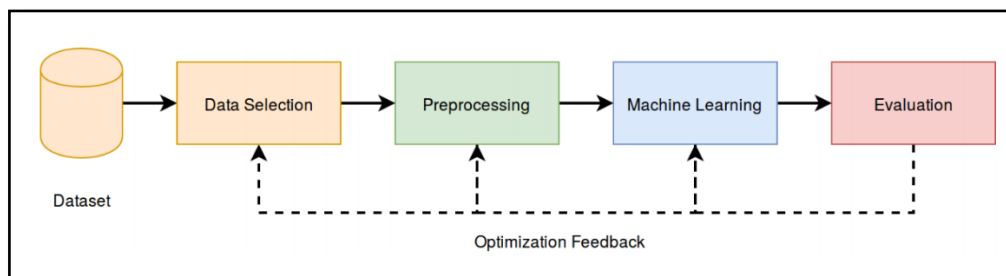


Abbildung 10: Angepasster Data Mining Ablauf zur Anomalieerkennung [14]

Der Datensatz (Dataset) ist die Gesamtheit der urbanen Daten. Nach einer Auswahl/Selektion (Data Selection) und Vorbereitung der Daten (Preprocessing) erfolgt eine automatische Auswertung mit Algorithmen des maschinellen Lernens (Machine Learning). Abschließend werden die Informationen zur Auswertung (Evaluation) dargestellt. Wesentlicher Bestandteil ist die Optimierung der Algorithmen (Machine Learning) durch ein stetiges Feedback.

Ausgehend von den unterschiedlichen Datentypen in urbanen Daten sind verschiedene Methoden zur Informationsgewinnung notwendig. Es müssen Daten aus Datenbanken (z. B. Open Data-Portale), Informationen von Webseiten oder Blogs, digitale Textdokumente, schriftliche Beiträge in sozialen Medien und mediale



Dateien (Bild, Video, Audio) ausgewertet werden. In Kapitel 5 werden die Methoden zur Auswertung beschrieben.

#### **4.2.5 Presentation layer – Die Visualisierung von urbanen Daten**

Die Visualisierung zielt einerseits auf die Darstellung der Informationen ab, andererseits ist die Visualisierung selbst eine Auswertungsmethode.

Eine geeignete visuelle Darstellung soll einem Anwender ohne Kenntnisse über die Algorithmen und Parameter die Möglichkeit geben, große Datenbestände auszuwerten. Besonders geeignet ist diese Methode zur Auswertung von unbekanntem Daten [62].

Klassische Visualisierungstechniken sind aus der deskriptiven Statistik bekannt und umfassen unter anderem Tabellen, Diagramme, Grafiken oder Boxplots. Diese Art und Weise der Darstellungsformen ist für kleinere Datenbestände geeignet [39,63]. Für die Darstellung in einer Sensorik-Leistelle sind diese Formen nützlich, jedoch erfordert die Zusammenführung von unterschiedlichen Informationen geeignetere Präsentationsformen. Zudem müssen die Präsentationsformen die Zusammenarbeit von Fachexperten und die Darstellung eines gesamten (Stadt-)Lagebildes ermöglichen [64].

Keim [62] beschreibt Interaktions- und Verzerrungstechniken als effektive Methoden zur visuellen Datenexploration. Interaktionstechniken werden eingesetzt um verschiedene Visualisierungstechniken miteinander zu verbinden. Zur Darstellung von Ausschnitten einer Datenmenge werden Verzerrungstechniken eingesetzt. Dabei werden die übrigen Daten mit einem geringeren Detaillierungsgrad dargestellt.

Zusätzlich kann zwischen dynamischen und interaktiven Ansätzen unterschieden werden. Bei den dynamischen Techniken wird die visuelle Darstellung automatisch angepasst. Bei der interaktiven Technik erfolgt eine Anpassung der Darstellungsform durch die Interaktion mit den Daten [62]. Auf diese Weise können Informationen aus unterschiedlichen Quellen verknüpft werden, ohne dass eine Veränderung der Datenquellen oder der Hintergrundsysteme notwendig ist [64]. Technische Möglichkeiten zur Visualisierung werden in Kapitel 6 vorgestellt.

Zur Auswertung der urbanen Daten werden in Kapitel 5 verschiedene Methoden der künstlichen Intelligenz, als Disziplin zur Automatisierung von intelligentem Verhalten und dem Lernen von Maschinen, vorgestellt.

## 5 Algorithmen: Methoden zur Auswertung

Dem Grundgedanken einer Sensorik-Leitstelle folgt der Ansatz, dass Gefahrenlagen im städtischen Raum automatisiert erkannt werden. Es erscheint naheliegend, Mitteilungen von Bürgern aus sozialen Netzwerken (beispielsweise ein Tweet auf Twitter) als einen ersten Hinweis auf eine Gefahrenlage zu nutzen [18]. Dabei sind Methoden des Text Mining, als eine Teildisziplin des Data Mining, notwendig. Data Mining (vgl. Kapitel 5.5) sowie Text Mining (vgl. Kapitel 5.6) und Web Mining (vgl. Kapitel 5.7) als Spezifikation des Data Minings, basieren auf unterschiedlichen Methoden der künstlichen Intelligenz. Neben den Mining-Verfahren können beispielsweise künstlichen neuronale Netzwerke (vgl. Kapitel 5.1) zur Auswertung genutzt werden. Das Anlernen dieser Netzwerke wird in Kapitel 5.2 und 5.4 beschrieben.

Für die Erkennung von urbanen Gefahrenlagen, im Sinne von Anomalien in Datenströmen, wird das System in drei Bereiche gegliedert (siehe Tabelle 3).

Tabelle 3: Drei Systemkomponenten einer Sensorik-Leitstelle

	<b>Kontinuierliche Auswertung</b>	<b>Trainingsumgebung</b>	<b>Anomalieerkennung</b>
<b>Ziele</b>	<ul style="list-style-type: none"> <li>• Stadt kennen lernen</li> <li>• Abbildung der Realität</li> <li>• Vergleichsebene für Feature Vektoren schaffen</li> </ul>	<ul style="list-style-type: none"> <li>• anhand von Trainingsdaten die Anomalieerkennung trainieren</li> <li>• Feature Vektoren trainieren</li> </ul>	<ul style="list-style-type: none"> <li>• Dynamische Kontextualisierung</li> <li>• Erkennen von besonderen Lagen/Notfallsituationen (Anomalien)</li> <li>• Bereitstellung von Wissen über Anomalien an die Trainingsumgebung</li> </ul>

### **Kontinuierliche Auswertung**

Zur Beschreibung des Bezugsgebietes werden im ersten Schritt die regulären Datenströme identifiziert. Mit diesen Daten wird für die spätere Identifikation von Abweichungen ein Abbild der Realität (Vergleichsebene) geschaffen. Die regulären Datenströme einer Stadt bestehen aus offenen (Verwaltungs-)Daten, technischen Sensordaten und Beiträgen aus sozialen Netzwerken. Zur Auswertung werden Methoden des Data Mining, der Texterkennung und der Bilderkennung eingesetzt.

### **Trainingsumgebung**

Anschließend wird das System in Zusammenarbeit mit Fachexperten und Trainingsdaten angelernt. Trainingsdaten können Textbeiträge aus sozialen Netzwerken, Bilder, Videos oder technischen Sensordaten sein, die eine reale Gefahrenlage abbilden. Das Ergebnis sind Feature Vektoren. In einem Feature Vektor sind parametrisierte Merkmale einer Gefahrenlage in vektorieller Weise

zusammengefasst. Mit den Vektoren ist das System in der Lage selbstständig Gefahrenlagen zu erkennen.

### **Anomalieerkennung**

Zur automatischen Erkennung von Gefahrenlagen wird das System in die Datenströme und Netzwerke einer Stadt integriert. Der stetige Abgleich mit den Feature Vektoren ermöglicht die Detektion von Anomalien. Im Falle einer Anomalie werden die Daten erfasst und visualisiert. Zusätzlich können ergänzende Informationen aus der kontinuierlichen Auswertung angezeigt werden. In der praktischen Anwendung wäre es denkbar, dass beispielsweise Hintergrundinformationen (z. B. das Kanalisationssystem) über das aktuelle Lagebild (z. B. Chemieunfall) gelegt werden.

Die Unterteilung des Systems in die drei beschriebenen Bereiche ermöglicht die Identifikation von erforderlichen Datensätzen und die zur Auswertung notwendigen Methoden. Zum grundlegenden Verständnis für das maschinelle Lernen wird in Kapitel 5.1 das Konzept der künstlichen neuronalen Netzwerke beschrieben.

## **5.1 Künstliche neuronale Netzwerke**

Das Lernen von Maschinen wird vom Lernen des Menschen inspiriert. Ähnlich dem neuronalen Netzwerk des menschlichen Gehirns wird versucht künstliche neuronale Netzwerke zu schaffen. Auf diese Weise können Maschinen komplexe Aufgaben lösen [65,66].

Zum besseren Verständnis wird ein künstliches neuronales Netzwerk mit einem biologischen Neuron verglichen. Eine menschliche Nervenzelle besteht aus mehreren Dendriten, einem Zellkörper, einem Zellkern, einem Axon und mehreren Synapsen.

Die Steuerung der Nervenzelle erfolgt durch den Zellkern. Ein ankommender Reiz wird von den Dendriten aufgenommen. Eine elektrische Potentialveränderung der Zellmembran ermöglicht die Reizweiterleitung über das Axon hin zu den Synapsen. Die Weitergabe des Reizes an nachgeschaltete Neuronen oder Muskelzellen erfolgt über den Spalt zwischen den Synapsen und den Dendriten. Zahlreiche Neuronen ermöglichen die Signalverarbeitung im menschlichen Gehirn [65,66].

Künstliche neuronale Netzwerke sind an den biologischen Modellen orientiert. Auf vielfältige Weise können künstliche Neuronen miteinander verbunden werden. Zu einem besseren Verständnis werden die künstlichen Neuronen in Schichten angeordnet. Jede einzelne Schicht repräsentiert eine gewisse Anzahl von lernfähigen Neuronen. Grundsätzlich besteht ein neuronales Netzwerk aus einer Eingabeschicht, einer verborgenen Schicht und einer Ausgabeschicht. Über die Eingabeschicht werden die zu verarbeitenden Daten in das System eingegeben. Die Verarbeitung erfolgt durch die künstlichen Neuronen der verborgenen Schicht. Die Auswertung wird über die Ausgabeschicht zurückgegeben. Abbildung 11 zeigt den exemplarischen Aufbau eines künstlichen Netzwerkes, bestehend aus Neuronen (Knoten) und deren Verbindungen (Graphen) [65,66].

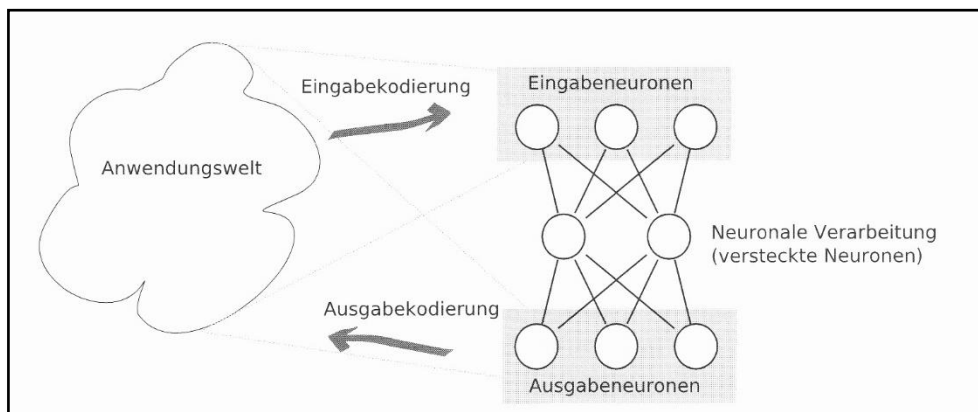


Abbildung 11: Exemplarischer Aufbau eines künstlichen Netzwerkes [65]

Die Netztopologie unterscheidet zwischen vorwärts gerichteten Netzwerken und Netzwerken mit Rückkopplungen (rekurrente Netze). Die Wahl der Netzstruktur hängt von den zu verwendenden Lernverfahren ab.

Vorwärtsgerichtete Netzwerke (Convolutional Neuronal Network) bestehen mindestens aus einer Eingabe- und Ausgabeschicht sowie einer oder mehrerer verborgenen Schichten. Das entscheidende Merkmal ist eine Verknüpfung der Neuronen in Richtung der Ausgabeschicht (siehe Abbildung 12).

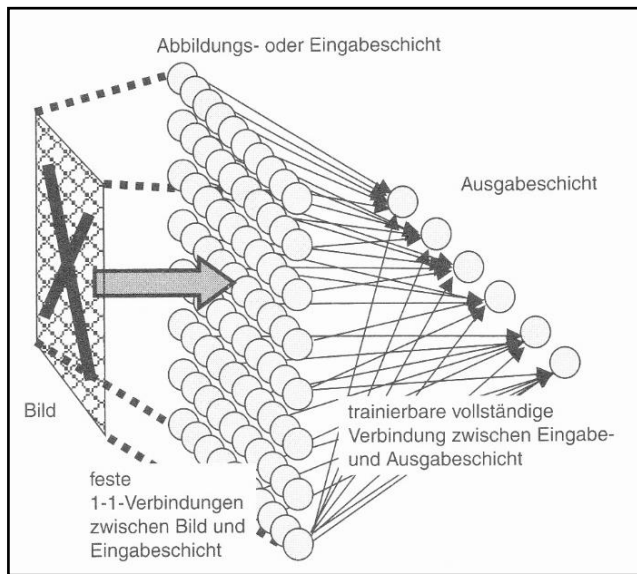


Abbildung 12: Aufbau eines vorwärtsgerichtete Netzwerkes [66]

Die Neuronen sind bei dieser Form des Netzes bereits untereinander verbunden.

Abbildung 13 stellt die unterschiedlichen Verbindungsmöglichkeiten dar.

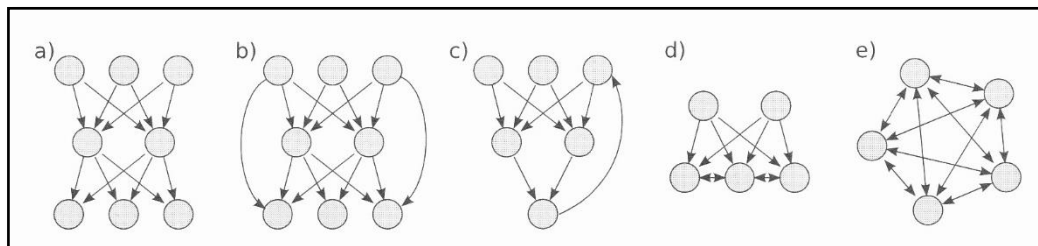


Abbildung 13: Verschiedene Netztopologien: a) vorwärts verkettet, b) mit Abkürzungen, c) mit Rückkopplungen, d) mit lateralen Verbindungen, e) voll vernetzt [65]

Rekurrente Netze (Recurrent Neuronal Network) sind besondere Netzformen, bei denen die Neuronen über rückwärtsgerichtete Graphen miteinander verbunden sind (siehe c) in Abbildung 13). Sogenannte Long short-termin memory (LSTM-)Techniken ermöglichen eine sequenzielle Verarbeitung und die Speicherung von Wissen im Sinne eines Gedächtnisses [65,66].



Besonders bei geringer oder keiner Kenntnis über das explizite Wissen zu einer Datenmenge, können neuronale Netzwerke zur Exploration eingesetzt werden. Weitere mögliche Anwendungsfelder sind [65]:

- **Klassifikation:** Ausgabe entspricht einer Klasse (Mustererkennung: Handschriften, Objekterkennung: Gesichter, Sprache)
- **Beurteilung:** Ausgabe entspricht einer Wertung
- **Assoziation:** Ausgabe entspricht einem Muster
- **Prognose:** Ausgabe entspricht einer Vorhersage
- **Steuerung:** Ausgabe entspricht einer Aktion

Eingesetzt werden die Methoden zur Text- oder Bildererkennung sowie zur Suche nach Mustern und Zusammenhängen in großen Datenbeständen (Data Mining). Künstliche Neuronale Netzwerke müssen vor dem Einsatz mit Trainingsdaten angelernet werden. In Kapitel 5.2 wird das Deep learning als Reihe von Techniken und Algorithmen zur Optimierung von künstlichen neuronalen Netzen beschrieben.

## 5.2 Deep learning

Speziell im Bereich der Sprach-, Gesichts- oder Texterkennung haben herkömmliche künstliche neuronale Netzwerke einen Nachteil, da sich die Prozesse nur bedingt mit mathematischen Modellen beschreiben lassen. Zur Auswertung von komplexen Daten werden vielzählige verborgene Schichten aufgebaut. Iterative Durchläufe dieser Schichten ermöglichen eine maschinelle Verarbeitung. Auf diese Weise können Objekte in Bildern automatisiert beschrieben werden [14]. Mit der Kenntnis zu neuronalen Netzen sowie ersten Lernaspekten werden in Kapitel 5.3 beispielhafte künstliche neuronale Netzwerke vorgestellt.

### 5.3 Beispielhafte künstliche neuronale Netzwerke

Je nach Anwendungszweck werden unterschiedliche Klassen bzw. Typen von künstlichen neuronalen Netzwerken gebildet. Für die Nutzung in einer Sensorik-Leitstelle sind Convolutional Neural Networks oder Self-Organizing Maps geeignete Architekturen.

Convolutional Neural Networks sind eine Form der künstlichen neuronalen Netzwerke zur Auswertung von Bild- und Audiodateien. Die Neuronen sind in einer zwei- oder dreidimensionalen Ordnung sortiert. Ein Bild, als eine Repräsentation in einer zwei- oder dreidimensionalen Matrix, wird von den Neuronen in kleine Teile (Schichten) zerlegt und separiert abgetastet. Die einzelnen Muster werden am Ende zusammengefügt, um ein oder mehrere Objekte zu erkennen [14].

Ein Anwendungsbeispiel ist die von Google entwickelte Cloud Vision API<sup>9</sup>, zur Identifizierung von Objekten und Gesichtern in Bildern [67].

Eine weitere Form sind Self-Organizing Maps oder selbstorganisierende Karten. Über eine Eingabeschicht werden Daten mit einer mehrdimensionalen Anordnung in das System eingegeben. Die darunterliegenden Neuronen können sowohl zwei- als auch dreidimensional angeordnet sein. Eine wesentliche Konzeption ist, dass nur die Neuronen angeregt werden, die den Eingabedaten am Nächsten liegen. Die Verbindungen der Neuronen bewirken, dass die anderen Neuronen in die Richtung der angeregten Neuronen ausgerichtet werden. Die Ausrichtung der Neuronen wird als topografische Abbildung dargestellt. Ein Anwendungsbeispiel ist die Clusterbildung im Text Mining [66]. In Kapitel 5.2 wurde das Deep Learning als eine Möglichkeit zum

---

<sup>9</sup> Weiterführende Informationen zur Cloud Vision API unter: <https://cloud.google.com/vision/>

Anlernen von neuronalen Netzen erläutert. In Kapitel 5.4 werden die drei wesentlichen Lernverfahren vorgestellt.

## 5.4 Maschinelles Lernen

Angelernt werden künstliche neuronale Netzwerke mit den Lernverfahren des maschinellen Lernens. Die verschiedenen Algorithmen werden in drei Lernparadigmen unterteilt [65]:

- **Überwachtes Lernen** (supervised learning)
- **Bestärkendes Lernen** (reinforcement learning)
- **Unüberwachtes Lernen** (unsupervised learning)

Überwachtes Lernen beschreibt das Lernen mit bekannten Eingabedaten und mit zu erwartenden Ausgabewerten. Ziel ist, dass das System aus einer Reihe von Rechengängen mit bekannten Ein- und Ausgabedaten selbständig Kriterien entwickelt und Assoziationen herstellt. Auf diese Weise können Systeme automatisiert Muster und Objekte in Bildern erkennen [65].

Eine Spezifikation des überwachten Lernens ist bestärkendes Lernen. Die Klassenzuordnung von Bildern kann vorab nicht genau beschrieben werden. Beim bestärkenden Lernen probiert das System verschiedene bekannte Algorithmen und lernt über das Feedback des Anwenders die richtigen Entscheidungen (in der Anwendung der Bilderkennung, der Zuordnung der richtigen Klasse) zu treffen [65,66].

Beim unüberwachten Lernen erfolgt kein Feedback des Anwenders an das System. Ausgehend von den Eingabedaten erzeugen die Algorithmen selbstständig ein Modell zur Beschreibung der Daten. Zur Unterscheidung der Muster werden Clustering-

Verfahren durchgeführt. Auf diese Weise lernt das System selbstständig wie es auf die Ausgaben zu reagieren hat.

In Kapitel 5.5 wird als ein Anwendungsbeispiel die Suche nach Informationen in großen Datenbeständen, dem klassischen Data Mining-Verfahren, beschrieben.

## 5.5 Data Mining

Mit dem Begriff Data Mining wird die Anwendung von statistischen Methoden auf große Datenbestände beschrieben. In Abgrenzung zum KDD-Prozess ist hier die eigentliche Auswertung der Daten gemeint.

Die Verfahren werden hauptsächlich bei strukturierten Daten angewendet, die beispielsweise über Datenbanken angeboten werden. Nach einer Aufbereitung der Textdaten können die Verfahren ebenso zum Text Mining eingesetzt werden.

Im Ergebnis erzeugen die Analyseverfahren ein Modell zur Identifikation von Mustern und Zusammenhängen in den Datenbeständen.

Klassischer Weise werden in einem ersten Schritt Ausreißer erkannt. Insbesondere in urbanen Datenbeständen können Ausreißer ein Hinweis auf Anomalien sein. Beispielsweise kann hierzu der von Breunig [68] vorgeschlagene Algorithmus „Local Outlier Factor“ eingesetzt werden. Dieser Ansatz basiert auf dem Gedanken, dass Ausreißer einen geringeren Abstand zu anderen Objekten mit einer dichteren Lage haben.

In einem weiteren Schritt werden die Daten in Teilmengen (Clustern) zusammengefasst. Die Idee hinter dem **Clustering** ist, dass ähnliche Objekte in Clustern zusammengefasst werden können. Objekte eines Clusters weisen ähnliche

Eigenschaften auf und können von anderen Clustern abgegrenzt werden. Der Vorteil des maschinellen Lernens (unüberwachtes Lernen) besteht darin, dass keine Ergebnisse oder eine Beispielmenge zum Vergleich vorgegeben werden müssen. Gängige Algorithmen zum Clustering können in partitionierende, hierarchische und dichte-basierte Clusterbildung unterteilt werden. Beim partitionierenden Vorgehen wird eine gegebene Menge von Objekten in zuvor festgelegte k-Cluster zerlegt. Vorteilhaft ist, dass jedes Objekt in höchstens einem Cluster enthalten ist. Dem hierarchischen Ansatz nach weisen Cluster zueinander eine geringe Distanz (Unterscheidung) auf. Der Vorteil liegt in der Bildung von Hierarchien zwischen den Clustern. Dichte-basierte Verfahren betrachten Objekte, die dicht beieinanderliegen und von Bereichen mit geringer Dichte getrennt werden. Die einzelnen Cluster können mit Namen beschrieben und als Klassen deklariert werden [39].

Eine weitere Möglichkeit zur Ordnung ist eine **Klassifikation**. Ähnlich wie beim Clustering werden die Objekte in Gruppen, in diesem Fall in Klassen, eingeteilt. In Abgrenzung zum Clustering werden die Gruppen (Klassen) vorgegeben und die Objekte den Klassen zugeordnet. Über Trainingsdaten muss das System angeleitet werden (überwachtes Lernen). Eingesetzt werden Trainingsmengen von Daten mit bekannten Klassenmerkmalen der Objekte, um ein Vergleichsebenen zu schaffen. Neben einer eindeutigen Zuordnung zu einer Klasse, kann die Zugehörigkeit zu einer Klasse ebenso als eine Wahrscheinlichkeit beschrieben werden [39].

Die numerische Vorhersage (**Regressionsanalyse**) berechnet ausgehend von einem gegebenen Datensatz mit Funktionswerten zukünftige Werte. Eine bekannte Methode zur numerischen Vorhersage ist die Lineare Regression. Im Gegensatz zur

diskreten Vorhersage durch eine Klassifizierung, können Vorhersagen in einem großen numerischen Wertebereich getroffen werden [39].

Abschließend wird die **Assoziationsanalyse** beschrieben, um Beziehungen zwischen den einzelnen Objekten zu identifizieren. Mit den Informationen können zukünftige Trends (Verhalten neuer Datensätze) vorhergesagt werden [39].

Eine Spezifikation des Data Mining ist die Auswertung von Textdokumenten. In Kapitel 5.6 wird die Suche nach Informationen in Textdaten als Text Mining vorgestellt.

## 5.6 Text Mining

Text Mining umfasst eine Zusammenstellung von Methoden und Algorithmen zur Auswertung von Textdokumenten. Texte weisen keine feste Strukturierung (vgl. Kapitel 3.1.2) wie Daten aus Datenbanken auf, obwohl Texte durch natürliche Strukturen wie Grammatik organisiert sind [39].

Mit Texten kann Wissen festgehalten, beibehalten und wiederbearbeitet werden. Ungefähr 80 Prozent der Daten liegen in einer textuellen Form vor. Zur Nutzung der Informationen aus Texten wurde die eigenständige Disziplin „Text Mining“ geprägt, welche auch als „Knowledge Discovery from Text“ beschrieben wird. Im Grunde beruht Text Mining auf einem interdisziplinären Ansatz, bestehend aus maschinellem Lernen, statistischen Verfahren, Information Retrieval (Informationsrückgewinnung) und Natural Language Processing (maschinelle Interpretation von menschlicher Sprache) [69]. Als Ergebnis werden die zu erfragenden Informationen aus den Texten

extrahiert. Diese können beispielsweise Schlüsselwörter, die Häufigkeit von einzelnen Wörtern oder eine erste Kategorisierung der Texte sein [39].

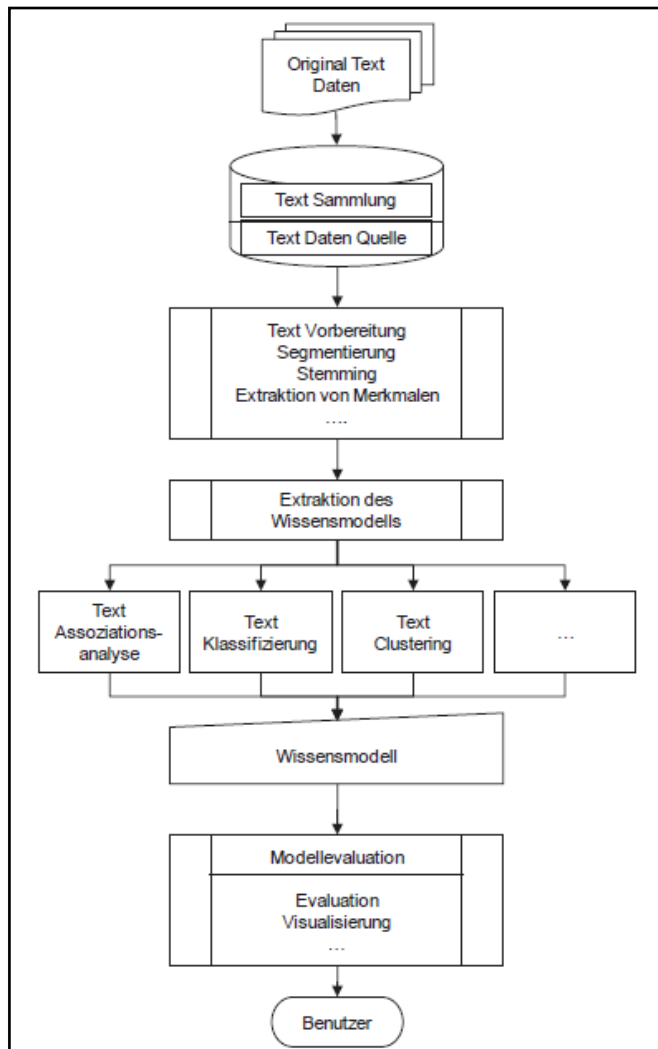


Abbildung 14: Ein idealtypischer Text Mining Ablauf [69]

Bevor die Texte ausgewertet werden können ist eine linguistische bzw. statistische Vorbereitung erforderlich. Hierzu zählt die Identifikation uninteressanter Wörter (stop words) oder die Reduzierung der Wörter auf ihren Wortstamm (Stemming) (vgl.

Kapitel 5.6.2) [39,69]. Die eigentliche Analyse findet mit dem reduzierten Wortbestand statt. Zur Analyse werden bekannte Verfahren, wie beispielweise Klassifikation oder Clustering, angewendet [39]. Ein idealtypischer Text Mining Ablauf wird in Abbildung 14 dargestellt. Als erster Ablaufschritt wird in Kapitel 5.6.1 die Auswahl der Textdaten beschrieben. Im nächsten Schritt folgen die in Kapitel 5.6.2 beschriebenen Maßnahmen zur Vorbereitung der Textdaten.

### **5.6.1 Auswahl der Textdaten**

Dateien für das Text Mining können einerseits vollständige Textdokumente andererseits einzelne Textfragmente sein. Datenquellen sind beispielsweise einzelne Dateien, Datenbanken, Webseiten oder soziale Medien. Unterschieden werden statische Dateien, die ihre ursprüngliche Struktur beibehalten und dynamische Dateien, die im Zeitverlauf bearbeitet werden können. Für eine spätere Verarbeitung werden die Textdaten in Datenbanken oder Document Warehouse gespeichert. Ein Document Warehouse ist einem Data Warehouse ähnlich, wie in Kapitel 4.2.2 vorgestellt.

### **5.6.2 Vorbereitung der Textdaten**

Vor der Auswertung ist eine Vorbereitung der Datenbestände notwendig. Die vorbereitenden Maßnahmen werden in drei Hauptkategorien unterteilt: Vorbereitungsarbeiten (Preparatory), automatische Spracherkennung (Natural Language Processing) und aufgabenorientierte Verfahren. Die Verfahren werden in Abbildung 15 hierarchisch dargestellt und in den folgenden Unterkapiteln näher beschrieben.



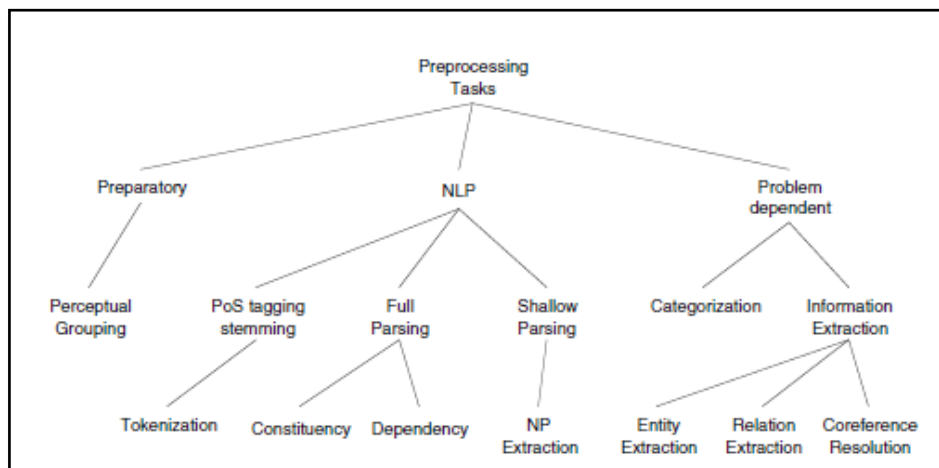


Abbildung 15: Mögliche Verfahren zur Vorverarbeitung von Textdaten [70]

Mit der Vorbereitung (Preprocessing) werden die unstrukturierten Rohdateien in ein strukturiertes, für Maschinen lesbareres Format umgewandelt. Die einzelnen Textbestandteile (Überschriften, Absätze, Zeilen, Tabellen) werden separiert und bereits formatierte Abschnitte (z. B. Angaben zum Autor (<Author>) oder Titel (<Title>)) identifiziert. Die wesentliche Vorbereitung erfolgt mit einer automatischen Spracherkennung. Die einzelnen Wörter werden separiert und als Token dargestellt (Tokenization) und einer Wortart zugeordnet (Part of Speech (POS) Tagging). Alternativ kann eine Syntaxanalyse (Parsing) durchgeführt werden. Das Ergebnis ist eine erste Aufbereitung der Texte, die noch keine Relevanz für die Aufgabenstellung haben [70].

#### *Stoppwörter und Stopplisten*

Texte enthalten eine Reihe von irrelevanten Wörtern, die für eine Auswertung eine geringere Bedeutung haben. Diese Wörter werden als Stoppwörter (stop words) bezeichnet, hierunter fallen Artikel, Konjunktionen oder Präpositionen. Zudem können Wörter mit einer geringeren Häufigkeit zu den Stoppwörtern gezählt werden.

Zur Identifikation können Bibliotheken mit Stoppwörtern eingesetzt werden. Die Reduzierung der Texte erhöht die inhaltliche Qualität der Auswertung und reduziert den Rechenaufwand [69].

#### *Tokenisierung*

Bevor der gesamte Text ausgewertet werden kann, wird der Text für Maschinen in kleine Teile zerlegt. Die einzelnen Satzstücke werden in Token zerlegt und definierte Satzzeichen (beispielsweise Punkte, Kommata) entfernt. Bei Token handelt es sich in der Regel um einzelne Wörter, die in ihrer Gesamtheit einen Satz ergeben können. Dennoch werden Token nicht als Wörter bezeichnet, sondern als eine Einheit bestehend aus einer Folge von Zeichen. Zusammengesetzt können Token eine semantische Einheit im Sinne eines Wortes bilden. Die Zerlegung in einzelne Wörter ist zunächst ein einfaches Vorgehen, da einzelne Token durch ein Leerzeichen getrennt und Satzzeichen ignoriert werden. Schwieriger wird es bei zusammenhängen Wortformen (beispielsweise eine Trennung durch Bindestriche) oder bei Interpunktionen innerhalb von Sätzen [69].

#### *Part-of-speech Tagging*

Mit dem Part-of-speech Tagging werden die Wörter bzw. die Tokens mit einer syntaktischen Wortart markiert. Diese können Markierungen als Artikel, Präpositionen, Substantive, Verben oder Adjektive sein. Insbesondere bei der Identifikation von Personen, Orten oder Organisationen kann das Tagging hilfreich sein. Als Referenzwerte werden Lexika oder manuell erstellte Datenbanken genutzt. Auf diese Weise kann ebenfalls geregelt werden, wie mit Doppeldeutigkeiten verfahren wird. Eine Doppeldeutigkeit kann mit folgenden Sätzen verdeutlicht werden: „Er verteidigt die Burg“ und „Die verteidigte Burg fiel“. Das Wort „verteidigt“ tritt einmal als Verb und einmal als Adjektiv auf [69]. Am Ende des Tagging entsteht

ein Korpus bestehend aus den Wörtern und den zugeordneten Tags [69,70]. In Abbildung 16 wird ein solcher Korpus dargestellt.

A Whole New Mind \$16.47 US, is a 2005 release from Daniel H. Pink that covers creative thinking and other aspects of success.
A_DT Whole_JJ New_NNP Mind_NNP \$_PUNC 16.47_CD US_NNP ,_PUNC is_VBZ a_DT 2005_CD release_NN from_IN Daniel_NNP H_NNP Pink_NNP that_WDT covers_VBZ creative_JJ thinking_NN and_CC other_JJ aspects_NNS of_IN success_NN .PUNC
Tags: CC = coordinating conjunction, CD = cardinal number, DT = determiner, IN = preposition/subordinating participle conjunction, JJ = adjective, NN = noun (singular or mass), NNS = proper noun (singular), NNP = noun (plural), PUNC = punctuation, VBZ = verb (3rd person, singular, present), WDT = wh-determiner

Abbildung 16: Ein Text vor und nach dem POS-Tagging [69]

#### *Lemmatisierung und Stemming*

Eine weitere Möglichkeit zur Vorbereitung von Texten folgt dem Ansatz, dass ein einzelnes Wort viele Formen annehmen kann und Wörter auf ihren Wortstamm reduziert werden können. Als Voraussetzung müssen die Wörter mit dem POS-Tagging markiert werden [69].

Werden Verben auf ihre Grundform und Nomen auf ihre singuläre Form zurückgeführt, wird dieses als Lemmatisierung beschrieben. Die neue Form des Wortes ist ein Lemma. Mit dieser Methode können die Suchergebnisse erweitert werden. Zudem ist bei einer Abfrage die korrekte Eingabe des Wortes von geringerer Relevanz [69]. Die Reduzierung des Wortes auf den Wortstamm wird als Stemming benannt. Wörter mit gleicher oder ähnlicher Bedeutung werden auf einen gemeinsamen Stamm reduziert. Das Ergebnis sind keine tatsächlichen Wörter der jeweiligen Sprache (Abbildung 17) [69].

<i>Word form</i>	<i>Lemma</i>	<i>Stem</i>
Lemma	lemma	lemm
Lemmas	lemma	lemm
Lemmatizer	lemmatizer	lemm
Lemmatizers	lemmatizer	lemm
Lemmatized	lemmatize	lemm
Lemmatizing	lemmatize	lemm

Abbildung 17: Vergleich von Lemmatisierung und Stemming [69]

#### *Textkategorisierung und Informationsextraktion*

Zur maschinellen Verarbeitung werden die Dokumente als eine Sammlung von sogenannten Features repräsentiert. Die Zuordnung der Feature kann entweder durch eine Textkategorisierung oder eine Informationsextraktion erfolgen. Beide Verfahren benötigen zur eindeutigen Beschreibung der Feature eine vorherige Ordnung mit dem beschriebenen POS-Tagging. Aufgrund der vielzähligen Wörter entsteht am Ende ein ‚n-dimensionaler‘ Featureraum [70].

Anschließend werden die Textdaten mehreren Kategorien zugeordnet. In der Regel werden die Kategorien manuell erstellt und haben einen kleineren Umfang. Zudem können Hierarchien berücksichtigt werden.

Eine weitere Strukturierung ist über die wesentlichen Informationen möglich. Die Informationsextraktion strukturiert die relevanten Informationen in einer Datenbank. Diese Zuordnung von Kategorien und wesentlichen Informationen ist eine weitere Form des „tagging“ (Abbildung 18) [70].

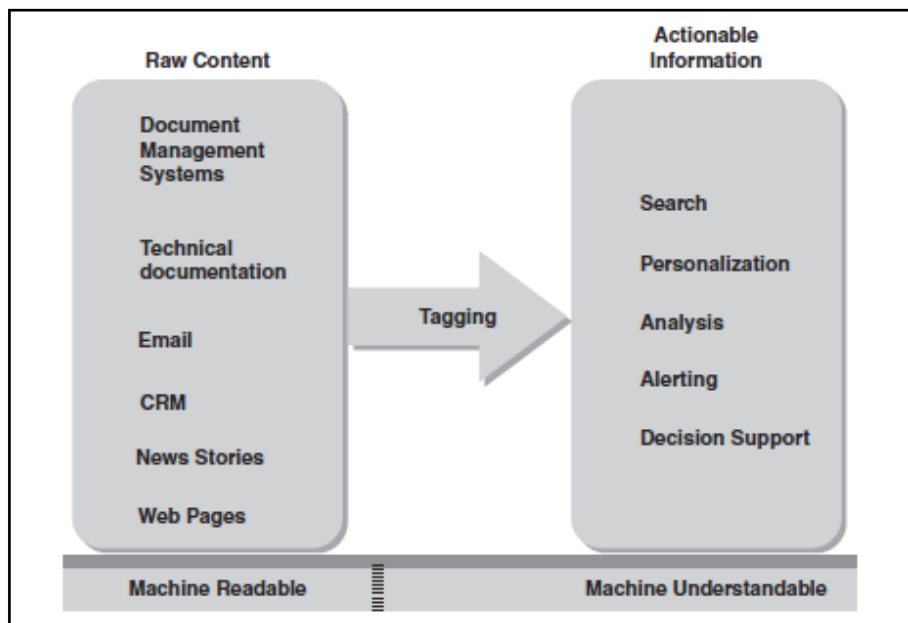


Abbildung 18: Maschinenlesbare Textdaten werden zu maschinenverständlichen Daten [70]

### 5.6.3 Transformation der Textdaten

Im folgenden Kapitel werden das Vektorraummodell und das Gewichtungsverfahren von Termen als weitere Möglichkeiten zur Vorbereitung von Textdaten beschrieben.

#### *Vektorraummodell und Gewichtungsverfahren*

Beim Vektorraummodell werden die Dokumente in einem hochdimensionalen Raum als Vektoren repräsentiert. Jeder Vektor entspricht dabei dem Gewicht seines Terms. Oftmals handelt es sich bei einem Term um ein Wort. Die Gewichtung wird über die Häufigkeit des Terms in einem Dokument definiert. Diese Vektoren können einerseits zur Repräsentation der Dokumente, andererseits zum Auffinden von Dokumenten eingesetzt werden. Ob ein Dokument, also ein Vektor, relevant für den Vergleich oder die Suchanfrage ist, erfolgt durch eine Bestimmung des Winkels zwischen einem Anfragevektor und dem Dokumentenvektor (Kosinus-Ähnlichkeitsmaß) (siehe Abbildung 19) [69].

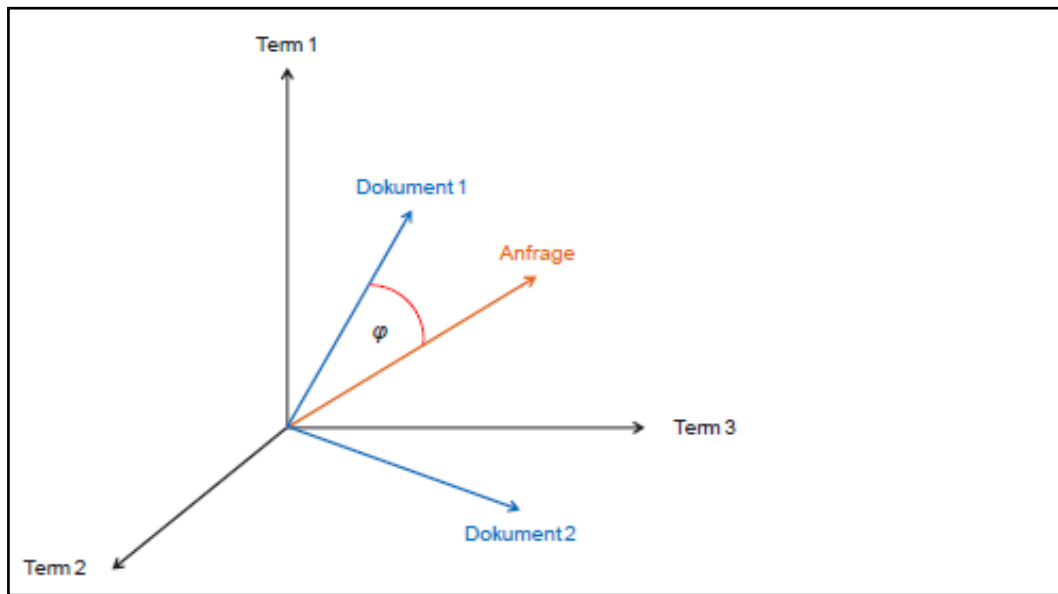


Abbildung 19: Darstellung von Textdaten im Vektorraum [69]

Das Konzept der Gewichtung von Termen findet ebenfalls Anwendung bei der Darstellung der Vorkommenshäufigkeit. Der Grundgedanke bei diesem Konzept liegt darin, dass die Häufigkeit eines Terms hinsichtlich der Signifikanz eine unterschiedliche Bedeutung hat. Dabei wird der Ansatz verfolgt, dass häufig und besonders selten vorkommende Wörter nahezu keine oder eine geringe Bedeutung für die Repräsentation des Textes haben. Abbildung 20 zeigt die exemplarische Verteilung einer Worthäufigkeit in Form einer Normalverteilung. Die Relevanz eines Wortes nach oben und nach unten wird durch die Linien C und D dargestellt [69].

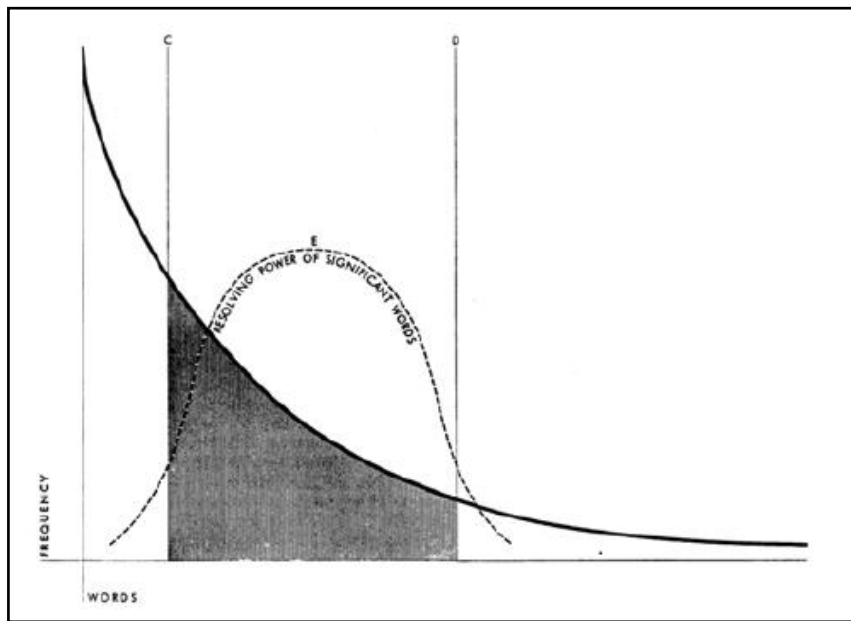


Abbildung 20: Darstellung einer Worthäufigkeit [69]

Für die Gewichtung eines Dokumentes bzw. von Textdaten wird das Wort bzw. die Wörter mit einer mittleren Verteilung als Repräsentant genutzt. Eine weitere Möglichkeit ist die Bestimmung über die inverse Dokumentenfrequenz. Hierbei werden Terme verwendet, die nur in wenigen Textdaten vorkommen, um somit die Daten voneinander separieren zu können [69].

Mit den vorgenannten Verfahren werden unstrukturierte Textdaten in strukturierte Daten umgewandelt und können mit den Verfahren des Data Mining (z. B. Clusteranalysen, Assoziationsanalysen, Klassifikationsmethoden) ausgewertet werden.

#### **5.6.4 Anwendungsbeispiel**

Ausgehend von der Idee, dass eine Mitteilung über den Microblogging Dienst Twitter auf eine Gefahrenlage hindeuten kann, wird im Folgenden eine Möglichkeit zur Anomalieerkennung nach Thom et al. [71] mit Tweets beschrieben.

Twitter zählt zu den berühmtesten Microblogging Diensten weltweit. Täglich werden bis zu 230 Millionen Beiträge (Tweets) von über 100 Millionen Nutzern erzeugt. Jeder Tweet ist weltweit öffentlich und kann georeferenziert geteilt werden. Neben der Darstellung über die Twitter-Webseite, können die Tweets über eine API in Echtzeit abgerufen werden. Neben dem Autor, dem Zeitstempel und dem Beitrag selbst, werden verlinkte Personen, Links auf Webseiten, der Ort sowie die Geokoordinaten bzw. der Ortsname dargestellt [71].



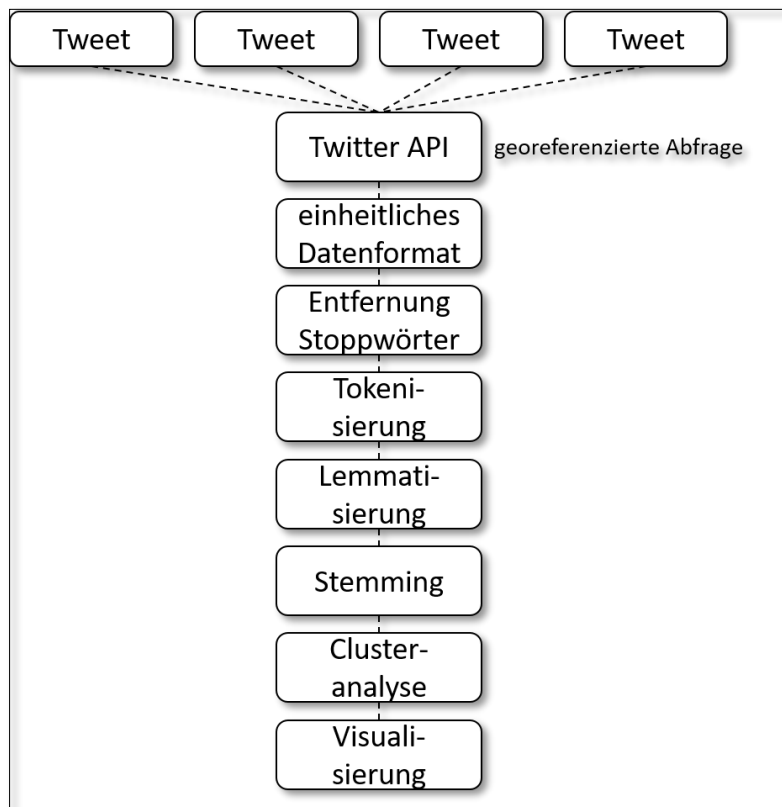


Abbildung 21: Anwendung von Text Mining Verfahren zur Anomalieerkennung

Für die Auswertung (Abbildung 21) werden nur Tweets berücksichtigt, die auf ein Bezugsgebiet (z. B. eine Stadt) angewendet werden können. Zur Selektion wird bei der Abfrage ein Geocode samt Umkreisradius definiert. Die reduzierten Tweets werden in einem lokalen Text Warehouse gespeichert. In der Datenbank werden die Nutzer-ID und der Beitrag getrennt voneinander abgelegt [71].

Anschließend werden die Textdaten aufbereitet. Die Stoppwörter werden entfernt, die einzelnen Wörter in Token getrennt und eine Lemmatisierung sowie Stemming durchgeführt. Mit diesen Verfahren können die Tweets in einem ersten Schritt im Umfang und auf wenige Begriffe reduziert werden [71].

Thom et al. [71] beschreiben, dass auf diese Weise fünf bis zehn Begriffe für die Analyse bestehen bleiben. Die einzelnen Begriffe stehen dabei nicht in einer Abhängigkeit zu dem ursprünglichen Tweet, sondern werden isoliert betrachtet. Dabei beinhaltet jeder Term neben dem Wort die Nutzer-ID sowie die Zeit- und Geoinformationen [71].

Im folgenden Schritt wird eine Clusteranalyse durchgeführt, um die Wörter zu bündeln. Im Wesentlichen ordnet der von Thom et al. [71] vorgestellte Algorithmus jedem Term ein Cluster zu. Jedem Cluster wird die zum Term zugehörige Geoinformation beigefügt. Das Cluster wird von der Geoinformation und dem Wort gekennzeichnet. Weitere erfasste Terme werden mit den vorhandenen Clustern abgeglichen. Der Term wird dem nächstgelegenen Cluster zugeordnet. Terme ohne Zuordnung zu einem bestehenden Cluster werden einem neuen Cluster zugeordnet [71].

Abschließend werden die Abstände der Cluster zueinander ermittelt. Ein größerer Abstand kann auf eine Anomalie hindeuten. Bei einer nicht signifikanten Abgrenzung der Cluster werden die vorhandenen Cluster in zwei neue Cluster aufgeteilt und die Abstände zueinander abgeglichen [71].

Abbildung 22 zeigt die Schritte zur Erkennung von Anomalien. Nach einer Tokenisierung und der Entfernung von Stoppwörtern mit der Programmbibliothek Lucence (Apache) (1) werden die Terme in Clustern (2) zusammengefasst. Zur Visualisierung wird eine mehrschichtige Darstellung (Layer) im Kartenformat verwendet. Die Cluster werden in einzelnen Layern (3) präsentiert. Die mehrfache Nennung von Termen wird als Anomalien in einer Gesamtkarte (4) präsentiert [71].

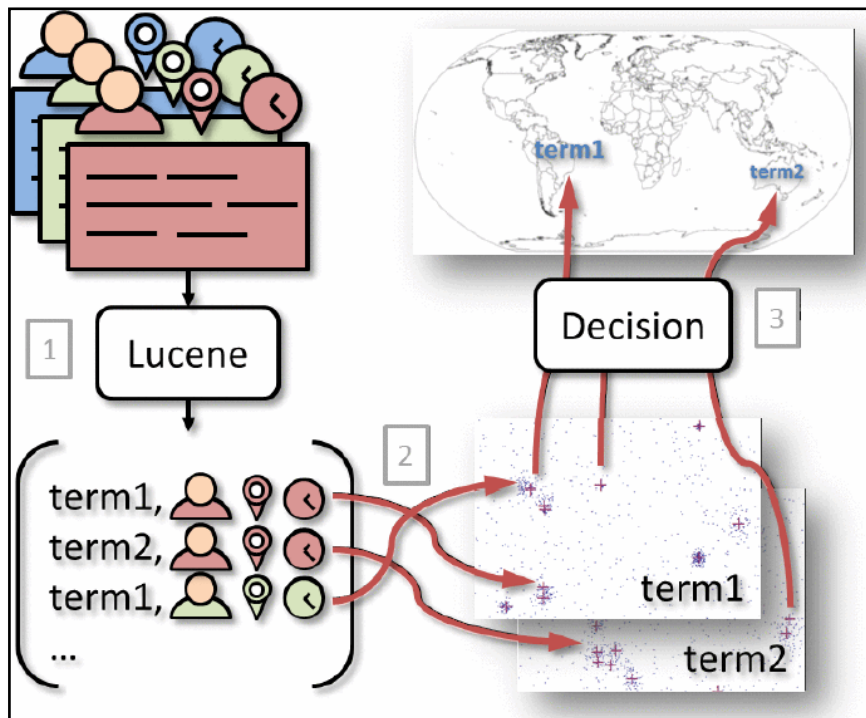


Abbildung 22: Anomalieerkennung mit Tweets [71]

Mit dem vorgestellten Verfahren haben Thom et al. [71] das Nutzungsverhalten von Twitter-Nutzern während eines Erdbebenereignisses am 23. August 2011 vor der US-Ostküste ausgewertet. Die Ergebnisse werden in Abbildung 23 dargestellt. Neben einer minutengenauen Analyse des Nutzungsverhaltens (linke Darstellung), werden die Schlagwörter (Terme) in den ersten Minuten des Ereignisses (mittlere Darstellung) und eine Suche nach Schadensmeldungen (rechte Darstellung) visualisiert [71].



Abbildung 23: Auswertung eines Erdbebens von der US-Ostküste [71]

Neben Text Mining ist Web Mining eine weitere Spezialisierung des Data Mining zur automatisierten Generierung von Informationen aus dem Internet. In Kapitel 5.7 wird das Web Mining vorgestellt.

## 5.7 Web Mining

Das Web Mining wird unterteilt in die Suche nach Texten und medialen Daten (Web Content Mining) und in die Untersuchung des Verhaltens der Nutzer (Web Usage Mining).

Beim Web Usage Mining werden die Protokolldateien der Server durchsucht, um das Online-Verhalten der Nutzer zu analysieren. Diese Informationen sind für das Online-Marketing interessant, da das Kaufverhalten der Nutzer analysiert werden kann.

Die Suche nach Inhalten von Webseiten erfolgt mit Methoden des sogenannten Information Retrieval. Unter diesem Begriff werden Verfahren zur automatisierten Suche nach Texten und medialen Dateien (komplexen Inhalten) beschrieben. Anders als beim reinen Data Mining werden keine neuen Erkenntnisse gesucht, sondern bestehende Daten zugänglich gemacht. Anschließend können die Daten zur Informationsgewinnung mit Methoden des Data Mining ausgewertet werden [72].

Gewonnen werden die Daten mit Crawlern, die sowohl das öffentliche Internet als auch das Deep Web und das Dark Web durchsuchen können.

## 6 Visualisierung: Technische Möglichkeiten

Ein Leitstand für eine Sensorik-Leitstelle muss technisch in der Lage sein, hochkomplexe und nicht vorhersagbare Gefahrenlagen zu erfassen und darzustellen. Verschiedene Informationsquellen sind in einem einheitlichen System zu visualisieren. Zudem muss die Zusammenarbeit verschiedener Akteure zur Gefahrenabwehr möglich sein. In Kapitel 6.1 wird an einem Praxisbeispiel ein interaktiver Gefahrenabwehr-Leitstand als eine mögliche Architektur beschrieben. Das in Kapitel 6.2 vorgestellte digitale Stadtmodell, als eine neuartige Präsentationsform von urbanen Daten, ergänzt die Beschreibungen zur Visualisierung von urbanen Daten aus Kapitel 4.2.5.

### 6.1 Interaktiver Gefahrenabwehr-Leitstand

Eine anwendungsorientierte Architektur eines Leitstandes für die Gefahrenabwehr wurde von Kowalczyk et al. [64] vorgestellt. Das System „PORTprotect“<sup>10</sup> basiert auf unterschiedlichen Arbeitsplatztypen: mehreren Einzelarbeitsplätzen (Desktop),

---

<sup>10</sup> Weiterführende Informationen zur technischen Umsetzung können hier abgerufen werden: <https://www.wps.de/portfolio-items/portprotect/?portfolioID=11881>

einem Planungstisch (Touch-Tisch), mehreren mobilen Anwendungen (Tablet) und einer Lagerdarstellung (Video-Wall) (Abbildung 24) [64].

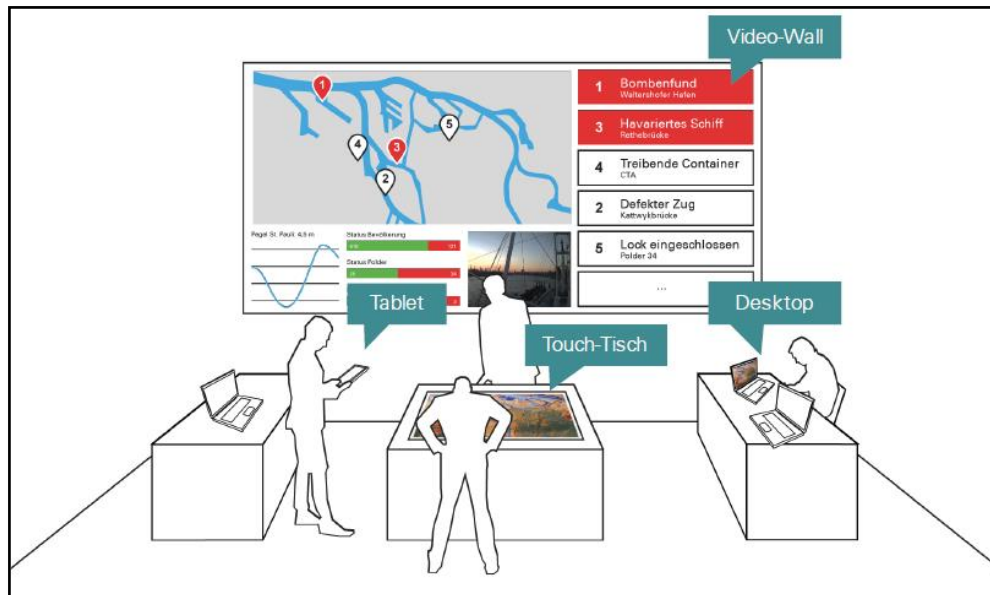


Abbildung 24: Unterschiedliche Arbeitsplatztypen im System PORTprotect [73]

Der Einzelarbeitsplatz ist für die Stabsmitarbeiter zur Erfassung und Bearbeitung der Lagemeldungen vorgesehen. Die Darstellung erfolgt georeferenziert in einer digitalen Karte. Die digitale Karte wird allen Akteuren über eine Leinwand (Video-Wall) angezeigt. Zentraler Bestandteil der Architektur ist der Planungstisch<sup>11</sup> (Touch-Tisch). Technisch realisiert wird der Touch-Tisch mit einem hochauflösenden Bildschirm samt Touchfunktion [64]. In Abbildung 25 wird die Touchfunktion und in Abbildung 26 eine Lagerdarstellung mit einer überlagerten Überflutungskarte dargestellt.

<sup>11</sup> Weiterführende Informationen sowie Videoaufnahmen des Planungstisches unter:  
<https://www.wps.de/portfolio-items/geneal/?portfolioID=18967>

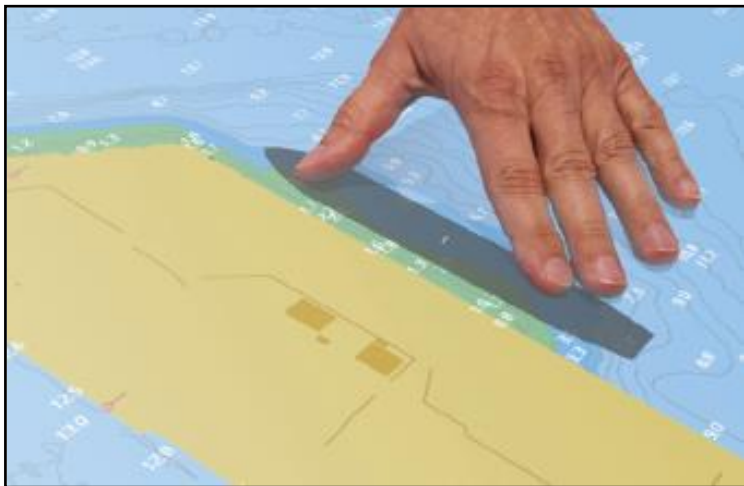


Abbildung 25: Verschieben eines Schiffmodells mit der Touchfunktion [73]



Abbildung 26: Darstellung einer Lagekarte mit einem überlagerten Überflutungslayer [64]

In regelmäßigen Lagebesprechungen kommen die Stabsmitglieder am Planungstisch zur Lagebesprechung zusammen. Ziel dieser Besprechungen ist eine Abstimmung über die nächsten Maßnahmen zur Gefahrenabwehr. Im Vordergrund steht die Interaktion der Teilnehmer, die nicht durch technische Schwierigkeiten beeinflusst werden soll. Ein wesentliches Merkmal des Planungstisches ist die intuitive Bedienung mit einfachen Interaktionsmodulen wie Schiebreglern und Drehknöpfen [64].



Technisch realisiert wird das System als eine sogenannte Service-Architektur. Bei dieser Architektur können verteilte Systeme miteinander verbunden werden. Ein wesentliches Merkmal ist die Bündelung von bestehenden Systemen, sogenannten Microservices, in einem übergeordneten Dienst. Ein Microservice für das System „PORTprotect“ ist beispielsweise der Pegelservice des Sturmflutwarndienstes, über den Prognosedaten der zu erwartenden Flutkurve übermittelt werden. Die Daten werden im Lagebild zusammengefasst und an der Leinwand dargestellt [64].

Das System „PORTprotect“ zur integrierten Darstellung von verschiedenen Informationen kann für eine Sensorik-Leitstelle der Gefahrenabwehr grundlegend übernommen werden. Besonders die Förderung der Gruppenarbeit an einem Planungstisch ist ein wesentliches Erfolgskriterium für die Bewältigung von Gefahrenlagen. Die unterschiedlichen Daten und Informationen zur Lage werden auf eine Leinwand (Video-Wall) projiziert. Zur Erfassung und Visualisierung der Daten wird ein Geoinformationssystem genutzt. Klassischer Weise werden die Informationen zweidimensional auf einer Karte dargestellt. Abbildung 27 stellt ein Anwendungsbeispiel zur Darstellung von Hochwassergefahren und -risikokarten dar.

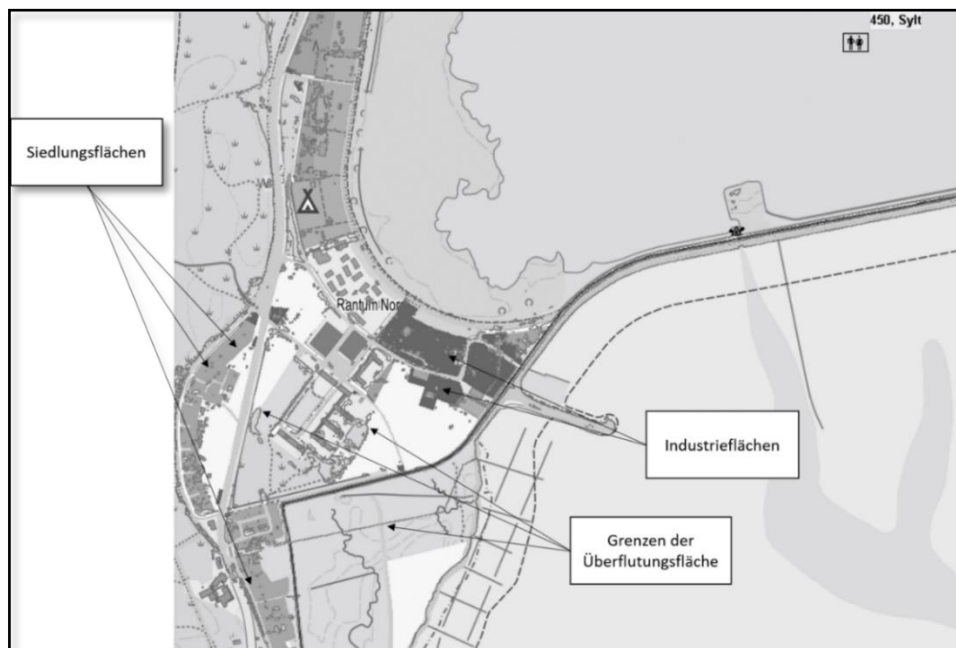


Abbildung 27: Darstellung von Hochwassergefahren (bearbeitete Darstellung) [74]

Bei dieser Form der Darstellung von Daten und Informationen steht nicht der Interaktionsaspekt im Vordergrund, sondern die räumliche Darstellung von Informationen. Zudem kann immer nur ein Ausschnitt detailliert dargestellt werden.

In Kapitel 6.2 wird ein digitales Stadtmodell zur Darstellung von urbanen Daten beschrieben.

## 6.2 Digitales Stadtmodell

Urbane Daten sind hochkomplexe Daten, die einen räumlich-zeitlichen Bezug zu einer Stadt haben. Die Daten werden an unterschiedlichen Orten durch statische Sensoren (technische Sensoren) oder durch dynamische Sensoren (soziale Sensoren) erfasst. Erst durch eine Interaktion mit diesen Daten können Informationen gewonnen werden [62]. Geoinformationssysteme alleine können diese Interaktion mit den

Anwendern nur bedingt ermöglichen. Eine transparente und intuitive Interaktion wird durch den Einsatz digitaler 3D-Stadtmodelle möglich [26].

Ein interaktives und digitales Stadtmodell wurde am MIT Media Lab mit dem Namen „CityScope“ entwickelt. Im Wesentlichen besteht das System aus einem Modelltisch und sogenannten Daten-Blöcken (Abbildung 28). Über einen Projektor werden Informationen auf die Daten-Blöcke, beispielsweise Flächennutzungspläne, projiziert.



Abbildung 28: CityScope Modell in Tongji, Shanghai [75]

Die wesentlichen Komponenten des System sind mehrere interaktive Datenblöcke (1), ein Modellierungstisch (2), ein oder mehrere Tischscanner (3), ein oder mehrere Control Screen (4), ein oder mehrere Tischprojektoren (5) sowie ein oder mehrere Daten-Screen(s) (6) (Abbildung 29) [76].

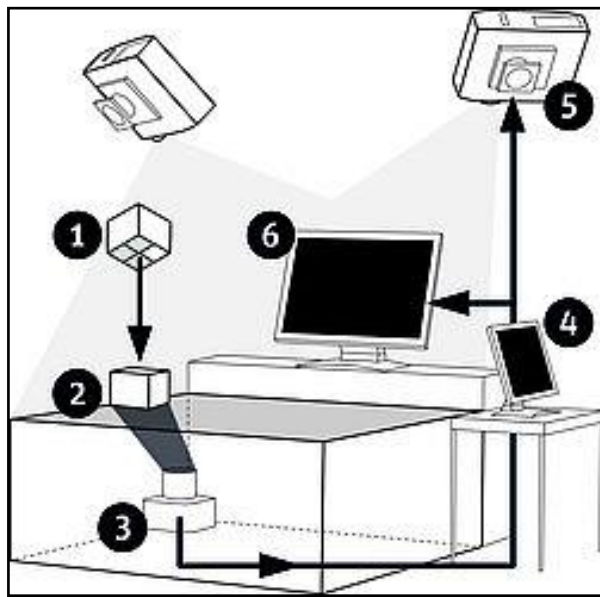


Abbildung 29: Systemkomponenten des CityScope [76]

Mit diesem System können komplexe urbane Daten auf transparente Weise dargestellt werden. Zudem ermöglicht diese Form der Darstellung eine Interaktion mit den Daten und liefert ein zeitnahes visuelles Feedback. Entwickelt wurde das System für eine moderne Form der Stadtentwicklung in Zusammenarbeit verschiedener Akteure [76,77].

Ein Anwendungsbeispiel ist das Kooperationsprojekt „Finding Places“ der HafenCity Universität und der Stadt Hamburg. Ziel des Projektes war die Identifikation von freien Flächen zur Errichtung von Flüchtlingsunterkünften. Für die Suche wurde mit dem CityScope-System ein Modelltisch geschaffen, auf dem Ausschnitte der Hamburger Stadt projiziert wurden. Basierend auf der Projektion konnten mögliche freie Flächen identifiziert werden. Zur Schaffung einer dreidimensionalen Struktur wurden Daten-Steine ergänzt. Auf diese Weise konnten freie Flächen mit Daten-Steinen markiert werden (Abbildung 30) [77,78].



Abbildung 30: Anwendung des CityScopes Systems im Projekt FindingPlaces (Bild: Walter Schießwohl) [79]

Die Tischscanner haben die identifizierten Flächen erfasst und rechnergestützt mit weiteren Informationen zur Realisierbarkeit angereichert [77,78].

An diesem Beispiel wird deutlich, dass urbane Daten zur Gefahrenabwehr dargestellt werden können. Zur Erkennung von Gefahrenlagen in einer Stadt wäre es denkbar, dass einzelne Mitteilungen aus sozialen Netzwerken (z. B. Twitter) als Punkte oder mehrere Tweets als eine Heatmap auf ein dreidimensionales Modell projiziert werden.

## 7 Diskussion

Urbane Daten sind eine Erscheinungsform von Big Data. Im Wesentlichen trifft diese Beschreibung zu, jedoch ist eine differenziertere Betrachtung von urbanen Daten notwendig. Urbane Daten sind nicht ausschließlich eine große Menge an Daten aus einer oder ähnlichen Quellen, sondern Datensätze aus unterschiedlichsten Quellen und Dateiformaten (vgl. Kapitel 2.2).

Psyllidis [S. 26] [80] unterscheidet urbane Daten in konventionelle und soziale Daten. **Konventionelle Daten** werden auf Anfrage als Information zu einem aktuellen Sachstand erzeugt (z. B. Einwohnerzahlen, Gästezahlen, Einwohner pro Quadratmeter). Das Ergebnis sind qualitativ hochwertige Datensätze mit einer hohen Glaubhaftigkeit sowie einer strukturierten Sortierung [80]. Die Datensätze haben eine feste Reihenfolge und definierte Attribute [66]. Diese Form der Strukturierung ermöglicht eine Analyse ohne aufwändige Vorbereitung der Datensätze. Die Schwelle diese Daten zu nutzen ist dadurch geringer. Ein deutlicher Nachteil konventioneller Daten ist das geringere Aktualisierungsintervall.

**Soziale Daten** werden -im Gegensatz zu starren konventionellen Daten- von der Bevölkerung spontan und in kürzeren Intervallen erzeugt. Sie sind gekennzeichnet

durch zeitlich-räumliche Merkmale und spiegeln subjektive Wahrnehmungen der Umgebung wieder. Die Datengenerierung unterliegt keinem einheitlichen Standard, sodass die Qualität aufgrund unterschiedlicher Datenformate geringer ist. Unterschiedliche (wissenschaftliche) Projekte nähern sich sozialen Daten mit verschiedenen Methoden der Erhebung und Auswertung [12,71]. Die fehlende Strukturierung der Daten in einem einheitlichen Format erfordert umfangreiche vorbereitende Maßnahmen (vgl. Kapitel 4.2.3) und erschwert die Auswertung der Daten. Die Glaubhaftigkeit der sozialen Daten wird im Vergleich zu konventionellen Daten niedriger eingestuft, weil sie über die Bürger subjektiv dargestellt werden [80].

Der Nutzen dieser Art von Daten für die Gefahrenabwehr gleicht den Vorbereitungsaufwand jedoch deutlich aus. Die Stärke der sozialen Daten -als Repräsentanten des kollektiven Wissens- ist die Darstellung der aktuellen Lebensumstände der Bevölkerung [10]. Soziale Daten haben das Potential unmittelbare Daten zu sein, die das soziale Geschehen -und somit auch Gefahrenlagen- in nahezu Echtzeit erfassen und weiterleiten können. Die zeitliche Dimension ist jedoch abhängig vom individuellen Verhalten zu einer Gefahrenlage.

Wie in Kapitel 2.2 beschrieben bestehen urbane Daten im Wesentlichen aus Daten von Sensoren, Open Data-Portalen oder internen Datenbeständen. Sensordaten haben sowohl eine bürgerliche als auch eine technische Quelle. Technische Quellen umfassen von Maschinen erzeugte Datensätze. Die von den Bürgern erzeugten Daten sind im Wesentlichen Mitteilungen in sozialen Netzwerken. Aktuelle Open Data-Portale beinhalten zum gegenwärtigen Zeitpunkt hauptsächlich prozessgesteuerte Datensätze als Repräsentation eines aktuellen Sachstands [80]. Interne Datenbestände können sowohl maschinell erzeugte Daten als auch

prozessgesteuerte Datensätze beinhalten. Soziale Daten, im Sinne einer bürgerlichen Quelle, sind bei diesen Datenbeständen nicht zu erwarten.

Insbesondere das systematische Erfassen von technischen sowie sozialen Sensordaten erfordert neue technische Konventionen. Für eine effektive Verarbeitung der Daten sind bereits bei der Entwicklung einer Smart City allgemeinbekannte Datenformate und -Schnittstellen zu beschreiben. Eine einheitliche Infrastruktur -im Sinne einer smarten Anwendung in der Gefahrenabwehr- erleichtert die Kollaboration mit den Bürgern und die Integration von Daten aus Datenbanken (z. B. Open Data-Portalen). Ein konkretes System, welches die unterschiedlichen Datenquellen integriert, wurde bislang nicht entwickelt [81].

Urbanen Daten sind große Datenbestände und werden mit den sogenannten fünf V, volume (Menge), velocity (Geschwindigkeit), variety (Vielfältigkeit), veracity (Glaubwürdigkeit) und value (Aussagekraft), charakterisiert (vgl. Kapitel 2.3) [33,34]. Diese Definition ist für den Umgang mit urbanen Daten jedoch nicht umfassend aussagekräftig. Psyllidis [S. 63-75] [80] hat in Anlehnung an Devlin [82] für urbane Daten eine Charakterisierung mit den Begrifflichkeiten diversity (Vielfalt), scale (Ausmaß), timeliness (Aktualität), structure (Datenstruktur), spatiotemporal resolution (räumlich-zeitlicher Bezug), semantic expressiveness (semantische Bedeutung), representativeness (Repräsentativität) und veracity (Aussagekraft) vorgestellt. Die Vielfalt (diversity) beschreibt, dass die Daten aus unterschiedlichen Quellen gewonnen werden. Dabei unterscheiden sich die Daten in der Qualität, Datenformat und -struktur sowie im Inhalt. Das Ausmaß (scale) steht im Zusammenhang mit der Datenmenge. Die Aktualität (timeliness) beschreibt die Aktualisierungsintervalle. Die Aktualität der Daten wird mit den fünf V nicht direkt



erfasst, lediglich die high-velocity (Geschwindigkeit) deutet auf eine Schnelligkeit der Daten hin. Die Datenstruktur (structure) beschreibt, dass urbane Daten in einem strukturierten, unstrukturierten oder semi-strukturierten Datenformat vorliegen können (vgl. Kapitel 3.1). Der räumlich-zeitliche Bezug (spatiotemporal resolution) verdeutlicht, dass urbane Daten neben Geokoordinaten ebenfalls Metadaten zu Adressen (Straßenname, Hausnummer, Postleitzahl, Bundesland, Stadt, Ortsteil) beinhalten können. Ein weiteres Merkmal ist die semantische Bedeutung der Daten. Insbesondere Daten aus sozialen Medien beinhalten Aussagen zu den Lebenssituationen der Bürger oder eine Zustands- bzw. Situationsbeschreibung einer Stadt. Mit der Repräsentativität (representativeness) wird erfasst, dass die Daten unter Umständen nur von einem Teil der Bevölkerung bereitgestellt werden. Insbesondere ist zu erwarten, dass Daten in sozialen Netzwerken von jüngeren Menschen erzeugt werden [80,82].

Zur Erkennung von Gefahrenlagen in urbanen Daten sind neben der Verfügbarkeit, die Relevanz und das Aktualisierungsintervall wesentliche Merkmale. Die Verfügbarkeit der Daten wird mit den „fünf V´s“ oder den zuvor beschriebenen neun Merkmalen nicht erfasst. Das Aktualisierungsintervall entspricht dem Merkmal timeliness (Aktualität). Die Merkmale semantische Aussagekraft (semantic expressiveness), Repräsentativität (representativeness) und Glaubwürdigkeit (veracity) können zur Beschreibung der Relevanz betrachtet werden.

### *Verfügbarkeit*

Die Darstellung der Datenökonomie am Beispiel der Stadt Hamburg wird in Kapitel 3.2 beschrieben. Die Auflistung von exemplarischen Datensätzen in Tabelle 2 veranschaulicht, dass zum jetzigen Zeitpunkt diverse Datensätze in einer Stadt

verfügbar sind. Den vorwiegenden Anteil haben prozessuale und maschinell erzeugte Datensätze. Einen kleineren Anteil bilden Daten aus sozialen Netzwerken. An dieser Stelle wird auf den orientierenden Charakter der Auflistung hingewiesen. Obwohl nutzergenerierte Daten in der Auflistung einen kleineren Anteil bilden, werden gerade diese zur Darstellung einer Echtzeit-Stadt zukünftig einen wesentlichen Anteil der Datensätze ausmachen [10,13,26]. Hierunter fallen Geodaten (Volunteered Geographic Information), Data Science Communities (wie luftdaten.info) und Geräte/Sensoren des Internet of Things. Ein treibender Faktor ist die zunehmende Datafizierung von unterschiedlichen Aspekten des öffentlichen Lebens [83].

Daten aus sozialen Medien sind eine weitere Quelle, jedoch oftmals nur lückenhaft auslesbar. Hinzu kommt, dass nicht alle soziale Medien eine öffentliche Schnittstelle anbieten. Eben die Erfassung dieser zeitlich-räumlich variierenden Echtzeitdaten erfordert neue technische Systeme zur Erfassung [26].

#### *Aktualisierungsintervall*

Verschiedene Quellen beschreiben, dass mit urbanen (sozialen) Daten die Prozesse einer Stadt in nahezu Echtzeit abgebildet werden können [13,24]. Zur Erkennung von Gefahrenlagen im städtischen Raum sind Echtzeitdaten mit kurzen Aktualisierungsintervallen erforderlich. Die in Tabelle 2 (vgl. Kapitel 3.2) aufgelisteten Datensätze zeigen, dass urbane Daten sowohl kurze (Live-Daten, minütlich) als auch lange (stündlich, monatlich, jährlich) Aktualisierungsintervalle umfassen. Prozessgesteuerte Daten weisen erwartungsgemäß geringere Aktualisierungsintervalle gegenüber maschinell erzeugten Datensätzen sowie sozialen Daten auf [80]. Ein wesentliches Merkmal von Sensorik-Leitstellen ist die Echtzeitauswertung von urbanen Daten zur Erkennung von Gefahrenlagen. Die aufgelisteten Datensätze (vgl.

Tabelle 2) verdeutlichen jedoch, dass urbane Daten die potentiellen Möglichkeiten zum jetzigen Zeitpunkt nicht in Echtzeit repräsentieren.

### *Relevanz*

Sensordaten werden mit standardisierten technischen Verfahren erhoben. Die messtechnischen Verfahren haben eine hohe Glaubwürdigkeit. Die Aussagekraft sozialer Daten ist geringer. Im Wesentlichen bestehen die Datensätze aus Wertepaaren (Zeitpunkt, Zeitraum, geographische Informationen, Messwerte). Eine aussagekräftige Auswertung kann nur erfolgen, wenn die Daten in einem Kontext betrachtet werden [80]. Beispielsweise sind Daten von Luftmessung (z. B. Feinstaub) erst im Vergleich mit Grenzwerten aussagekräftig.

Soziale urbane Daten, im engeren Sinne Daten aus sozialen Netzwerken, entstehen in einem Kontext zur Umwelt. Die Datensätze werden als Reaktion auf ein Ereignis ad hoc mit begleitenden Informationen generiert (hohe Semantik). Ein Nachteil der sozialen Daten ist, dass die sozialen Medien von bestimmten Bevölkerungsgruppen genutzt werden. Es ist zu erwarten, dass vorwiegend jüngere und technikaffine Bevölkerungsgruppen die neuen Medien nutzen. Ein weiterer Nachteil ist, dass die Glaubwürdigkeit der sozialen Daten schwieriger zu bewerten ist [80].

Eine wesentliche Herausforderung ist die Bewertung der Semantik und der Aussagekraft der Mitteilungen aus sozialen Netzwerken. In Kapitel 5.6 wird die Clusteranalyse als ein mögliches Verfahren zur Auswertung von Textdaten vorgestellt. Die unterschiedlichen Begriffe werden einzelnen Clustern zugewiesen. Das Resultat sind gewichtete Cluster (siehe Abbildung 23) [71]. Neben dem vorgestellten Verfahren werden in der Literatur lexikonbasierte Methoden und

Verfahren des maschinellen Lernens als geeignete Verfahren zur Bewertung der Semantik beschrieben [15,16,84].

Lexikonbasierte Verfahren beruhen auf einem Wörterbuch (Lexikon) mit eindeutigen Schlüsselwörtern, die beispielsweise eine Gefahrenlage beschreiben. Ähnliche Verfahren sind aus der Sentiment-Analyse bekannt [15,16]. Bei diesem Verfahren werden den Wörtern zur Charakterisierung unterschiedliche Gewichtungen zugewiesen. Die Gewichtung kann sowohl automatisiert als auch manuell erfolgen [16]. Die Wertung der Texte erfolgt nach Kategorien (positiv, negativ, objektiv) [15] oder einer Skaleneinteilung [16]. Herausforderungen sind die Erstellung eines geeigneten Wörterbuches und die Gewichtung der Begriffe [84]. Zur Erstellung eines Wörterbuches für die Gefahrenabwehr könnten das Wörterbuch für Bevölkerungsschutz und Katastrophenhilfe der Ständigen Konferenz für Katastrophenvorsorge und Katastrophenschutz oder das BBK-Glossar verwendet werden. Den Begrifflichkeiten der beiden Publikationen können hohe Gewichtungen zugeordnet werden. Eine weitere Möglichkeit zur Erstellung ist die Auswertung von vergangenen Mitteilungen zu Gefahrenlagen in sozialen Netzwerken [85]. Auf diese Weise können eindeutigen Sätze, Wörter und Muster von realen Gefahrenlagen identifiziert werden [46,86]. Die Auswahl von positiven Mitteilungen kann nach den drei folgenden Heuristiken durchgeführt werden [86]:

1. Mitteilungen mit einer eindeutigen Geoinformation oder einer räumlichen Beschreibung zum Notfallort sind positiv
2. Mitteilungen mit einem Bild oder einer Videoaufnahme von der Gefahrenlage sind positiv
3. Werden Mitteilungen weitergeleitet, ist die ursprüngliche Mitteilung positiv

Die Speicherung der positiven Mitteilungen erfolgt in einer Datenbank. Xu et al. [S. 6] [86] nennen mögliche Kategorien zur Ordnung der Mitteilungen: die aktuelle Lagesituation, Schadensinformationen, Einsatz (Gefahrenabwehr), zukünftige Situation und die Hilfebedürftigkeit der Bevölkerung.

Einerseits können die positiven Mitteilungen für die Erstellung eines Wörterbuches, andererseits als Trainingssatz für maschinelles Lernen eingesetzt werden. Mit den Verfahren des überwachten Lernens wird ein Klassifizierungsalgorithmus mit positiven Mitteilungen zur Erkennung von Gefahrenlagen angelernt [85]. In der Literatur werden sowohl Support Vector Machine [85] als auch Naive Bayes [87] als geeignete Klassifikatoren des maschinellen Lernens beschrieben.

Eine grundsätzliche Anforderung an eine Sensorik-Leitstelle zur Erkennung von Gefahrenlagen im urbanen Raum ist die Einbindung von verschiedenen Datenquellen (vgl. Kapitel 4.1) [51,52]. Die vorliegende Masterthesis konzentriert sich auf die Auswertung von textuellen Daten im Sinne von Mitteilungen der Bürger in sozialen Netzwerken, da diese einen ersten Hinweis auf eine Gefahrenlage geben können. Eine umfangreiche Lagebeurteilung erfordert jedoch die Einbeziehung von unterschiedlichsten Informationen. Zum einen können erste Hinweise auf eine Gefahrenlage bestätigt werden. Zum anderen wird sichergestellt, dass Gefahrenlagen erkannt werden, die nicht über die sozialen Netzwerke oder technische Sensoren wahrgenommen werden. Dies bedeutet, dass nicht nur soziale und technischen (im Sinne von Messeinheiten) Sensoren einbezogen werden, sondern im systemischen Aufbau die Möglichkeit besteht, weitere Sensoren („Everything as a sensor“) einzubeziehen. Hierzu ist das in Kapitel 4.2.1 beschriebene „Sensing layer“ mit offenen Programmierschnittstellen zu gestalten. Beispielsweise

können zusätzlich Videodaten aus Überwachungskameras zur Erkennung von Gefahrenlagen ausgewertet werden. Xu et al. [46] werteten zur Erkennung von Gefahrenlagen sowohl Videodaten von Überwachungskameras als auch Mitteilungen aus sozialen Netzwerken aus. Die Integration weiterer Sensoren erfordert wiederum angepasste Algorithmen zur Auswertung. Künstliche neuronale Netzwerke, wie beispielsweise Convolutional Neural Networks, Self-organizing maps (vgl. Kapitel 5.3) oder Support Vector Machine, haben das Potential unterschiedlichste Datenquellen zu erfassen und auszuwerten [67,88,89].

Die vorliegende Masterthesis zeigt neue Wege der Erkennung von Gefahrenlagen in urbanen Daten auf. Durch die Freie und Hansestadt Hamburg [19] wurden mit der Einführung und Umsetzung des Transparenzgesetzes die technischen sowie rechtlichen Grundvoraussetzungen für eine smarte Anwendung in der Gefahrenabwehr (smartGAb) geschaffen. Offen bleibt in welchem Ausmaß die vorhandenen Daten in der praktischen Anwendung zur Erkennung von Gefahrenlagen genutzt werden können. Zudem ist zum jetzigen Zeitpunkt unklar wie hoch die Bereitschaft der Bürger zur Mitwirkung an smartGAb ist. Die technische Umsetzbarkeit sowie die Bereitschaft der Bürger zur aktiven Mitwirkung muss der Ausgangspunkt für weitere wissenschaftliche Studien sein. Abschließend kann festgehalten werden, dass die Stadt Hamburg ein geeigneter deutscher Pilotstandort für smartGAb ist.

In Kapitel 8 werden die wesentlichen Aussagen und Ergebnisse zusammengefasst und offene Forschungsaspekte aufgezeigt.

## 8 Fazit

Mit der vorliegenden Arbeit wurde aufgezeigt, dass mit einer algorithmischen Auswertung von Daten aus sozialen Netzwerken, Sensornetzwerken und Open Data-Portalen Gefahrenlagen im urbanen Raum erkannt werden können.

Zur Beschreibung der im urbanen Raum erzeugten Datensätze wurden urbane Daten vorgestellt. Es konnte gezeigt werden, dass urbane Daten im Wesentlichen der Definition von Big Data entsprechen und mit den fünf „V“ beschrieben werden können, jedoch wegen der diversen Quellen und Datentypen eine aussagekräftige Definition notwendig ist. Hierzu wurde die Behauptung, dass zur algorithmischen Erkennung von Gefahrenlagen eine Konzentration auf die Verfügbarkeit, die Relevanz und das Aktualisierungsintervall ausreichend ist, mit den Charaktermerkmalen Vielfalt, Ausmaß, Aktualität, Datenstruktur, räumlich-zeitlicher Bezug, semantische Bedeutung, Repräsentativität und Aussagekraft diskutiert.

Am Beispiel der Freien und Hansestadt Hamburg wurde die Datenökonomie einer Stadt vorgestellt. Die Stadt Hamburg hat mit der Einführung des Hamburgischen Transparenzgesetzes eine Vorreiterrolle im Bereich der offenen Daten eingenommen. Obwohl erste prozessgesteuerte und maschinell erzeugte Datensätze

vorliegen, ist die Bereitstellung von sozialen urbanen Daten erweiterungsfähig. Eine exemplarische Auflistung der unterschiedlichen Datensätze zeigt hauptsächlich prozessuale und maschinelle Daten auf. Bei den sozialen Netzwerken wurden zunächst ein Microblog-Dienst und zwei Fotocommunities als geeignete Ressourcen für eine Echtzeitauswertung ausgewählt. Mit der vorliegenden Arbeit konnte die Datenökonomie einer Stadt nicht vollumfänglich beurteilt werden. Die Auflistung hat einen orientierenden Charakter und ist nicht repräsentativ.

Ebenso wird deutlich, dass die Erhebung von urbanen Daten neue (technische) Konventionen erfordert. Weitere Vorhaben sollten die Informations- und Datenflüsse in einer Stadt mit dem Ziel der Erhebung von Echtzeitdaten untersuchen.

Die Anwendungsbeispiele aus New York City und Rio de Janeiro zeigen, dass ausgewählte Prozesse einer Stadt auf Daten basierend repräsentiert werden können und urbane Daten zur frühzeitigen Erkennung Gefahrenlagen geeignet sind. Die beiden Praxisbeispiele sind im Grunde Sensorik-Leitstellen, die die Mensch-Computer-Interaktion in Alltags- und Krisensituationen ermöglichen. Eine Herausforderung für Sensorik-Leitstellen ist die Erfassung von komplexen, zusammengesetzten räumlich-zeitlichen Daten (urbane Daten). Zudem muss eine kooperative Auswertung der Daten sowie eine Visualisierung möglich sein.

Eine wesentliche Erkenntnis ist, dass zur Datenexploration sowohl maschinelle Methoden als auch digitale Stadtmodelle mit Interaktions- und Verzerrungstechniken notwendig sind. Das vorgestellte System einer Sensorik-Leitstelle beruht auf einer Vier-Schichten-Architektur und ist ein Vorschlag für eine Referenzarchitektur zur Erfassung von urbanen Daten. Zur Überprüfung der Nutzbarkeit sind praktische Versuche mit realen Daten oder Trainingsdaten notwendig.



Zur maschinellen Auswertung von urbanen Daten sind verschiedene Methoden der künstlichen Intelligenz und des maschinellen Lernens notwendig. Für die vorliegende Arbeit wurde angenommen, dass Mitteilungen von Bürgern in sozialen Netzwerken ein erster Hinweis auf Gefahrenlagen sein können. Die Methoden des Text Minings können zur Auswertung von Textdaten (z. B. Mitteilungen in sozialen Netzwerken) eingesetzt werden. Ein wesentlicher Teil der Analysearbeit liegt in der Vorbereitung der Daten. Als ein technisches Verfahren wurde die Clusteranalyse an einem Anwendungsbeispiel vorgestellt. In der Literatur werden zudem lexikonbasierte Methoden und durch überwachtetes Lernen angelernete künstliche neuronale Netzwerke (Feature Recurrent Neuronal Network und Convolutional Neuronal Network) als geeignete Verfahren beschrieben. Insbesondere sind diese Verfahren aus der semantischen Analyse bekannt. Urbane Daten bestehen jedoch nicht nur aus Textdateien. Hinzu kommen beispielsweise statistische Daten, Beiträge aus sozialen Medien, Bilder und Videoaufzeichnungen. Es ist zu zeigen, dass die Verfahren ebenfalls für eine (semantische) Analyse unterschiedlicher Datentypen zur Erkennung von Gefahrenlagen einzusetzen sind. Ein praktischer Versuch war im Rahmen der theoretischen Arbeit nicht umsetzbar.

Urbane Daten, insbesondere die der Bürger, besitzen das Potential zukünftig aussagekräftige Indikatoren der Gefahrenabwehr zu werden. Ihre Bearbeitung von der Erhebung bis zur Visualisierung birgt vielseitige Herausforderungen und Chancen, die im Sinne der Sicherheit der Bürger weiter im Fokus wissenschaftlicher Arbeit bleiben muss. Als ein letzter Aspekt wurde deutlich, dass die Bürger aktiv zu ihrer Sicherheit beitragen können und dazu weitere Forschungsarbeiten -z. B. Akzeptanz der Bereitstellung persönlicher Daten zur Gefahrenabwehr- anzuregen sind.

Trotz der positiven Effekte einer Sensorik-Leitstelle durch das Nutzen urbaner Daten -verbessertes Erkennen und Beurteilen von Gefahrenlagen im urbanen Raum- wird an dieser Stelle darauf hingewiesen, dass im Umgang mit urbanen Daten datenschutzrechtliche Aspekte nicht unbeachtet bleiben dürfen. Eine differenzierte Betrachtung war im Rahmen dieser Arbeit nicht möglich und unabdingbar in weiteren Forschungen zu bewerten.

Diese Masterthesis zum Thema Erkennung von Gefahrenlagen zeigt auf, dass eine Sensorik-Leitstelle -als eine smarte Anwendung in der Gefahrenabwehr (smartGAb)- in der Lage ist, die bisherige Digitalisierung der Freien und Hansestadt Hamburg richtungsweisend zu ergänzen und zu optimieren.

# Abkürzungsverzeichnis

API	Programmierschnittstelle (application programming interface)
App	Anwendungssoftware (Application software)
BBK	Bundesamt für Bevölkerungsschutz und Katastrophenhilfe
CSV	Kommaseparation (comma-separated values)
GIS	Geoinformationssystem
HTML	Hypertext Markup Language
IBM	International Business Machines Corporation
IoT	Internet of Things
JSON	JavaScript Object Notation
KDD	Knowledge Discovery in Databases
LSTM	Long short-termin memory
MIT	Massachusetts Institute of Technology
NINA	Warn-App des Bundesamtes für Bevölkerungsschutz und Katastrophenhilfe
PDF	Portable Document Format
POS	Part-of-speech
smartGAb	Smarte Anwendung der Gefahrenabwehr
URL	Uniform Resource Locator
vfdb	Vereinigung zur Förderung des Deutschen Brandschutzes e. V.
VGI	Volunteered Geographic Information

# Abbildungsverzeichnis

Abbildung 1: Sechs Dimensionen und dazugehörige Handlungsfelder einer Smart City nach Giffinger et al. [22].....	19
Abbildung 2: Nutzungsverhalten der Fotoplattform "Flickr" durch Touristen [24] ..	21
Abbildung 3: Urbane Daten - Datenquellen in Smart Cities.....	23
Abbildung 4: 14 Kategorien des Transparenzportals der Stadt Hamburg [48] .....	33
Abbildung 5: Datenformate im Transparenzportal der Stadt Hamburg [48] .....	34
Abbildung 6: Häufige Suchanfragen im Transparenzportal (Mai 2018) [49].....	35
Abbildung 7: Schichtenarchitektur für eine Sensorik-Leitstelle .....	43
Abbildung 8: Data Mining - ein interdisziplinärer Ansatz nach Cleve und Lämmel [39] .....	46
Abbildung 9: Ablauf einer Informationsgewinnung aus großen Datenbeständen [39] .....	47
Abbildung 10: Angepasster Data Mining Ablauf zur Anomalieerkennung [14] .....	48
Abbildung 11: Exemplarischer Aufbau eines künstlichen Netzwerkes [65] .....	55
Abbildung 12: Aufbau eines vorwärtsgerichtete Netzwerkes [66] .....	56
Abbildung 13: Verschiedene Netztopologien: a) vorwärts verkettet, b) mit Abkürzungen, c) mit Rückkopplungen, d) mit lateralen Verbindungen, e) voll vernetzt [65] .....	56
Abbildung 14: Ein idealtypischer Text Mining Ablauf [69] .....	63
Abbildung 15: Mögliche Verfahren zur Vorverarbeitung von Textdaten [70] .....	65
Abbildung 16: Ein Text vor und nach dem POS-Tagging [69] .....	67
Abbildung 17: Vergleich von Lemmatisierung und Stemming [69].....	68

---

Abbildung 18: Maschinenlesbare Textdaten werden zu maschinenverständlichen Daten [70].....	69
Abbildung 19: Darstellung von Textdaten im Vektorraum [69] .....	70
Abbildung 20: Darstellung einer Worthäufigkeit [69] .....	71
Abbildung 21: Anwendung von Text Mining Verfahren zur Anomalieerkennung ...	73
Abbildung 22: Anomalieerkennung mit Tweets [71].....	75
Abbildung 23: Auswertung eines Erdbebens von der US-Ostküste [71] .....	76
Abbildung 24: Unterschiedliche Arbeitsplatztypen im System PORTprotect [73] ....	79
Abbildung 25: Verschieben eines Schiffmodells mit der Touchfunktion [73] .....	80
Abbildung 26: Darstellung einer Lagekarte mit einem überlagerten Überflutungslayer [64].....	80
Abbildung 27: Darstellung von Hochwassergefahren (bearbeitete Darstellung) [74] .	82
Abbildung 28: CityScope Modell in Tongji, Shanghai [75].....	83
Abbildung 29: Systemkomponenten des CityScope [76] .....	84
Abbildung 30: Anwendung des CityScopes Systems im Projekt FindingPlaces (Bild: Walter Schießwohl) [79].....	85

# Tabellenverzeichnis

Tabelle 1: Beispiele für strukturierte, semi- und unstrukturierte Daten [43] .....	29
Tabelle 2: Beispielhafte Datenquellen.....	36
Tabelle 3: Drei Systemkomponenten einer Sensorik-Leitstelle.....	52

# Literaturverzeichnis

1. Nichts geht mehr ohne Vernetzung und Digitalisierung / vfdb-Jahresfachtagung in Bremen: Innovation und Tradition gehören zusammen [Internet]. presseportal.de. [zitiert 4. Mai 2018]. Verfügbar unter: <https://www.presseportal.de/pm/126597/3644070>
2. Stroop R, Strickmann B, Horstkötter H, Kuhlbusch T, Hartweg H-R, Kerner T. Smartphone-basierte First-Responder-Alarmierung „Mobile Retter“. Notarzt. 2015;31(05):239–45.
3. Bundesamt für Bevölkerungsschutz und Katastrophenhilfe - Warn-App NINA [Internet]. [zitiert 4. Mai 2018]. Verfügbar unter: [https://www.bbk.bund.de/DE/NINA/Warn-App\\_NINA.html](https://www.bbk.bund.de/DE/NINA/Warn-App_NINA.html)
4. Universität Siegen. EmerGent – Emergency Management in Social Media - Computerunterstützte Gruppenarbeit und Soziale Medien [Internet]. Computerunterstützte Gruppenarbeit und Soziale Medien | Universität Siegen. [zitiert 4. Mai 2018]. Verfügbar unter: <https://www.csw.uni-siegen.de/projekte/emergent/>
5. The City of New York. Road Map for the Digital City Achieving New York City's Digital Future [Internet]. New York, NY; 2011. Verfügbar unter: <http://www.nyc.gov/html/media/media/PDF/90dayreport.pdf>
6. INFORMS. NYPD Domain Awareness System (DAS) [Internet]. INFORMS. [zitiert 18. April 2018]. Verfügbar unter: <https://www.informs.org/Impact/O.R.-Analytics-Success-Stories/NYPD-Domain-Awareness-System-DAS>
7. Deakin M. Intelligent cities as smart providers: CoPs as organizations for developing integrated models of eGovernment Services. Innov Eur J Soc Sci Res. Juni 2012;25(2):115–35.

8. Jaekel M. Smart City wird Realität [Internet]. Wiesbaden: Springer Fachmedien Wiesbaden; 2015 [zitiert 18. Juni 2017]. Verfügbar unter: <http://link.springer.com/10.1007/978-3-658-04455-8>
9. Cukier K, Mayer-Schoenberger V. The rise of big data: How it's changing the way we think about the world. *Foreign Aff.* 2013;92(3):28–40.
10. Goodchild MF. Citizens as Voluntary Sensors: Spatial Data Infrastructure in the World of Web 2.0. *Int J Spat Data Infrastruct Res.* 2007;2(2):24–32.
11. Ganti RK, Ye F, Lei H. Mobile crowdsensing: current state and future challenges. *IEEE Commun Mag.* 2011;49(11).
12. Xu Z, Liu Y, Yen N, Mei L, Luo X, Wei X, u. a. Crowdsourcing based Description of Urban Emergency Events using Social Media Big Data. *IEEE Trans Cloud Comput.* 2016;1–1.
13. Kitchin R. The real-time city? Big data and smart urbanism. *GeoJournal.* 2014;79(1):1–14.
14. Assendorp JP. Masterarbeit - Deep learning for anomaly detection in multivariate time series data - im Studiengang Master of Science Informatik - Hochschule für Angewandte Wissenschaften Hamburg [Internet]. HAW Hamburg; 2017. Verfügbar unter: <http://users.informatik.haw-hamburg.de/~ubicomp/arbeiten/master/assendorp.pdf>
15. Denecke K. Using SentiWordNet for multilingual sentiment analysis. In *IEEE*; 2008 [zitiert 9. August 2018]. S. 507–12. Verfügbar unter: <http://ieeexplore.ieee.org/document/4498370/>
16. Thelwall M, Buckley K, Paltoglou G. Sentiment in Twitter events. *J Am Soc Inf Sci Technol.* Februar 2011;62(2):406–18.
17. I.B.M. Takes 'Smarter Cities' Concept to Rio de Janeiro - The New York Times [Internet]. [zitiert 4. Mai 2018]. Verfügbar unter: <https://www.nytimes.com/2012/03/04/business/ibm-takes-smarter-cities-concept-to-rio-de-janeiro.html>
18. De Longueville B, Smith RS, Luraschi G. Omg, from here, i can see the flames!: a use case of mining location based social networks to acquire spatio-temporal data on forest fires. In *Seattle: ACM*; 2009. S. 73–80.



19. Das Hamburgische Transparenzgesetz [Internet]. hamburg.de. [zitiert 18. Juni 2017]. Verfügbar unter: <http://transparenz.hamburg.de/das-hmbtg/>
20. Freie und Hansestadt Hamburg. Die Digitalisierung der großen Stadt – Chancen für Wirtschaftskraft, Kommunikation und öffentliche Dienstleistungen [Internet]. 2015. Verfügbar unter: <https://www.hamburg.de/contentblob/9260384/ed1cb41d024dbef3f62bd9cd834ca838/data/strategie-deutsch.pdf>
21. Saskia Sassen: «Die Global City ist ein brutaler Ort» [Internet]. 2012 [zitiert 5. Juni 2017]. Verfügbar unter: <https://www.woz.ch/-2ea1>
22. Giffinger R, Fertner C, Kramar H, Kalasek R, Pichler-Milanovic N, Meijers E. Smart cities. Ranking of European medium-sized cities, Final Report, Centre of Regional Science, Vienna UT [Internet]. 2007 [zitiert 6. Mai 2017]. Verfügbar unter: [http://www.smart-cities.eu/download/smart\\_cities\\_final\\_report.pdf](http://www.smart-cities.eu/download/smart_cities_final_report.pdf)
23. Jaekel M, Bronnert K. Moderne Städte sind smart. In: Die digitale Evolution moderner Großstädte [Internet]. Springer Fachmedien Wiesbaden; 2013 [zitiert 5. Juni 2017]. S. 9–20. Verfügbar unter: [http://link.springer.com/chapter/10.1007/978-3-658-00171-1\\_2](http://link.springer.com/chapter/10.1007/978-3-658-00171-1_2)
24. Sagl G, Resch B, Hawelka B, Beinat E. From social sensor data to collective human behaviour patterns: Analysing and visualising spatio-temporal dynamics in urban environments. In: Proceedings of the GI-Forum. 2012. S. 54–63.
25. Meier A, Zimmermann H-D. Digitales Entwicklungsmodell smarterer Städte. In: Smart City. Springer; 2016. S. 3–18.
26. Jin J, Gubbi J, Marusic S, Palaniswami M. An information framework for creating a smart city through internet of things. IEEE Internet Things J. 2014;1(2):112–21.
27. Fröschle N. Engineering von Neuen Bürgerbeteiligungsinstrumenten am Beispiel von Elektromobilität, Feinstaub und Luftreinhaltung. HMD Prax Wirtsch. August 2017;54(4):502–17.
28. ODL-Info - Startseite [Internet]. [zitiert 18. April 2018]. Verfügbar unter: <https://odlinfo.bfs.de/DE/index.html>

29. Hamburger Luftmessnetz – FHH [Internet]. hamburg.de. [zitiert 18. April 2018]. Verfügbar unter: <http://luft.hamburg.de/>
30. Offene Daten > Open Knowledge Foundation Deutschland [Internet]. [zitiert 18. Juni 2017]. Verfügbar unter: <https://okfn.de/themen/offene-daten/>
31. von Lucke J. Innovationsschub durch Open Data, Datenportale und Umsetzungswettbewerbe. na; 2011.
32. Was ist Open data [Internet]. hamburg.de. [zitiert 18. Juni 2017]. Verfügbar unter: <http://transparenz.hamburg.de/was-ist-open-data/>
33. Laney D. 3D data management: Controlling data volume, velocity and variety [Internet]. Stamford: META Group; 2001 S. 4. (META Group research note). Report No.: 949. Verfügbar unter: <https://blogs.gartner.com/douglaney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf>
34. Panneerselvam J, Liu L, Hill R. Introduction to Big Data. In: Akhgar B, Saathoff GB, Arabnia HR, Hill R, Staniforth A, Bayerl PS, Herausgeber. Application of big data for national security: a practitioner's guide to emerging technologies. Waltham: Butterworth-Heinemann; 2015. S. 3–13.
35. Beyer M. Gartner Says Solving'Big Data'Challenge Involves More Than Just Managing Volumes of Data. Gart Arch Orig On. 2011;10.
36. New York City Police Department and Microsoft Partner to Bring Real-Time Crime Prevention and Counterterrorism Technology Solution to Global Law Enforcement Agencies [Internet]. News Center. 2012 [zitiert 4. August 2017]. Verfügbar unter: <https://news.microsoft.com/2012/08/08/new-york-city-police-department-and-microsoft-partner-to-bring-real-time-crime-prevention-and-counterterrorism-technology-solution-to-global-law-enforcement-agencies/>
37. Lu X, Brelsford C. Network Structure and Community Evolution on Twitter: Human Behavior Change in Response to the 2011 Japanese Earthquake and Tsunami. Sci Rep [Internet]. 2015 [zitiert 18. April 2018];4(6773). Verfügbar unter: <http://www.nature.com/articles/srep06773>

38. Cassa CA, Chunara R, Mandl K, Brownstein JS. Twitter as a sentinel in emergency situations: lessons from the Boston marathon explosions. *PLOS Curr Disasters*. 2013;5.
39. Cleve J, Lämmel U. *Data mining*. München: De Gruyter Oldenbourg; 2014. 306 S.
40. Chae J, Thom D, Bosch H, Jang Y, Maciejewski R, Ebert DS, u. a. Spatiotemporal social media analytics for abnormal event detection and examination using seasonal-trend decomposition. In *IEEE; 2012* [zitiert 19. April 2018]. S. 143–52. Verfügbar unter: <http://ieeexplore.ieee.org/document/6400557/>
41. Fischer Y. *Wissensbasierte probabilistische Modellierung für die Situationsanalyse am Beispiel der maritimen Überwachung*. Bd. 23. KIT Scientific Publishing; 2016.
42. Neue Geodaten und Internetnutzung. *Raumforsch Raumordn*. Februar 2013;71(1):83–4.
43. Gibson H. Acquisition and Preparation of Data for OSINT Investigations. In: Akhgar, Babak, Bayerl PS, Sampson F, Herausgeber. *Open source intelligence investigation*. New York, NY: Springer Berlin Heidelberg; 2016.
44. Hamburger Luftmessnetz – Stationsinfos / Datenarchiv – Archivierte Messdaten – FHH [Internet]. [hamburg.de](http://luft.hamburg.de/clp/archivierte-daten/). [zitiert 19. April 2018]. Verfügbar unter: <http://luft.hamburg.de/clp/archivierte-daten/>
45. Renschler CS, Frazier AE, Arendt LA, Cimellaro GP. A framework for defining and measuring resilience at the community scale: The PEOPLES resilience framework. 2010.
46. Xu Z, Mei L, Lu Z, Hu C, Luo X, Zhang H, u. a. Multi-modal Description of Public Security Events using Surveillance and Social Data. *IEEE Trans Big Data* [Internet]. 2017 [zitiert 19. April 2018]; Verfügbar unter: <http://ieeexplore.ieee.org/document/7831493/>
47. Xu Z, Luo X, Liu Y, Choo K-KR, Sugumaran V, Yen N, u. a. From latency, through outbreak, to decline: detecting different states of emergency events using web resources. *IEEE Trans Big Data*. 2016;
48. Mühlbacher C, Schmidhuber L, Hilgers D. Open Government Data: Offene Daten in der Stadt Hamburg. *Verwalt Manag*. 2017;23(6):318–24.

49. Suche nach beliebten Themen [Internet]. hamburg.de. [zitiert 2. Mai 2018]. Verfügbar unter: <http://transparenz.hamburg.de/beliebtethemen/>
50. Murjahn R, Tegtmeyer S. Open Data / Transparenzportal Hamburg – Grundlagen, Umsetzung, Erfahrungen, Auswirkungen. *Zfv – Z Für Geodäsie Geoinformation Landmanagement*. 2016;(5/2016):330–335.
51. Turoff M, Chumer M, Van de Walle B, Yao X. The design of a dynamic emergency response management information system (DERMIS). *JITTA J Inf Technol Theory Appl*. 2004;5(4):1–35.
52. Carver L, Turoff M. Human-computer interaction: the human and computer as a team in emergency management information systems. *Commun ACM*. 2007;50(3):33–8.
53. Goodspeed R. Digital knowledge technologies in planning practice: from black boxes to media for collaborative inquiry. *Plan Theory Pract*. Oktober 2016;17(4):577–600.
54. Herold H, Lurz B, Wohlrab J. *Grundlagen der Informatik*. 2., aktualisierte Aufl. München: Pearson; 2012. 800 S. (IT-Informatik).
55. Shkapenyuk V, Suel T. Design and implementation of a high-performance distributed Web crawler. In *IEEE Comput. Soc*; 2002 [zitiert 22. April 2018]. S. 357–68. Verfügbar unter: <http://ieeexplore.ieee.org/document/994750/>
56. Castrillo-Fernández O. European Publi Sector Information Platform - Topic Report No. 2015 / 10 - Web Scraping: Applications and Tools [Internet]. 2015 [zitiert 22. April 2018]. Verfügbar unter: [https://www.europeandataportal.eu/sites/default/files/2015\\_web\\_scraping\\_applications\\_and\\_tools.pdf](https://www.europeandataportal.eu/sites/default/files/2015_web_scraping_applications_and_tools.pdf)
57. Chaulagain RS, Pandey S, Basnet SR, Shakya S. Cloud Based Web Scraping for Big Data Applications. In *IEEE*; 2017 [zitiert 22. April 2018]. S. 138–43. Verfügbar unter: <http://ieeexplore.ieee.org/document/8118431/>
58. Panneerselvam J, Liu L, Hill R. Requirements and Challenges for Big Data Architectures. In: *Application of Big Data for National Security*. Elsevier; 2015. S. 131–9.

59. Tiwari R, Winters J. The death of strategic plan: questioning the role of strategic plan in self-initiated projects relying on stakeholder collaboration. *Int Plan Stud.* 3. April 2017;22(2):161–71.
60. Bettencourt LMA. The Uses of Big Data in Cities. *Big Data.* März 2014;2(1):12–22.
61. Fayyad U, Piatetsky-Shapiro G, Smyth P. From data mining to knowledge discovery in databases. *AI Mag.* 1996;17(3):37–54.
62. Keim DA. Datenvisualisierung und Data Mining. In: Vortragsfolien zu: 9 Fachtagung Datenbanksysteme in Büro, Technik und Wissenschaft (BTW ' 01). Oldenburg: Deutsche Informatik Akademie (DIA); 2001.
63. Using Data Visualization to Find Insights in Data - The Data Journalism Handbook [Internet]. [zitiert 18. Juni 2017]. Verfügbar unter: [http://datajournalismhandbook.org/1.0/en/understanding\\_data\\_7.html](http://datajournalismhandbook.org/1.0/en/understanding_data_7.html)
64. Kowalczyk M, Züllighoven H, Saxe S, Breitling H. Das Leitstandsystem „PORTprotect“ - Architektur eines Katastrophenschutz-Leitstandes. OBJEKTSpektrum. 2017;03/2017.
65. Boersch I, Heinsohn J, Socher R. Wissensverarbeitung: eine Einführung in die künstliche Intelligenz für Informatiker und Ingenieure. 2. Aufl. München: Elsevier, Spektrum Akad. Verl; 2007. 379 S.
66. Lämmel U, Cleve J. Künstliche Intelligenz: mit 51 Tabellen, 43 Beispielen, 118 Aufgaben, 89 Kontrollfragen und Referatsthemen. 4., aktualisierte Aufl. München: Hanser; 2012. 332 S.
67. Vision API – Analyse von Bildinhalten [Internet]. Google Cloud. [zitiert 25. April 2018]. Verfügbar unter: <https://cloud.google.com/vision/?hl=de>
68. Breunig MM, Kriegel H-P, Ng RT, Sander J. LOF: identifying density-based local outliers. In *ACM*; 2000. S. 93–104.
69. Sharafi A. Knowledge Discovery in Databases [Internet]. Wiesbaden: Springer Fachmedien Wiesbaden; 2013 [zitiert 29. April 2018]. Verfügbar unter: <http://link.springer.com/10.1007/978-3-658-02002-6>

70. Feldman R, Sanger J. The text mining handbook: advanced approaches in analyzing unstructured data [Internet]. Cambridge; New York: Cambridge University Press; 2007 [zitiert 29. April 2018]. Verfügbar unter: <http://www.books24x7.com/marc.asp?bookid=23164>
71. Thom D, Bosch H, Koch S, Worner M, Ertl T. Spatiotemporal anomaly detection through visual analysis of geolocated Twitter messages. In: IEEE Pacific [Internet]. Songdo: IEEE; 2012 [zitiert 30. April 2018]. S. 41–8. Verfügbar unter: <http://ieeexplore.ieee.org/document/6183572/>
72. Manning CD, Raghavan P, Schütze H. Introduction to information retrieval. New York: Cambridge University Press; 2008. 482 S.
73. WPS – WORKPLACE SOLUTIONS GMBH. Visionäre Leitstandentwicklung - Für jeden Arbeitsplatz die richtige Anwendung [Internet]. [zitiert 4. Mai 2018]. Verfügbar unter: <https://files.wps.de/pdf/wps-visionaere-leitstandentwicklung.pdf>
74. Ministerium für Energiewende, Landwirtschaft, Umwelt, Natur und Digitalisierung. Hochwasserkarten [Internet]. [zitiert 4. Mai 2018]. Verfügbar unter: <http://zebis.landsh.de/webauswertung/>
75. Massachusetts Institute of Technology. City Science Lab Shanghai [Internet]. MIT Media Lab. [zitiert 4. Mai 2018]. Verfügbar unter: <https://www.media.mit.edu/projects/city-science-lab-shanghai/overview/>
76. HafenCity Universität Hamburg (HCU). Tools - CityScope [Internet]. [zitiert 4. Mai 2018]. Verfügbar unter: <https://www.hcu-hamburg.de/research/csl/forschung/tools/>
77. Noyman A, Holtz T, Kröger J, Noennig JR, Larson K. Finding Places: HCI Platform for Public Participation in Refugees' Accommodation Process. *Procedia Comput Sci.* 2017;112:2463–72.
78. HafenCity Universität Hamburg. CityScopes | FINDINGPLACES.hamburg [Internet]. [zitiert 4. Mai 2018]. Verfügbar unter: <https://findingplaces.hamburg/hintergrund/so-funktioniert-das-modell.html>
79. Stadtportal hamburg.de. Hamburger Stadtdialog „FindingPlaces“ am 11. Mai [Internet]. hamburg.de. [zitiert 4. Mai 2018]. Verfügbar unter:

- <http://www.hamburg.de/pressearchiv-fhh/5982422/2016-05-04-pr-finding-places-stadtdialog/>
80. Psyllidis A. REVISITING URBAN DYNAMICS THROUGH SOCIAL URBAN DATA: Methods and tools for data integration, visualization, and exploratory analysis to understand the spatiotemporal dynamics of human activity in cities. S.l.: TU DELFT; 2016. 332 S.
  81. da Silva WM, Alvaro A, Tomas GHRP, Afonso RA, Dias KL, Garcia VC. Smart cities software architectures: a survey. In ACM Press; 2013 [zitiert 5. August 2018]. S. 1722. Verfügbar unter: <http://dl.acm.org/citation.cfm?doid=2480362.2480688>
  82. Devlin B. Business unIntelligence: insight and innovation beyond analytics and big data. Basking Ridge, New Jersey: Technics Publications, LLC; 2013. 419 S.
  83. Kishino Y, Takeuchi K, Shirai Y, Naya F, Ueda N. Datafying city: Detecting and accumulating spatio-temporal events by vehicle-mounted sensors. In: 2017 IEEE International Conference on Big Data (Big Data) [Internet]. Boston, MA: IEEE; 2017 [zitiert 3. September 2018]. S. 4098–104. Verfügbar unter: <http://ieeexplore.ieee.org/document/8258428/>
  84. Öztürk N, Ayvaz S. Sentiment analysis on Twitter: A text mining approach to the Syrian refugee crisis. *Telemat Inform.* April 2018;35(1):136–47.
  85. Sakaki T, Okazaki M, Matsuo Y. Tweet Analysis for Real-Time Event Detection and Earthquake Reporting System Development. *IEEE Trans Knowl Data Eng.* April 2013;25(4):919–31.
  86. Xu Z, Zhang H, Sugumaran V, Choo K-KR, Mei L, Zhu Y. Participatory sensing-based semantic and spatial analysis of urban emergency events using mobile social media. *EURASIP J Wirel Commun Netw.* Dezember 2016;2016(1):9.
  87. Bermingham A, Smeaton AF. Classifying sentiment in microblogs: is brevity an advantage? In ACM Press; 2010 [zitiert 9. August 2018]. S. 1833. Verfügbar unter: <http://portal.acm.org/citation.cfm?doid=1871437.1871741>
  88. Snoek CGM, Worring M, van Gemert JC, Geusebroek J-M, Smeulders AWM. The challenge problem for automated detection of 101 semantic concepts in multimedia. In: Proceedings of the 14th annual ACM international conference on Multimedia - MULTIMEDIA '06 [Internet]. Santa Barbara, CA, USA: ACM

Press; 2006 [zitiert 3. September 2018]. S. 421. Verfügbar unter:  
<http://portal.acm.org/citation.cfm?doid=1180639.1180727>

89. Yanagawa A, Chang S-F, Kennedy L, Hsu W. Columbia university's baseline detectors for 374 Iscom semantic visual concepts. Columbia Univ ADVENT Tech Rep. 2007;222–2006.



## Erklärung zur selbstständigen Bearbeitung

*Hiermit versichere ich, dass ich die vorliegende Arbeit ohne fremde Hilfe selbstständig verfasst und nur die angegebenen Hilfsmittel benutzt habe.*



*Hamburg, den 4.9.2018*