

Bachelorarbeit

im Studiengang Computerlinguistik
an der Ludwig- Maximilians- Universität München
Fakultät für Sprach- und Literaturwissenschaften

Aufbau eines semantischen Netzes zum Bereich „Naturkatastrophen“

vorgelegt von
Kateryna Hami

Betreuer:	Prof. Dr. Klaus U. Schulz
Prüfer:	Prof. Dr. Klaus U. Schulz
Bearbeitungszeitraum:	28. März - 05. Juni 2023

Selbstständigkeitserklärung

Ich erkläre hiermit, dass ich die vorliegende Arbeit selbständig angefertigt, alle Zitate als solche kenntlich gemacht sowie alle benutzten Quellen und Hilfsmittel angegeben habe.

München, den 05. Juni 2023

.....
Kateryna Hamii

1 Abstract

Ziel der Arbeit ist die Erstellung einer Ontologie und eines semantischen Netzes für die computergestützte linguistische Analyse von Texten aus dem Themenbereich Naturkatastrophen. Zu den Hauptaufgaben dieser Forschung gehören die Korpusbildung, die Entwicklung von Ausdruckskategorien, die Bildung von Häufigkeitslisten von Wörtern und die Extraktion von domänenspezifischen mehrdeutigen Lexemen. Wichtig ist auch die Zusammenfassung des Fachwortschatzes nach Synonymen und Flexionsvarianten.

Darüber hinaus müssen für den Aufbau einer Ontologie die Beziehungen zwischen den Konzepten definiert werden, und für ein semantisches Netz muss eine Struktur mit hierarchischen Beziehungen entwickelt werden. Dies erfordert eine manuelle Analyse und die Berücksichtigung der Besonderheiten von Naturkatastrophen. Schließlich muss das gesamte semantische Netz mit Hilfe von Identifikationsnummern für Konzepte und Beziehungen formell kodiert werden.

Der beschriebene Ansatz zum Aufbau einer Ontologie und eines semantischen Netzes für die computergestützte linguistische Analyse von Texten im Themenbereich Naturkatastrophen wird es uns ermöglichen, ein effizientes Werkzeug für die Verarbeitung und Analyse von Textinformationen zu erhalten, das den spezifischen Anforderungen dieses Bereichs entspricht.

Danksagung

An dieser Stelle möchte ich mich bei allen bedanken, die mich bei der Erstellung dieser Bachelorarbeit unterstützt und motiviert haben.

An erster Stelle möchte ich mich bei Professor Dr. Klaus U. Schulz bedanken, der meine Bachelorarbeit betreut und begutachtet hat. Ich bedanke mich für Ihre hilfreichen Hinweise und konstruktive Kritik während der Erstellung dieser Arbeit.

Danken möchte ich auch meinem Freund, der mich mit viel Geduld, Interesse und Freundlichkeit unterstützt hat. Ich möchte mich für zahlreiche interessante Diskussionen und Programmierhilfen bedanken, die maßgeblich dazu beigetragen haben, dass diese Bachelorarbeit in ihrer jetzigen Form entstanden ist.

Schließlich möchte ich mich bei meinen Eltern und meiner Gastfamilie bedanken, deren Unterstützung mein Studium ermöglicht hat und die immer ein offenes Ohr für mich hatten.

Inhaltsverzeichnis

1	ABSTRACT	5
2	HINTERGRUND & MOTIVATION	11
2.1	Semantisches Netz zur Erforschung von Naturkatastrophen	11
2.2	Zielsetzung und Forschungsfragen der Arbeit	11
3	THEORIE UND GRUNDLAGEN.....	13
3.1	Einführung in semantische Netze.....	13
3.1.1	Definition von semantischen Netzen und Ontologien	13
3.1.2	Semantische Netze in Datenanalyse und Wissensrepräsentation	17
3.1.3	Anwendungen von semantischen Netzen in verschiedenen Bereichen	18
3.2	Naturkatastrophen und ihre Folgen	18
3.2.1	Ein Überblick über Naturkatastrophen und ihre Auswirkungen auf die Umwelt und die Gesellschaft.....	18
3.2.2	Klassifizierung von Naturkatastrophen und deren Merkmale	19
3.2.3	Ursachen und Prävention von Naturkatastrophen.....	21
4	ERSTELLUNG EINES SEMANTISCHEN NETZES	23
4.1	Erfassung und Definition des fachspezifischen Vokabulars	23
4.2	Einordnung und Strukturierung des Vokabulars	25
4.3	Gesamtkodierung und Erstellung des semantischen Netzes	28
4.4	Evaluierung des semantischen Netztes	31
5	DISKUSSION UND ZUSAMMENFASSUNG	33
5.1	Bewertung der Stärken und Schwächen des semantischen Netzes und der vorgestellten Methoden.....	33
5.2	Zusammenfassung der wichtigsten Ergebnisse und Erkenntnisse	33
6	SCHLUSSFOLGERUNGEN	35
	Ausblick auf zukünftige Forschungsarbeiten und Anwendungsmöglichkeiten	35
7	BIBLIOGRAPHY	36

2 Hintergrund & Motivation

2.1 Semantisches Netz zur Erforschung von Naturkatastrophen

Naturkatastrophen sind Ereignisse, die aufgrund natürlicher Prozesse wie Erdbeben, Stürme, Überschwemmungen oder Waldbrände auftreten und oft weitreichende Auswirkungen auf die Umwelt und die betroffene Bevölkerung haben. In den letzten Jahrzehnten hat die Häufigkeit und Intensität solcher Ereignisse zugenommen, was zu einem wachsenden Interesse an der Verbesserung der Katastrophenbewältigung und dem Schutz von Menschenleben und Eigentum geführt hat.

Angesichts der zunehmenden Zahl von Naturkatastrophen und von Menschen verursachten Unglücken bemühen sich viele Organisationen verstärkt um das Katastrophenmanagement, um neue Lösungen für die Notfallhilfe zu entwickeln. Eine entscheidende Komponente bei der Katastrophenbewältigung ist die schnelle Entdeckung und Integration von Informationen, um effektive Entscheidungen treffen zu können. Hierbei kann die Integration von geografischen Informationen ein wichtiges Wissen im Zusammenhang mit der Katastrophe liefern, die den Fortschritt der Hilfsoperationen erleichtert und eine bessere Schadensbewertung ermöglicht.

In den letzten Jahren hat sich die Ontologie zu einem der wichtigsten Interessensgebiete in der geografischen Informationssystem entwickelt. Ontologien spezifizieren direkt oder indirekt die Zusammensetzung, Struktur und grundlegenden Eigenschaften der vereinfachten Welten, die unsere Modelle darstellen und klären die beabsichtigten Bedeutungen der verwendeten Begriffe. Sie können zur Identifikation und Assoziation von Konzepten in einem bestimmten Bereich, deren Eigenschaften und Beziehungen verwendet werden. Die Darstellung von Wissen durch Ontologien ist eine Lösung, um verstecktes Wissen zu identifizieren und semantische Lücken zu schließen. Ontologie wurde als eine effektive Lösung zur Überwindung von semantischen Problemen erkannt. (H.Bouyerbou, K.Bechkoum, R.Lepage, 2019)

Die Erstellung eines semantischen Webs im Bereich der Naturkatastrophen ist für die wirksame Bewältigung dieser Ereignisse unerlässlich. Durch ein semantisches Web werden alle relevanten Informationen und Ressourcen zusammengeführt, so dass die Entscheidungsträger rasch präzise Maßnahmen ergreifen und die Reaktionszeiten verkürzen können. Dies ist von großer Bedeutung für den wirksamen Schutz von Leben und Eigentum.

2.2 Zielsetzung und Forschungsfragen der Arbeit

Die Forschung im Bereich der Ontologieentwicklung und der computergestützten linguistischen Textanalyse ist in der modernen Informationswelt sehr wichtig. Einer der interessantesten Bereiche auf diesem Gebiet ist die Erstellung einer Ontologie für den Themenbereich Naturkatastrophen, die das Potenzial hat, das Verständnis und die Verarbeitung von Informationen über diese Naturphänomene erheblich zu verbessern.

Die Formalisierung von Konzepten und Beziehungen allein reicht jedoch für die erfolgreiche Umsetzung einer solchen Ontologie nicht aus. Sie erfordert die Entwicklung spezifischer sprachlicher Ausdrücke, die das Wesen dieser Konzepte und Beziehungen in Texten genau wiedergeben. Es ist notwendig, alle möglichen Varianten des Wortgebrauchs zu berücksichtigen, wie Synonyme, Komposita und flektierte Varianten sowie Mehrwortlexeme, die in einem bestimmten Themenbereich vorkommen.

Es ist auch anzumerken, dass die Formalisierung einer Ontologie auf der Grundlage von Mehrfachrelationen zu Schwierigkeiten bei der sprachlichen Disambiguierung führen kann. Daher kann für die computergestützte linguistische Textanalyse ein alternativer Ansatz in Betracht gezogen werden, der auf der Erstellung eines semantischen Netzes mit einer einzigen hierarchischen Beziehung beruht. Dies wird eine klarere Darstellung von allgemeinen Konzepten und Themen im Zusammenhang mit spezielleren Themen zu Naturkatastrophen ermöglichen.

Die Entwicklung einer Ontologie und dafür eines semantischen Netzes für den Themenbereich Naturkatastrophen und die computergestützte linguistische Textanalyse sind also wichtige Aufgaben, die spezielle Kenntnisse und Fähigkeiten erfordern. Diese Arbeit ist darauf ausgerichtet, diese Probleme möglichst effizient zu lösen, um die Verarbeitung von Informationen über Naturkatastrophen zu verbessern und zur Weiterentwicklung des Fachgebiets beizutragen.

3 Theorie und Grundlagen

3.1 Einführung in semantische Netze

3.1.1 Definition von semantischen Netzen und Ontologien

Ein semantisches Netz ist eine grafische Notation zur Darstellung von Wissen in Form von Mustern miteinander verbundener Knoten und Bögen. (Sowa, 1991)

Semantische Netze wurden ursprünglich für die künstliche Intelligenz und die maschinelle Übersetzung entwickelt, haben aber auch Anwendungen in der Philosophie, Psychologie und Linguistik gefunden. Sie bieten eine grafische Möglichkeit, Wissen auf der Grundlage eines deklarativen Ansatzes auszudrücken, der es ermöglicht, Informationen zu speichern und zu automatisieren, so dass semantische Netze als Mittel für Schlussfolgerungen über Wissen verwendet werden können.

Ontologien hingegen sind eine Art von Vokabeldarstellung, die auf einen bestimmten Bereich oder ein bestimmtes Fachgebiet spezialisiert ist. Sie beschreiben die Konzepte und ihre Beziehungen im Zusammenhang mit diesem Bereich. Ontologien werden häufig in der künstlichen Intelligenz, im semantischen Web, in der Softwaretechnik, in der biomedizinischen Informatik, in der Bibliothekswissenschaft und in der Informationsarchitektur verwendet, um das Wissen über die Welt oder einen bestimmten Bereich zu systematisieren und darzustellen.

(Abdel-Badeeh M.Salem, Marco Alfonse, 2008)

Eine Ontologie ist eine detaillierte Spezifikation eines bestimmten Wissensbereichs, die ein Vokabular von Begriffen, die in diesem Bereich verwendet werden, und eine Reihe von logischen Beziehungen enthält, die die Beziehungen zwischen diesen Begriffen beschreiben. Mit anderen Worten, es handelt sich um eine strukturierte Darstellung von Konzepten, ihren Beziehungen und logischen Regeln, die es uns ermöglichen, Informationen zu verstehen und darüber zu kommunizieren.

Ontologien überschneiden sich in vielen Aspekten mit dem bekannten Konzept eines Thesaurus in der Informatik und Linguistik. Sie ermöglichen es jedoch, Konzepte so darzustellen, dass sie für die automatische Verarbeitung durch Computer geeignet sind. Heute gibt es viele verschiedene Ontologien, die in unterschiedlichen Sprachen beschrieben und mit unterschiedlichen Themenbereichen verbunden sind. Sie unterscheiden sich im Umfang, in den Ausdrucksmöglichkeiten, im Zweck, im Grad der Formalisierung und in anderen Merkmalen.

(Рогущина, 2016)

Eine Ontologie bildet zusammen mit einer Reihe spezifischer Klassen die Grundlage für eine Wissensbasis. Der Hauptzweck der Erstellung besteht darin, ein gemeinsames Verständnis der Struktur von Informationen zwischen Menschen oder Software-Agenten zu teilen. Dies vereinfacht die Wiederverwendung von Wissen in einem bestimmten Fachgebiet, macht die zugrunde liegenden Annahmen explizit, trennt Fachwissen von operativem Wissen und ermöglicht die Analyse von Fachwissen. Eines der allgemeinen Ziele der Entwicklung von Ontologien ist es, ein gemeinsames Verständnis der Struktur von Informationen zwischen Menschen oder Software-Agenten zu teilen.

(Abdel-Badeeh M.Salem, Marco Alfonse, 2008)

Der Begriff "semantisch" bezieht sich auf die Bedeutung von etwas und die Semantik ist ein wissenschaftlicher Bereich, der sich mit der Untersuchung der Beziehung zwischen Symbolen und den Objekten befasst, die sie bezeichnen. Die Semantik beschäftigt sich insbesondere damit, wie die Bedeutung von Zeichen erkannt und interpretiert wird. Ein semantisches Netzwerk ist ein Modell, das die Beziehungen zwischen den Objekten in einem bestimmten Fachgebiet darstellt. Es wird in der Regel als Graph dargestellt, bei dem die Knotenpunkte die Objekte und die Kanten die Beziehungen zwischen ihnen repräsentieren. Semantische Netzwerke werden in verschiedenen

Bereichen eingesetzt, wie zum Beispiel in der künstlichen Intelligenz, der Sprachverarbeitung und der Datenanalyse. (ШТУЧНИЙ ІНТЕЛЕКТ В ЗАДАЧАХ КІБЕРБЕЗПЕКИ, 2021)

Der Grundgedanke ist, dass die Bedeutung eines Begriffs in erster Linie durch seine assoziativen Beziehungen zu anderen Begriffen beschrieben werden kann. Dies bedeutet, dass ein Begriff nicht isoliert betrachtet wird, sondern in Beziehung zu anderen Begriffen steht. Diese Beziehungen sind jedoch nicht immer hierarchisch oder nach anderen Ordnungskriterien geordnet. In solchen Fällen ist es oft besser, die Beziehungen zwischen Begriffen als Netz darzustellen.

Um zu veranschaulichen, wie Assoziationen zwischen Begriffen funktionieren, führte Dengel ein kleines Experiment durch. Er bat verschiedene Menschen, bestimmte Wörter zu nennen, und fragte sie dann, ob sie spontan den Begriff nennen könnten, der ihnen als erstes einfiel. Dies führte häufig zu assoziativen Antworten, die subjektive Aspekte widerspiegeln. So wurde zum Beispiel auf das Wort "schwarz" häufig "weiß" geantwortet oder auf den Begriff "Nacht" in vielen Fällen der Begriff "Tag". Assoziationen können also eine Verbindung zwischen zwei völlig unabhängigen Begriffen ausdrücken, die eigentlich nichts miteinander zu tun haben sollten, aber unter bestimmten Bedingungen in eine lose Verbindung gebracht werden können.

Generell ermöglichen semantische Netze eine flexiblere und verfeinerte Darstellung der Bedeutung von Begriffen, da sie nicht streng hierarchischen oder anderen Ordnungskriterien folgen müssen. Stattdessen können sie assoziative Beziehungen zwischen Begriffen aufzeigen, was ein tieferes Verständnis ihrer Bedeutung ermöglicht. (Dengel, 2012)

Semantische Netze abstrahieren vorhandenes Wissen und nutzen dafür einen markierten Graphen. Objekte, wie Themen, Ereignisse, Personen oder deren Eigenschaften werden durch Knoten repräsentiert, während Kanten assoziative Beziehungen zwischen Objekten und ihren Eigenschaften beschreiben. Diese Assoziationen sind oft bidirektional, werden aber oft nur in einer bestimmten Richtung benötigt. Wenn der Zugriff nur auf ein bestimmtes Objekt möglich ist, spricht man von einer gerichteten Beziehung und die Kanten im Graphen sind gerichtet. Durch die Verbindungen entlang der gerichteten Kanten wird jede Information, die zur Darstellung eines Objekts benötigt wird, direkt zugänglich. Das Wissen wird gewonnen, indem man das Netz nach Verbindungen durchsucht. Das hinzugefügte Bild (Abb.) veranschaulicht diese Konzepte und gibt einen visuellen Eindruck davon, wie ein semantisches Netzwerk aufgebaut sein kann.

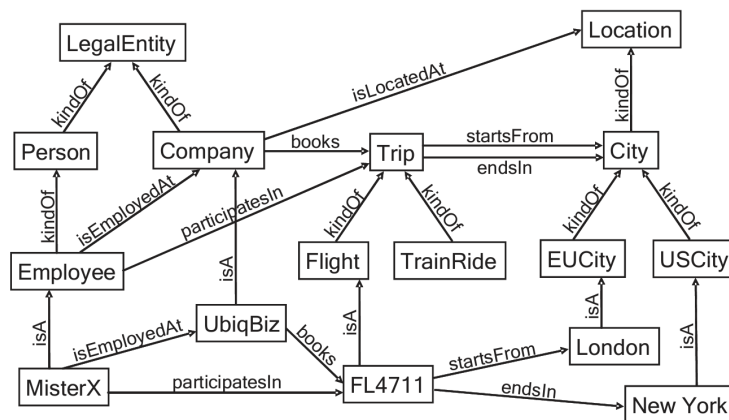


Abbildung 3.1, A Semantic Network for Business Trips

(S.Grimm, P.Hitzler, A.Abecker)

Das Buch "Semantische Technologien: Grundlagen. Konzepte. Anwendungen." von Andreas Dengel beschreibt die grundlegenden Prinzipien semantischer Netze und ihre Bedeutung für die Wissensrepräsentation und Datenanalyse.

Dengel erklärt, dass semantische Netze das Wissen in Form von Knoten und Attributen von Kanten darstellen. Im Allgemeinen wird versucht, Objekte und Eigenschaften mit Beziehungen zu verknüpfen, die verschiedene Formen der Assoziation erfüllen. (Dengel, 2012). Dengel betont die Bedeutung verschiedener Relationstypen, die auf mathematischen Eigenschaften wie Reflexivität, Symmetrie und Transitivität basieren.

Insbesondere beschreibt Dengel die Vererbungsrelation als eine grundlegende Form der Beziehung zwischen Wörtern in der Semantik. Diese Relation verbindet einen spezifischen Begriff, der als Unterbegriff (auch Hyponym genannt) bezeichnet wird, mit seinem übergeordneten Oberbegriff (Hyperonym). Die Vererbungsrelation ist transitiv und asymmetrisch und kann am besten mit dem Ausdruck "*ist-ein*" beschrieben werden.

Dengel erwähnt auch die Aggregation als eine weitere Form der Wissensabstraktion in semantischen Netzen. Aggregation drückt eine starke semantische Beziehung zwischen zwei unabhängigen Objekten aus und ist typisch für Domänen, die sich mit physischen Objekten oder Mengen befassen, wie z.B. Komponenten oder Länder. Dabei werden "*ist ein Teil von*"-Beziehungen verwendet, auch Partonomie oder Meronymie genannt. Die Umkehrung dieser Beziehung wird Holonomie genannt und entspricht der Beziehung "*enthält*" oder "*hat einen Teil*". Während die "*ist-ein(e)*"-Relation gut verstanden ist und umfangreiche Herleitungssysteme existieren, ist die "*ist Teil von*"-Relation schwieriger zu verarbeiten und ihre Semantik hängt vom Fachgebiet ab. Eine andere Form der Beziehung laut Dengel besteht darin, dass einige Begriffe die gleiche Bedeutung haben oder sich auf die gleiche Sache beziehen. Beispielsweise sind "IEC" und "Internationale Elektrotechnische Kommission" synonym, da sie auf dasselbe Konzept verweisen. In ähnlicher Weise ist eine Handynummer "gleich" mit einer Mobiltelefonnummer, weil sie sich auf dieselbe Sache beziehen. Solche Beziehungen sind Äquivalenzbeziehungen, denn sie sind reflexiv (jeder Begriff ist mit sich selbst äquivalent), symmetrisch (wenn $A = B$, dann $B = A$) und transitiv (wenn $A = B$ und $B = C$, dann $A = C$). Eine weitere wichtige Form der Beziehung sind Antonyme, d. h. Wörter, die entgegengesetzte Bedeutungen haben. Zum Beispiel sind "schwarz" und "weiß" Antonyme, weil sie sich auf entgegengesetzte Farben beziehen. Obwohl die Antonymie symmetrisch ist, erfüllt sie nicht die Eigenschaften der Reflexivität und Transitivität.

Beziehungen zwischen Objekten können nicht nur durch komparative Beziehungen, sondern auch durch kausale Abhängigkeiten ausgedrückt werden. Diese Beziehung wird als Kausalbeziehung bezeichnet und häufig durch die Beziehung "*folgt aus*" dargestellt. Solche Beziehungen sind transitiv, aber nicht symmetrisch, da es eine feste Zeitabhängigkeit gibt. (Dengel, 2012)

Semantische Netze sind eine Form der Wissensdarstellung, die aus Knoten und den Verbindungen zwischen ihnen besteht. Sie sind durch eine grafische Darstellung leichter zu verstehen als durch eine logische Form. Darüber hinaus spart diese Methode Speicherplatz, da Objekte und ihre Eigenschaften nur einmal in Form von Knoten benannt werden können. Das Hinzufügen neuer Aussagen zum semantischen Netz erfolgt durch das Hinzufügen neuer Verbindungen. Bei der Modellierung komplexer Sachverhalte können semantische Netze jedoch sehr kompliziert werden und schnell wachsen. Für die Suche in solchen Netzen wird ein Aktivierungsmechanismus in Form einer Breadth-First-Suche verwendet. Bei diesem Verfahren ist es nicht möglich, die Reihenfolge des semantischen Netzes zu überprüfen.

Typen von semantischen Netzen

Semantische Netze sind ein leistungsfähiges Instrument zur Darstellung von Wissen. Wissen kann deklarativ und grafisch dargestellt werden, wodurch Informationen gespeichert, automatisiert und leicht verstanden werden können. Verschiedene Arten von semantischen Netzen, wie Definitionsnetze, Aussagenetze, Implikationsnetze, ausführbare Netze, Lernnetze und hybride Netze, bieten unterschiedliche Ansätze und Fähigkeiten zur Darstellung und zum Verständnis von Wissen. (Abdel-Badeeh M.Salem, Marco Alfonse, 2008)

- Definitionsnetze (definitional networks, Abb.3.2): stellen eine Hierarchie der Verallgemeinerung oder Unterordnung zwischen einem Konzepttyp und seinen Subtypen her. Die Informationen in diesen Netzen werden als notwendigerweise wahr angesehen, da die für einen Supertyp definierten Eigenschaften an alle seine Subtypen vererbt werden.
- Assertionsnetze (assertional networks, Abb. 3.3): sind für die Behauptung von Propositionen bestimmt. Informationen in Behauptungsnetzen werden als bedingt wahr angesehen, wenn sie nicht mit einem Modaloperator gekennzeichnet sind. Einige assertive Netze werden zur Modellierung konzeptueller Strukturen in der Semantik natürlicher Sprachen verwendet.

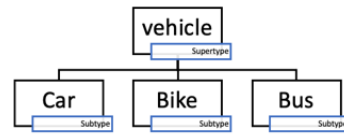


Abb.3.2, definitional networks

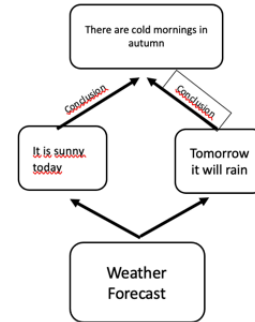


Abb. 3.3, assertional networks

- Implikatorische Netzwerke (implicational networks, Abb 3.4): eine besondere Art von logischen semantischen Netzen sind, die Implikation als Hauptbeziehung zur Verbindung von Knoten verwenden. Diese Beziehung zwischen den Knoten spiegelt die Ursache-Wirkungs-Beziehung wider. Je nach Kontext und Interpretation können diese Netze auch als Vertrauensnetze, kausale Netze, Bayes'sche Netze oder Wahrheitsbeweisungssysteme bezeichnet werden.
- Ausführbare Netze (executable networks, Abb. 3.5): Ein Netz kann eine Prozedur haben, die nach bestimmten Textmustern in Dokumenten sucht. Wenn eine Übereinstimmung gefunden wird, führt das Netz eine Ausgabe durch oder übermittelt eine Nachricht an andere Komponenten des Systems.

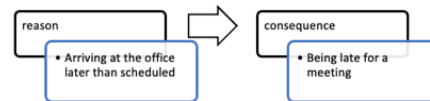


Abb. 3.4, implicational networks



Abb. 3.5, executable networks

- Lernende Netze (learning networks, Abb.3.6): Sie bauen ihre Darstellungen auf oder erweitern sie, indem sie sich Wissen aus Beispielen aneignen. Diese Netze können ihre Struktur und die den Knoten und Bögen zugeordneten numerischen Werte ändern, um sich an neue Informationen anzupassen. Neuronale Netze sind ein Beispiel für semantische Netze, die auf Implementierungen neuronaler Netze aufbauen und anhand von Trainingsbeispielen lernen, Muster zu erkennen oder andere Aufgaben auszuführen.

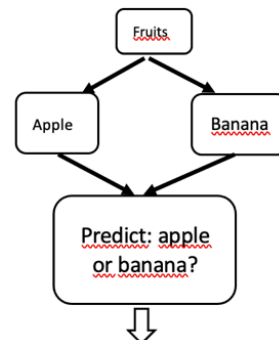


Abb. 3.6, learning networks

- Hybride Netze (hybrid networks): Sie kombinieren zwei oder mehr der oben genannten Methoden, entweder in einem einzigen Netz oder in getrennten, aber interagierenden Netzen. Einige hybride Netze wurden entwickelt, um Hypothesen über menschliche kognitive

Mechanismen zu testen, während andere entwickelt wurden, um die Effizienz von Computern zu steigern.

Darüber hinaus unterscheiden sich semantische Netze in der Anzahl der Typen und werden in homogene (mit einem einzigen Typ) und heterogene (mit verschiedenen Typen) unterteilt. Es gibt wiederum binäre (Beziehungen, die zwei Objekte verbinden) und n-äre (die mehr als zwei Konzepte verbinden) Beziehungen. Nach der Art der Knotenstruktur werden sie in einfache (die Knoten haben keine eigene Struktur) und hierarchische Typen (sie haben eine eigene Struktur in Form eines Netzes) unterteilt. Hierarchische Netze können das Netz in Teilnetze untergliedern. (СЕМАНТИЧНІ МЕРЕЖІ ПСІШ ТА ЇХ КЛАСИФІКАЦІЯ, 2022)

3.1.2 Semantische Netze in Datenanalyse und Wissensrepräsentation

Semantische Netze und Ontologien sind wichtige Instrumente für die Entwicklung und Anwendung von Informationssystemen. Sie ermöglichen die Darstellung von Wissen in Form von strukturierten Daten, was die Kommunikation zwischen verschiedenen Computersystemen und Menschen erleichtert.

Zu den wichtigsten Vorteilen der Verwendung semantischer Netze gehören:

- Erleichterung der Kommunikation: Die Verwendung von semantischen Netzen ermöglicht das Verständnis und den Transfer von Wissen zwischen verschiedenen Computersystemen und Personen, die im selben Bereich arbeiten.
- Flexibilität und Wiederverwendbarkeit: Semantische Netze erleichtern es, Wissen zu ändern und zu erweitern, ohne das gesamte System neu schreiben zu müssen. Darüber hinaus können bestehende semantische Netze in verschiedenen Projekten wiederverwendet werden.
- Sicherstellung der Kompatibilität: Die Verwendung des gleichen semantischen Netzes gewährleistet die Interoperabilität zwischen verschiedenen Computersystemen und ermöglicht einen einfachen Datenaustausch zwischen ihnen.
- Modellierungssicherheit und Lösungsflexibilität: Semantische Netze ermöglichen es, Wissen in Form von konzeptuellen Modellen darzustellen, was ihre Zuverlässigkeit erhöht und einen flexibleren Ansatz zur Lösung von Problemen ermöglicht.
- Benutzerfreundlichkeit: Semantische Netze ermöglichen es den Nutzern, mit Hilfe einer Suchfunktion leicht die benötigten Informationen zu finden.
- Transparenz für das technische Personal: Semantische Netze sorgen für Transparenz beim Verständnis, wie ein System funktioniert und welches Wissen es nutzt. Dies ermöglicht es dem technischen Personal, Probleme im Zusammenhang mit der Systementwicklung und -wartung effizienter zu lösen.

Die Verwendung von semantischen Netzen und Ontologien hat jedoch auch ihre Grenzen. Einige davon sind:

- Komplexität der Entwicklung: Die Entwicklung von semantischen Netzen und Ontologien kann komplex und zeitaufwändig sein. Um sie zu entwickeln, muss man über fundierte Kenntnisse in dem betreffenden Fachgebiet verfügen.
- Konformitätsprobleme: Wenn das System verschiedene Semantische Netze verwendet, kann es zu Problemen bei der Datenübereinstimmung kommen. So kann beispielsweise ein Begriff in verschiedenen Ontologien unterschiedliche Bedeutungen haben.
- Probleme bei der Unterstützung: Wenn semantische Netze und Ontologien schlecht konzipiert oder dokumentiert sind, kann es Probleme mit ihrer Wartung und künftigen Entwicklung geben.

Zusammenfassend lässt sich sagen, dass die Verwendung von semantischen Netzen und Ontologien ein wichtiges Instrument für die Entwicklung und Anwendung von Informationssystemen ist, jedoch ein gewisses Maß an Fachwissen und Aufmerksamkeit für Details erfordert, um erfolgreich zu sein.

3.1.3 Anwendungen von semantischen Netzen in verschiedenen Bereichen

Semantische Netze sind ein wirksames Mittel zur Darstellung von Wissen, das die Struktur von Konzepten und die Beziehungen zwischen ihnen in Form eines Graphen darstellt. Ein wichtiger Aspekt dieses Ansatzes ist die Hervorhebung der Verbindungen zwischen verschiedenen Informationseinheiten und Wissensbeständen. Es ist zu beachten, dass die Informationen zu jedem Konzept um den Netzknoten gruppiert sind, der es repräsentiert.

Heute wird der Anwendungsbereich semantischer Netze aktiv erweitert. Sie werden zur Beantwortung von Fragen, zur Analyse von Lern-, Gedächtnis- und Argumentationsprozessen und zur Beurteilung des Zustands von Patienten auf der Grundlage medizinischer Untersuchungsdaten eingesetzt. Diese Netze werden häufig in Expertensystemen eingesetzt.

Es wurde festgestellt, dass die Robotik und die Medizin die wichtigsten Anwendungsbereiche für semantische Netze sind. In der Robotik helfen sie bei der Lösung von Problemen der Mustererkennung und der Bewegungssteuerung, und in der Medizin werden sie zur Diagnose und Behandlung verschiedener Krankheiten eingesetzt.

Semantische Netze sind ein unverzichtbares Instrument der künstlichen Intelligenz, da sie Maschinen helfen, die Bedeutung von Sprache zu verstehen und zu interpretieren, was eine Schlüsselkomponente der natürlichen Sprachverarbeitung (NLP) ist. Semantische Netze ermöglichen es, Wissen in einer Form darzustellen, die Maschinen leicht verstehen und verarbeiten können, wodurch NLP-Aufgaben erleichtert werden. Indem sie die Beziehungen zwischen verschiedenen Konzepten und Objekten verstehen, können Maschinen Sprache genauer interpretieren und bessere Antworten geben. Sie werden in der künstlichen Intelligenz in einer Vielzahl von Bereichen eingesetzt. Sie helfen bei der Beantwortung von Fragen, der Suche nach Informationen, der Erstellung von Chatbot-Systemen und der Organisation und Verwaltung von Wissen in Organisationen. (Nova, 2023)

Die Verwendung semantischer Netze in der Biologie und Bioinformatik ermöglicht es, biologische Systeme und Prozesse zu modellieren, Wechselwirkungen zwischen verschiedenen biologischen Objekten zu untersuchen und ihre Rolle in biologischen Prozessen herauszufinden. Solche Netze können in der Genetik verwendet werden, um die Beziehungen zwischen Genen und ihren Funktionen zu untersuchen, sowie in der Molekularbiologie, um molekulare Interaktionen und die Rolle verschiedener Moleküle in biologischen Prozessen zu untersuchen. (M.Hsing, and A.Cherkasov, 2006)

Semantische Netze finden ihre Anwendung in der Cybersicherheit, wo sie zur Analyse von Netzaktivitäten und zur Erkennung von anomalem Verhalten eingesetzt werden. Auf diese Weise können mögliche Cyberangriffe erkannt und verhindert werden, und es kann schnell auf Sicherheitsvorfälle reagiert werden. Mit Hilfe semantischer Netze können Daten über den Netzverkehr und andere netzbezogene Parameter analysiert werden, wodurch anormale Aktivitäten erkannt und Sicherheitsbedrohungen identifiziert werden können. Dieser Ansatz für die Cybersicherheit ermöglicht einen wirksameren Netzschutz und verringert das Risiko von Cyberangriffen und Datenverletzungen. (He, 2014)

Generell sind semantische Netze ein leistungsfähiges Instrument für die Darstellung von Wissen und die Analyse von Informationen in verschiedenen Tätigkeitsbereichen.

(Андреев, 2021)

3.2 Naturkatastrophen und ihre Folgen

3.2.1 Ein Überblick über Naturkatastrophen und ihre Auswirkungen auf die Umwelt und die Gesellschaft

Naturkatastrophen sind außergewöhnliche Naturereignisse, die eine starke Zerstörungskraft haben und in den Gebieten, in denen sie auftreten, erhebliche Schäden verursachen. Sie stören das normale Funktionieren der Bevölkerung, verursachen die Zerstörung von Gebäuden und Strukturen, gefährden Menschenleben und führen zum Verlust von Menschenleben, Tieren und Sachwerten. Angesichts des Einflusses der vom Menschen verursachten Faktoren und des Urbanisierungsprozesses nimmt das

Ausmaß der Folgen von Naturkatastrophen ständig zu, so dass das Problem der Verhütung von Notfällen und der Beseitigung oder Minimierung ihrer Folgen von großer Bedeutung ist.

Die Kenntnis der Ursachen, der Entwicklungsdynamik und der Art der Einflussfaktoren von Naturkatastrophen trägt dazu bei, die Bedrohung für das Leben und die Gesundheit von Menschen zu verringern, materielle und wirtschaftliche Verluste in der Wirtschaft zu verhindern oder zu verringern und eine Reihe von Präventions-, Rettungs- und Notfallmaßnahmen im Zusammenhang mit Naturkatastrophen wirksam umzusetzen.

Wissenschaftlichen Untersuchungen zufolge ist unser Planet etwa 4,54 Milliarden Jahre alt. In dieser langen Zeitspanne des Bestehens der Erde haben sich verschiedene Epidemien, Katastrophen und Naturkatastrophen ereignet. Die meisten dieser Ereignisse hatten nur minimale Auswirkungen auf lebende Organismen. Schwache Erdbeben, die kaum zu spüren sind, leichte Krankheiten, die schnell wieder verschwinden, ohne dass Menschen zu Schaden kommen, oder Brände, die schnell wieder erlöschen - all das ist alltäglich geworden und erregt nicht die allgemeine Aufmerksamkeit der Menschheit.

In jüngster Zeit ist jedoch ein alarmierender Trend zu beobachten, dass die Zahl der Naturkatastrophen und der vom Menschen verursachten Katastrophen, der Naturgefahren und ihres Ausmaßes stetig zunimmt. Dies hat negative Auswirkungen auf die sozioökonomische Entwicklung der Länder. Zu Beginn des 21. Jahrhunderts hat dieses Problem bereits ein solches Ausmaß erreicht, dass es die Sicherheit der Staaten und ihrer Bevölkerung erheblich beeinträchtigt.

Der Schutz der Bevölkerung vor verschiedenen Katastrophen und Naturgefahren ist für viele Länder der Welt eine der wichtigsten Prioritäten. Die jüngsten Ereignisse zeigen, dass Naturkatastrophen keine Landesgrenzen haben und keine Nationalitäten berücksichtigen. Vor allem in den letzten Jahren hat die Zahl der Naturkatastrophen im Zusammenhang mit dem Klimawandel, Überschwemmungen, Schlammlawinen, Erdbeben, Dürren und Bränden zugenommen. Diese Katastrophen breiten sich immer mehr aus und verursachen erhebliche Schäden.

Die Veränderungen der natürlichen Umwelt, die zunehmende seismische Aktivität, das wachsende Ausmaß und die zunehmende Leistungsfähigkeit der technischen Systeme sowie die aktiven Eingriffe des Menschen in die Natur - all diese Faktoren erhöhen die Risiken von Natur-, Umwelt- und Industriekatastrophen. Die wirtschaftlichen Verluste durch diese Katastrophen sind so hoch, dass sie das Niveau der entwickelten Länder übersteigen und eine Bedrohung für die Bevölkerung, die Natur und die Weltwirtschaft darstellen.

Naturkatastrophen sind durch unerwartete und rasche Veränderungen der Umwelt gekennzeichnet, die zu einer großen Zahl von Opfern und erheblichen Sachschäden führen. Es handelt sich um Notfälle, die infolge von Naturkatastrophen entstehen und in bestimmten Gebieten eine gefährliche Situation schaffen. Solche Katastrophen können zu menschlichen Opfern, Gesundheits- und Umweltschäden, erheblichen materiellen Verlusten und einer Störung des normalen Funktionierens der Bevölkerung führen. In verschiedenen Ländern werden Forschungsarbeiten durchgeführt, um eine Datenbank über Naturkatastrophen und ihre Verteilung auf der ganzen Welt zu erstellen. Anhand dieser Daten werden entsprechende Maßnahmen zum Schutz der Bevölkerung und der Wirtschaft vor schweren Verlusten und Schäden durch Naturkatastrophen entwickelt. (Азаров С. И., Задунай О. С., 2019)

Die Vorhersage der Umweltfolgen von Naturkatastrophen ist eine äußerst wichtige Aufgabe, da sich ihre Auswirkungen erst nach Jahrzehnten zeigen können. Dazu können ein Verlust der Produktivität von Ökosystemen, Veränderungen im Wasserkreislauf und die Störung wichtiger Umweltparameter gehören. Leider gibt es bis heute keine umfassenden Analysemethoden, mit denen Naturkatastrophen und ihre Auswirkungen auf die Umwelt vollständig untersucht werden könnten. Daher sind die wissenschaftlichen Entwicklungen in diesem Bereich nach wie vor relevant und erfordern weitere Aufmerksamkeit. (И.Кравець, kein Datum)

3.2.2 Klassifizierung von Naturkatastrophen und deren Merkmale

Eine Naturkatastrophe ist eine natürliche Gefahr, die zum Verlust von Menschenleben und zu erheblichen wirtschaftlichen Schäden führt. Solche Folgen werden häufig durch Erdbeben,

Erdrutsche, Lawinen, Erdbeben, Gletscherschmelze, Überschwemmungen, Vulkanausbrüche, Waldbrände, Gewitter, Tornados, Stürme, Hitzewellen, Wirbelstürme usw. verursacht. (Азаров С. И., Задунай О. С., 2019)

Der Begriff "Naturkatastrophe" wird häufig mit dem Begriff "Umweltsicherheit" in Verbindung gebracht, der sich aus der Notwendigkeit ergibt, das Risiko für die Bevölkerung eines Gebiets im Hinblick auf mögliche Schäden an Gesundheit, Eigentum und Infrastruktur infolge von Umweltveränderungen zu bewerten. Veränderungen können sowohl durch natürliche als auch durch anthropogene Faktoren verursacht werden. Naturkatastrophen können gefährlich sein, weil sie unerwartet auftreten und ein Gebiet schnell verwüsten können, indem sie Gebäude, Eigentum und Infrastrukturen zerstören. Darüber hinaus können solche Katastrophen Folgen haben, die sich weiter ausbreiten, wie Hungersnöte, Krankheiten und Epidemien. Auch anthropogene Umweltveränderungen können katastrophale Folgen haben, z. B. Luft-, Wasser- und Bodenverschmutzung, die sich negativ auf die menschliche Gesundheit und die Ökosysteme auswirken können.

Naturkatastrophen weisen allgemeine Muster auf, die sich nachvollziehen lassen. Zunächst einmal zeichnet sich jede Art von Naturkatastrophe durch einen bestimmten Ort ihres Auftretens aus. Je stärker ein Naturereignis ist, desto seltener tritt es auf. Außerdem kann man vor jeder Naturkatastrophe bestimmte Signale wahrnehmen, die ihr vorausgehen. Zwar sind Naturkatastrophen unerwartet, aber sie lassen sich vorhersagen. Darüber hinaus gibt es passive und aktive Schutzmaßnahmen, die dazu beitragen können, die Risiken von Naturkatastrophen zu verringern. Die allgemeine Vielfalt der Ursachen von Naturkatastrophen macht es schwierig, sie vorherzusagen und zu verhindern, was zu erheblichen Verlusten führt.

Naturkatastrophen werden wie folgt unterteilt:

1. geologisch: Erdbeben, Vulkanausbrüche, Erdbeben, Karst;
2. meteorologisch: Wirbelsturm, Sturm, Tornado, Sturmböen, Regensturm, Hagel, starker Schneefall, Eis, strenger Frost, Lawine, extreme Hitze, Dürre, Staub, Staubsturm, Naturbrand;
3. hydrologisch: Überschwemmungen, Murenabgänge, frühzeitige Vereisung, frühzeitige Waldüberflutung;
4. Meeresbedingte Naturkatastrophen: tropischer Wirbelsturm (Taifun), Tsunami (Seebeben), schwerer Sturm, schwierige Eisverhältnisse;
5. heliophysikalisch: biomagnetische Stürme, erhöhte Sonnenaktivität, Verletzung der Ausbreitungsbedingungen für Radiowellen (magnetische Stürme);
6. biologisch (biologisch und sozial): Epidemien - Massenerkrankungen von Menschen,
 1. Epizootien - Massenerkrankungen von Tieren, Epiphytotika - Massenerkrankungen von
 2. von Pflanzen;
 7. Natürliche Brände: Wald, Torf, Steppe;
 8. Weltraum: Asteroiden, Kometen, Strahlung, interplanetarische Schwerkraft.

Die Untersuchung der Bedingungen für das Auftreten extremer Naturphänomene ermöglicht es uns, die Gesetze der Interaktion zwischen biologischen und physikalischen Teilsystemen der Umwelt zu identifizieren und zu formulieren, wodurch zumindest eine statistische Vorhersage von Naturgefahren möglich wird. (Азаров С. И., Задунай О. С., 2019)

Eine Naturkatastrophe ist ein großflächiges Naturereignis oder ein Prozess, der zu einer Bedrohung für das Leben und die Gesundheit von Menschen, zur Zerstörung oder Beschädigung von Eigentum und Bestandteilen der natürlichen Umwelt führen kann. Die Wirtschaft des Landes leidet unter Naturkatastrophen, denn sie führen zur Zerstörung von Produktionsanlagen, zum Verlust von Sachwerten und zum Verlust von Menschenleben. Darüber hinaus schaffen Naturkatastrophen ungeeignete Lebensbedingungen für die Bevölkerung, was zu einer massiven Ausbreitung von Infektionskrankheiten führen kann. Die Zahl der von Naturkatastrophen betroffenen Menschen kann erheblich sein, und die Art der Schäden kann sehr unterschiedlich sein. Überschwemmungen, tropische Stürme, Dürren und Erdbeben verursachen die größten wirtschaftlichen Schäden, da sie für das Leben und die Gesundheit der Menschen am gefährlichsten sind. Die Analyse der Entwicklung der Naturgefahren zeigt, dass trotz des wissenschaftlichen und technischen Fortschritts der Schutz

von Menschen und materiellen Ressourcen vor Naturgefahren nicht besser wird. Jedes Jahr steigt die Zahl der durch Naturkatastrophen getöteten Menschen in der Welt um 4,3 %, die Zahl der Verletzten um 8,6 % und die Höhe der Sachschäden um 10,4 %. Im Allgemeinen verliert jeder hunderttausendste Mensch auf der Erde sein Leben durch Naturkatastrophen, und in den letzten hundert Jahren waren es jährlich 16.000 Menschen. (Азаров С. И., Задунай О. С. , 2019)

3.2.3 Ursachen und Prävention von Naturkatastrophen

Naturkatastrophen haben schon immer eine wichtige Rolle in der globalen Umweltdynamik gespielt. In der Vergangenheit wurden Naturkatastrophen und verschiedene Naturkatastrophen in Übereinstimmung mit natürlichen Trends beobachtet. Seit dem 19. Jahrhundert wird ihre Dynamik jedoch von anthropogenen Faktoren beeinflusst. Mit der Destabilisierung des Klimas hat die Zahl der Katastrophen im Allgemeinen zugenommen. (Азаров С. И., Задунай О. С. , 2019)

Naturkatastrophen unterschiedlicher Art treten aus verschiedenen Gründen auf, z. B. Bodenerosion, seismische Aktivität, tektonische Bewegungen, atmosphärischer Druck, Meeresströmungen usw. Es handelt sich um Naturphänomene, die seit Beginn der Existenz der Erde zu beobachten sind und immer wieder schwere Schäden und Verluste an Menschenleben verursachen. Viele dieser Naturkatastrophen lassen sich auf Ungleichgewichte in der Umwelt zurückführen, z. B. auf Luft-, Lärm- und Wasserverschmutzung. Es ist jedoch wichtig, darauf hinzuweisen, dass die Naturkatastrophen nicht allein auf die Modernisierung zurückzuführen sind, da diese Phänomene auch in früheren Epochen auftraten, als die Menschen noch nicht den heutigen Stand der Technik erreicht hatten.

Natürliche Prozesse wie seismische Aktivität, tektonische Bewegungen und vulkanische Aktivität sind die Hauptursachen für Naturkatastrophen. Erdbeben können schwere Schäden und den Verlust von Menschenleben verursachen. Auch Vulkanausbrüche und Taifune werden mit seismischer Aktivität in Verbindung gebracht. Die Aktivität des Mondes beeinflusst die Meereswellen, insbesondere bei Vollmond. Dies wurde durch den tödlichsten Tsunami der Geschichte im Dezember 2004 bestätigt. Ein Erdbeben vor der birmanischen Küste löste eine Reihe von verheerenden Tsunamis aus, die mehr als 230 000 Menschen töteten und erhebliche Schäden verursachten.

Veränderungen der Meeresströmungen können zu Veränderungen der Wassertemperatur führen, die eine weltweite Nahrungsmittelknappheit bedrohen und für Fische und Meereslebewesen katastrophale Folgen haben. Veränderungen der Strömungen können auch die Intensität und Häufigkeit von Stürmen beeinflussen. Tornados, die durch das Zusammentreffen von Hoch- und Tiefdruckgebieten entstehen, können für die Bevölkerung gefährlich und zerstörerisch sein, insbesondere in der Tornado Alley. Der atmosphärische Druck bestimmt das Auftreten von Gewittern, Regen und Wirbelstürmen. Überschwemmungen und Wirbelstürme verursachen in Küstenstädten große Schäden und gefährden Leben und Wohlbefinden.

Naturkatastrophen sind eine Reihe von Naturphänomenen, die die Gesundheit und das Wohlergehen der Menschen bedrohen. Sie haben ihre Ursachen in der normalen Aktivität der Erde, aber in letzter Zeit hat die Zahl der Katastrophen aufgrund der Modernisierung und des Wachstums des menschlichen Wissens und der Technologie zugenommen. Umweltverschmutzung, Bergbau, Abholzung, globale Erwärmung und Bauarbeiten können Überschwemmungen, Erosion und Erdbeben verursachen.

Über einen langen Zeitraum hinweg sind Wissenschaftler zu dem Schluss gekommen, dass die Zunahme der hydrometeorologischen Katastrophen auf das Zusammenspiel natürlicher und anthropogener Faktoren zurückzuführen ist. Das Hauptproblem ist die globale Erwärmung, durch die die Temperatur der Ozeane und der Atmosphäre ansteigt, was zu intensiveren Stürmen wie Wirbelstürmen und Überschwemmungen aufgrund schmelzender Gletscher führt. Die Zahl der Gebäude in überschwemmungsgefährdeten Regionen nimmt zu, wodurch die Gefahr von Sturm- und Küstenüberschwemmungen in diesen Städten und Dörfern steigt. (Natural Disasters- Cause and Effects, kein Datum)

Die Fortschritte bei der Wettervorhersage und die Entwicklung von Frühwarnsystemen haben jedoch große Sprünge bei der Verringerung der Auswirkungen von Naturkatastrophen gemacht. Der Einsatz

seismischer und bodengestützter Sensoren, von Satelliten, Flugzeugen und schwimmenden Bojen auf See hat die Systeme zur Erkennung und Klassifizierung der Kräfte von Naturphänomenen, die zu Katastrophen führen können, verbessert, so dass diese im Voraus gewarnt werden können, bevor sie Schäden verursachen. Eine besondere Rolle spielen dabei die nationalen Wetterdienste, die das Wettergeschehen überwachen und vor Stürmen und anderen Klimaphänomenen warnen, die ein bestimmtes Gebiet betreffen können. Mit ihren über das ganze Land verteilten Stationen sind diese Dienste in der Lage, die lokalen Wetterbedingungen mehrmals am Tag zu messen. Die Daten werden zur Entwicklung von Wettermodellen und zur Vorhersage der Stärke und des Standorts von Stürmen verwendet, und zwar mehrere Tage, bevor sie sich bestimmten Gebieten nähern.

Darüber hinaus überwachen nationale Regierungsstellen und internationale Organisationen wie die China Earthquake Administration, die Japan Meteorological Agency und andere spezifische physikalische Kräfte, die Zerstörung und tödliche Naturkatastrophen verursachen können. Sie arbeiten untereinander und mit internationalen Organisationen zusammen, um Warnungen herauszugeben, Sicherheitsstandards zu entwickeln und die Risiken im Zusammenhang mit Naturgewalten zu bewerten, die mehrere Länder oder die ganze Welt betreffen. Einige Frühwarnsysteme haben sogar einen weltraumgestützten Aspekt, wie z. B. das Near-Earth Object Observation System der Europäischen Weltraumorganisation (ESA) und die Asteroidenwarnsysteme der NASA, die dazu beitragen, die Risiken von Kollisionen mit gefährlichen Objekten aus dem Weltraum zu erkennen und vorherzusagen. (Metych, 2023)

4 Erstellung eines semantischen Netzes

Im Rahmen dieses Kapitels wird der praktische Teil der Arbeit behandelt. Dabei werden verschiedene Schritte durchgeführt, um das Ziel zu erreichen. Dazu gehört die Erfassung und Definition des fachspezifischen Vokabulars, gefolgt von der Einordnung und Strukturierung dieses Vokabulars. Anschließend wird ein semantisches Netz erstellt, das die Beziehungen zwischen den Konzepten widerspiegelt. Schließlich wird das semantische Netz evaluiert, um seine Qualität und Effektivität zu überprüfen. Dieses Kapitel schafft einen praktischen Rahmen für die Implementierung semantischer Netze und trägt zum erfolgreichen Abschluss der Arbeit bei.

4.1 Erfassung und Definition des fachspezifischen Vokabulars Korpusaufbau

Der vorliegende Textkorpus beschäftigt sich mit dem Thema „Naturkatastrophen“. Bei der Erstellung dieses Korpus wurden Artikel aus der weltberühmten Enzyklopädie Wikipedia¹ als Grundlage verwendet. Diese Seiten lieferten einen umfassenden Überblick über das Thema und spielten eine wichtige Rolle bei der Festlegung der grundlegenden Prinzipien des Semantischen Netzes. Sie dienten dazu, einfache und klare logische Verbindungen zwischen verschiedenen Konzepten herzustellen.

Der Hauptteil des Korpus besteht aus Nachrichtenartikeln, die über die aktuellsten Naturkatastrophen in der Welt berichten, hauptsächlich aus dem Zeitraum von 2019 bis 2023. Als Informationsquelle wurden verschiedene Nachrichtensender wie Spiegel.de² und MunichRE³ verwendet. Die wichtigsten Daten kommen von einer plattformübergreifenden Nachrichtenmarke, nämlich dem Bayerischen Rundfunk⁴. Diese Informationsquelle ist aufgrund ihrer Aktualität und Relevanz für die Ereignisse ausgewählt. Dieser Ansatz bei der Erstellung des Korpus ermöglicht es uns, frische und relevante Informationen über Naturkatastrophen aufzunehmen. Die Nachrichtenartikel spiegeln verschiedene Aspekte von Katastrophen, deren Ursachen, Folgen und Präventionsmaßnahmen wider. Auf diese Weise können wir unser Verständnis dieses Themas erweitern und es im Zusammenhang mit den aktuellen Ereignissen in der Welt betrachten.

Der auf Wikipedia-Artikeln und Nachrichtenmaterial basierende Korpus ist somit eine wertvolle Forschungsressource. Er bietet Zugang zu Informationen über Naturkatastrophen und ermöglicht die Nutzung des semantischen Webs zur Analyse und Untersuchung dieses Themas im Kontext aktueller Ereignisse, die sich in Nachrichtenartikeln widerspiegeln.

Unser Korpus besteht insgesamt aus 12 Unterthemen, die den Schlüsselwörtern entsprechen und eine Gesamtgröße von etwa 4 Mb aufweisen. Um die Größe des Korpus zu kontrollieren und die Verarbeitung zu erleichtern, haben wir uns dazu entschieden, ihn in Unterthemen aufzuteilen. Diese Entscheidung wurde getroffen, da bestimmte Modelle, einschließlich der Lemmatisierung in spaCy⁵, Schwierigkeiten bei der Verarbeitung großer Textmengen hatten. Die Aufteilung in Unterthemen ermöglichte es uns auch, Schlüsselwörter für jedes Thema zu identifizieren und vereinfachte die Analyse und Bearbeitung des Textmaterials.

Anwendung von Crawler

Für Korpusaufbau haben wir eine Vielzahl von Such-, Sortier- und Retrievalverfahren aus der Computerlinguistik eingesetzt. Dazu wurden verschiedene Crawler und Parser eingesetzt, die es

¹ <https://de.wikipedia.org/wiki/Naturkatastrophe>

² <https://www.spiegel.de/>

³ <https://www.munichre.com/de.html>

⁴ <https://www.br.de/nachrichten/>

⁵ <https://spacy.io/>

ermöglichten, die erforderlichen Daten aus den Webseiten zu extrahieren. Ein Crawler ist ein wesentlicher Bestandteil der automatisierten Informationssammlung. Die Hauptkomponenten eines jeden Crawlers sind:

- Der Loader der Daten, die Daten können als HTML, JSON oder jedes andere Datenformat dargestellt werden. Ein optionaler Schritt während des Ladens kann die Ausführung der geladenen Daten sein, ein Beispiel kann Javascript sein, das in das HTML eingefügt wird und die Informationen auf der Seite anreichert.
- Parser, um die sinnvollen Informationen aus dem geladenen Stapel zu extrahieren.
- Collector und Output Builder, verantwortlich für die Bereitstellung der endgültigen Ausgabe.

Das Laden der Daten ist der erste und zugleich wichtigste Schritt beim Aufbau eines Crawlers. Die Hauptschwierigkeiten beim Aufbau eines Crawlers können statische und dynamische Seiten sein. Das heißt, es gibt Seiten, die direkt bei der ersten Anfrage an den Server mit allen benötigten Daten versorgt werden. Da einige Web-Seiten ein einzigartiges Markup mit verschiedenen HTML-Elementen aufweisen, wurde ein eigener Crawler für jede Seite entwickelt. Dies half, sauberen Text zu erhalten und das Vorhandensein von Werbung, Bildbeschreibungen und anderen unerwünschten Informationen zu minimieren.

Einer der positiven Aspekte ist, dass viele Seiten eine ähnliche Struktur haben, was es einfacher machte, die richtigen Informationen und Elemente zu finden, um sie in den Korpus aufzunehmen. Für die Wikipedia-Seiten wurde ein Programm **crawler_for_wiki.py** entwickelt, das auf dem BeautifulSoup-Parser⁶ basiert.

Als erstes stellt die Request-Bibliothek eine HTTP-Anfrage an die angegebene URL. Nach dem Senden der Anfrage empfängt der Code die Antwort von der Webseite.

```
url = 'https://de.wikipedia.org/wiki/Naturkatastrophe'  
response = requests.get(url)
```

Anschließend wird die Antwort von der BeautifulSoup-Bibliothek verarbeitet, die es ermöglicht, HTML-Code zu parsen und strukturierte Daten von einer Webseite zu erhalten. Mit "html.parser" erstellt der Code ein Soup-Objekt, das die HTML-Struktur der Seite enthält und die weitere Analyse, Extraktion und Verarbeitung der erforderlichen Daten aus der Seite ermöglicht

```
soup = BeautifulSoup(response.content, "html.parser")
```

Diese einfache Lösung ermöglichte, Links zu den gewünschten Artikeln anzugeben, die dann schnell und einfach hochgeladen und in die angegebene Datei geschrieben wurden. Das Programm navigierte automatisch zu den angegebenen Ressourcen, erkannte und bereinigte den Text von unnötigen HTML-Elementen und Informationen und fügte ihn dann unserem Korpus hinzu.

```
content = soup.find('div', class_='mw-parser-output')  
text = content.get_text()
```

Allerdings hatte dieser Ansatz den Nachteil, dass er zeitaufwendig war und die Seiten selbst nicht gleichzeitig eine große Menge an Inhalten lieferten. In gewisser Weise kann dieser Prozess mit der "Copy-Paste"-Methode verglichen werden.

Für dynamische Seiten ist es nicht möglich, die "requests"-Bibliothek zu verwenden, um den endgültigen Zustand der Seite zu erhalten, da in der ersten Antwort nur ein Teil des Markups empfangen wird, der Rest wird durch Javascript und asynchrone HTTP-Anfragen geladen. Dafür

⁶ <https://www.crummy.com/software/BeautifulSoup/bs4/doc/>

sollte ein anderes Tool verwendet werden, das über Interpretationsfähigkeiten verfügt, so dass es Netzwerkanfragen ausführen und auch das Markup mit den geladenen Daten aktualisieren kann.

Um erfolgreich Informationen von BR24-Seiten abzurufen, musste eine modifizierte Version des oben verwendeten Algorithmus entwickelt und angewandt werden, die die Besonderheiten und Komplexitäten im Zusammenhang mit dynamischen Inhalten auf diesen Seiten berücksichtigt. In solchen Situationen, in denen sich Inhalte ändern oder nach dem ersten Laden der Seite geladen werden, wird die Automatisierung von Webbrowser-Aktionen zu einem wichtigen Werkzeug, und dabei hilft Selenium⁷.

Mit Selenium kann man die Benutzeraktionen emulieren, einschließlich der Interaktion mit Seitenelementen, des Ausfüllens von Formularen, des Scrollens und anderer Aktionen, wodurch man die Möglichkeit hat, aktuelle und vollständige Informationen zu erhalten. In unserer Anwendung **search_parser_selenium.py** hat der Browser selbst alle Elemente geladen, bearbeitet und sie direkt in die Datei geschrieben. Unsere einzige Aufgabe bestand darin, ihm die Suchanfragen zu übermitteln. Der CSS-Selektor sucht das Inhaltselement, das den Inhalt des Artikels enthält.

```
article_content_selector = '#articlebody[class^="ArticleModuleText_content"]'
```

"WebDriverWait" soll darauf warten, dass das Element mit dem angegebenen CSS-Selektor auf der Webseite vorhanden ist, und fährt dann mit dem Rendern der Seite fort.

```
WebDriverWait(driver, 10).until(EC.presence_of_element_located((By.CSS_SELECTOR, article_content_selector)))
```

Dieses leistungsstarke Tool wurde bei der Erfassung von Daten aus BR24-Webseiten unverzichtbar. Der Prozess begann mit einer Website-Suche unter Verwendung von Schlüsselwörtern, die sorgfältig ausgewählt und manuell festgelegt wurden. Anschließend wurden die gefundenen Links zu Seiten, die mit dem gewünschten Thema zusammenhängen, organisiert und in ein Dokument in einem geeigneten Format wie JSON geschrieben. Dies erleichterte den Übergang zur nächsten Phase, in der bereits Selenium-Tools zum Einsatz kamen.

Mithilfe von Selenium und seinen Automatisierungsfunktionen konnten wir jeden Link effizient verarbeiten und die erforderlichen Informationen extrahieren. Ein speziell für die Verarbeitung von BR24-Seitendaten entwickelter und abgestimmter Algorithmus sorgte für eine reibungslose Interaktion mit dynamischen Inhalten und eine hohe Abrufgenauigkeit.

Dieser automatisierte Ansatz unter Verwendung von Selenium reduzierte die manuelle Arbeit erheblich und minimierte den Zeitaufwand für die Erstellung des Datensatzes. Dadurch konnten wir uns auf komplexere Aufgaben konzentrieren, wie die Verarbeitung und Vorverarbeitung der erhaltenen Daten, die Analyse und Modellierung von Naturkatastrophen mithilfe eines semantischen Netzes.

Der Einsatz von Selenium in Verbindung mit unserem modifizierten Algorithmus ermöglichte es uns, relevante und aktuelle Informationen von den BR24-Webseiten zu erhalten, die eine solide Grundlage für die Erstellung eines hochwertigen und aussagekräftigen Datenkorpus zu Naturkatastrophen bilden.

4.2 Einordnung und Strukturierung des Vokabulars

Nachdem der Korpus gesammelt worden war, bestand die nächste Aufgabe dieser Arbeit darin, das Thema Naturkatastrophen weiter zu erforschen und zu vertiefen. Dieser Schritt war wichtig für den Aufbau eines semantischen Netzes und die Schaffung einer konsistenten Kette von Konzepten. Dieser Ansatz ermöglichte es dem Benutzer, die Informationen Schritt für Schritt aufzunehmen und das Thema in einer verständlichen Weise darzustellen. Um dieses Ziel zu erreichen, wurden sowohl die Grundlagen aus Wikipedia als auch relevante Artikel aus dem Internet zur Gruppierung der Begriffe

⁷ <https://www.selenium.dev/documentation/webdriver/>

herangezogen. Eine ausführliche Recherche zum Thema Naturkatastrophen ermöglichte es uns, ein umfassendes Verständnis für die verschiedenen Aspekte von Katastrophen zu entwickeln. Es wurden Informationen über die Ursachen, Folgen und Präventionsmaßnahmen gesammelt und ausgewertet. Durch die Heranziehung zusätzlicher Ressourcen und Quellen war es auch möglich, aktuelle Ereignisse und Entwicklungen im Zusammenhang mit Naturkatastrophen zu berücksichtigen.

In der nächsten Phase der Bildung des semantischen Netzes wurde eine Suche nach Begriffen in einem Textkorpus durchgeführt. Geplant war die Erstellung von Listen verschiedener Wortarten, darunter Bigramme, Trigramme und Viergramme, mit Angabe der Häufigkeit ihres Auftretens. Das Hauptziel bestand darin, die Häufigkeit der Verwendung dieser Wortkombinationen im Zusammenhang mit unseren Schlüsselwörtern zu ermitteln.

Um die höchste Genauigkeit ohne übermäßige Datenmenge zu erreichen, wurden alle Subkorpora auf die Grundform des Wortes reduziert, d.h. lemmatisiert. Zu diesem Zweck wurde ein kleines Programm `lematizer.py` auf der Grundlage der `spaCy`-Bibliothek für die Verarbeitung natürlicher Sprache entwickelt, die hilfreiche Werkzeuge für die Tokenisierung, Lemmatisierung, Entitätsextraktion und andere textuelle Operationen bietet. Die `spaCy`-Bibliothek enthält verschiedene Sprachmodelle, die auf großen Datenmengen trainiert wurden und eine effiziente Verarbeitung von Texten in verschiedenen Sprachen ermöglichen.

Nach der Zusammenstellung der Texte wurden verschiedene Vorverarbeitungsschritte durchgeführt, um sie für die weitere Analyse vorzubereiten. Zunächst erfolgte eine Umwandlung aller Buchstaben in Kleinbuchstaben, um die Konsistenz zu gewährleisten. Die Entfernung von Satz- und Leerzeichen vereinfachte die Textstruktur.

Um den Text zu bereinigen und seine Qualität zu verbessern, wurde die Bibliothek `NLTK`⁸ (Natural Language Toolkit) verwendet. Diese Bibliothek diente dazu, Stoppwörter zu entfernen. Stoppwörter sind häufig vorkommende Wörter wie Präpositionen, Konjunktionen und Artikel, die in der Regel wenig aussagekräftige Informationen enthalten und für die Analyse nicht relevant sind. Das Entfernen dieser Stoppwörter ermöglichte es uns, den Text auf den relevanten Inhalt zu konzentrieren. Darüber hinaus setzten wir eine Liste mit irrelevanten Verben, Adjektiven und Substantiven zusammen und entfernten sie aus dem Text. Diese Wörter haben wenig Einfluss auf die Bedeutung und wurden als unwichtig angesehen. Durch die Entfernung dieser irrelevanten Wörter erhöhten sich die Genauigkeit und Zuverlässigkeit der anschließenden Analyse.

Die Vorverarbeitung des Textes ist ein notwendiger Schritt zur Gewinnung hochwertiger und relevanter Informationen. Durch die Anwendung dieser Schritte wurde der Text bereinigt und auf seinen Hauptinhalt reduziert, wodurch eine solide Grundlage für die weitere Analyse und Forschung geschaffen wurde.

Der nächste Schritt bestand darin, Listen mit Wortkombinationen unterschiedlicher Länge zu erstellen, um den Text besser verstehen zu können. Wir extrahierten Bigramme (Wortpaare), Trigramme (Tupel aus drei Wörtern) und Viergramme (Tupel aus vier Wörtern). Dazu haben wir ein speziell entwickeltes Programm `bigramm_regex_freq.py` (auch für Tri- und Viergramme entsprechend) auf jeden Subkorpus angewendet, das auf der `NLTK`-Bibliothek basiert. `NLTK` bietet eine breite Palette von Werkzeugen für die Verarbeitung natürlicher Sprache und ermöglicht die Erstellung von Phrasen auf der Grundlage der gegebenen Textdaten. Die generierten Listen enthielten lemmatisierte Wörter und Phrasen, z. B. Bigramme, und bestanden aus etwa 500-700 Wortpaaren pro Unterthema. Diese Zusammenstellung reduzierte die Datenmenge erheblich und beschleunigte die weiteren Verarbeitungsschritte. Durch die einheitliche Strukturierung der Textdaten waren wir in der Lage, die Informationen effektiver zu analysieren und für die weitere Verwendung nutzbar zu machen.

⁸ <https://www.nltk.org/>

Word	Frequency
('afad', 'katastrophenschutzbehörde'),	12
('katastrophe', 'mensch')	3
('folge', 'katastrophe')	2

Tabelle 4.1: Wortpaare mit ihren Häufigkeiten

Word	Frequency
('erdbeben', 'syrien', 'türkei')	44
('erdbeben', 'schweren', 'türkei')	7
('erdbeben-tote', 'versorgung', 'überlebenden')	5

Tabelle 4.2: Trigramme mit ihren Häufigkeiten

Word	Frequency
('hitze', 'pausen', 'schatten', 'tipps'),	14
('acht', 'extreme', 'hitzewellen', 'klimakrise')	4
('dürre', 'hitze', 'nehmen', 'waldbrände')	4

Tabelle 4.3: Viergramme mit ihren Häufigkeiten

Die Erstellung von Listen mit Phrasen ermöglichte es uns, über einzelne Wörter hinauszugehen und Verbindungen zwischen verschiedenen Begriffen zu erkennen.

Dieser Ansatz ermöglichte uns eine effizientere Nutzung der Daten und ein tieferes Verständnis des Textes. Die Erstellung von Listen mit Wortkombinationen war also ein wichtiger Schritt zur weiteren Analyse und Interpretation des Korpus.

Korpusinspektion

Zunächst erstellten wir kleine Basisdatenbanken auf der Grundlage unserer Tabellenkalkulationen und gingen dann zu einer der zeitaufwändigsten Phasen des Projekts über: der Suche nach relevanten und notwendigen Wörtern und deren Auswahl. Unser Ziel war es, spezifische Wörter und Phrasen auszuwählen, die sich direkt auf das Thema Naturkatastrophen beziehen. Diese Wörter sollten in der Lage sein, das Thema kurz und bündig zu beschreiben, ohne ganze Sätze zu erfordern. Es handelte sich um Aufbau eines Netzes mit nur 6 Knoten, das die Hauptbereiche des Themas beschrieb, nämlich: „Art von Naturkatastrophe“, „Naturkatastrophenursache“, „Katastrophenvorsorge und -planung“, „Rettung und Soforthilfe“, „Katastrophenschaden“; und als Ausgangspunkt für die Entwicklung der nachfolgenden Unterthemen diene.

Bei der Erstellung unserer Ontologie beschlossen wir auch, die Häufigkeit des Auftretens bestimmter Wörter und Phrasen in unseren Korpora zu berücksichtigen. Wie die Analyse gezeigt hat, liefert die Häufigkeit zwar einige aussagekräftige Merkmale, spielt aber keine entscheidende Rolle. Als wir alle Wortlisten bis zum Ende durchgingen, stellten wir fest, dass ebenso wichtige und aussagekräftige Ausdrücke die Häufigkeit N=1 haben, aber nicht ausgeschlossen werden können, da sie ein wesentlicher Bestandteil des Themas sind. Außerdem waren die Viergrammlisten unbrauchbar, weil fast identische Informationen bereits in den Bi- und Trigrammen enthalten waren.

Ein Großteil dieser Arbeit war eng mit dem Text selbst verbunden. In einigen Fällen war die genaue Bedeutung der ermittelten Ausdrücke nicht sofort klar, so dass wir uns den Kontext des Textes genauer ansehen mussten. Oft reichte es aus, ein Schlüsselwort im Text zu finden und den umgebenden Satz zu analysieren, um eine klarere Vorstellung von der Bedeutung und dem Kontext des Ausdrucks zu bekommen.

Durch die Berücksichtigung des Textzusammenhangs konnte man sich ein vollständigeres Bild von den Wörtern und Sätzen machen und entscheiden, ob sie in das semantische Netz aufgenommen werden sollten oder nicht. Ziel war es, sicherzustellen, dass die ausgewählten Ausdrücke tatsächlich zur Beschreibung und Kategorisierung des Themas Naturkatastrophen beitragen.

Da der Korpus hauptsächlich aus Nachrichtenartikeln besteht, die keine wissenschaftliche und theoretische Grundlage haben, war es notwendig, einige Begriffe zu gruppieren und manuell hierarchische Struktur aufbauen. Zu diesem Zweck wurden Google und [openthesaurus.de](https://www.openthesaurus.de/)⁹ als Hilfsmittel verwendet, um relevante Begriffe zu finden. Dadurch wird sichergestellt, dass die im semantischen Netz verwendeten Daten mit einer kurzen bis mittleren Erklärung versehen sind. Diese Begriffe sollten in der Lage sein, eine kurze Erklärung von 2-3 Wörtern für jeden Konzept zu liefern. Der Hauptvorteil dieses Ansatzes besteht darin, dass der Benutzer im Voraus eine Vorstellung von dem Thema erhält, mit dem er im semantischen Netz arbeiten wird. So lassen sich schnell logische Verbindungen zwischen verschiedenen Konzepten herstellen, und es entsteht ein verständliches Gesamtbild.

Durch die Gruppierung von Konzepten und die Verwendung klar definierter Erklärungen wird die Zugänglichkeit und Verständlichkeit des semantischen Netzes verbessert. So können die Nutzer schneller und effizienter Informationen finden und Zusammenhänge herstellen.

Bei der Erstellung der semantischen Netzstruktur war es besonders wichtig, den Wortformen und ihren Bedeutungen Aufmerksamkeit zu schenken. Dies war angesichts des linguistischen Aspekts und des Zwecks der Arbeit äußerst wichtig. Um Fehler zu vermeiden und genaue und zuverlässige Daten über die Formen der einzelnen Wörter zu erhalten, haben wir verschiedene Wörterbücher zu Hilfe genommen, nämlich das Wörterbuch der deutschen Rechtschreibung¹⁰ und das Übersetzungswörterbuch Lingvo¹¹. Die Berücksichtigung der verschiedenen Wortformen ermöglichte es uns, die grammatikalischen, semantischen und lexikalischen Aspekte des Themas vollständiger und genauer darzustellen. Dies führte zu einer genaueren Analyse und einem besseren Ergebnis der Textdaten. Die Nutzung externer Ressourcen wie Google und [openthesaurus.de](https://www.openthesaurus.de/) ermöglicht uns den Zugriff auf einen breiten Wissens- und Terminologiepool. Dadurch wird sichergestellt, dass die im semantischen Netz verwendeten Begriffe korrekt und genau sind.

4.3 Gesamtkodierung und Erstellung des semantischen Netzes

Der abschließende Schritt bei der Erstellung eines Lexikons für das Semantic Web war die Kodierung der Daten. Dieser Schritt war sehr notwendig, um die Informationen in einem Format darzustellen, das Computersysteme verstehen können. Wir haben uns für das JSON-Format entschieden, weil es uns eine effiziente Speicherung und Übertragung von Daten im semantischen Netz ermöglicht, das das Hauptziel unserer Arbeit ist.

Ursprünglich bestand das semantische Netz lediglich aus Konzepten (Abb. 4.1), die in einer hierarchischen Form mittels JSON dargestellt wurden. Dadurch waren die Beziehungen zwischen den Knoten bereits erkennbar, und man konnte eine grobe Vorstellung davon bekommen.

⁹ <https://www.openthesaurus.de/>

¹⁰ <https://www.duden.de/>

¹¹ <https://www.lingvolive.com/>

```

}
} "Naturkatastrophe": {
}   "Art von Naturkatastrophe": {
}     "Geologische": {...},
}     "Meteorologische": {
}       "Hurrikan": {...},
}       "Tornado": {...},
}       "Dürre": {...},
}       "Waldbrand": {...}
}     },
}     "Hydrologische": {...},
}     "Klimabedingte": {...},
}     "Biologische": {...},
}     "Extraterrestrische": {...}
}   },
}   "Naturkatastrophenursache": {...},
}   "Katastrophenvorsorge und -planung": {
}     "Frühwarnsystem": {...},
}     "Evakuierung": {...},
}     "Katastrophennotfallplan": {...},
}

```

Abb. 4.1 grobe Darstellung des semantischen Netzes

Am Ende hat das Lexikon insgesamt 369 Knoten beinhaltet, die als Konzepte dienen. Jeder Knoten wird in einem JSON-Objekt dargestellt, das verschiedene Attribute enthält, um die Eigenschaften des Konzepts umfassend zu beschreiben. Dazu gehören eine eindeutige ID, der Name des Konzepts, Synonyme, mögliche Wortformen, der Konzepttyp (extern oder intern) und eine übergeordnete ID, die die Beziehung zu anderen Konzepten angibt (Abb. 4.2). Synonyme werden in Form eines Arrays dargestellt und werden verwendet, um alternative Bezeichnungen oder Begriffe darzustellen, die in der Literatur oder in Bezug auf das Konzept benutzt werden. Eine solche Vielfalt von Ausdrücken ermöglicht eine vollständigere Darstellung der grammatikalischen, semantischen und lexikalischen Aspekte des Themas. Ein wichtiger Aspekt bei der Strukturierung des semantischen Netzes war die Unterscheidung zwischen internen und externen Knotenpunkten. Interne Knoten sind eng mit dem Thema verbunden und bilden die Schlüsselbegriffe des Lexikons. Externe Knoten hingegen dienen dazu, Verbindungen zu anderen Wissensgebieten oder zur "Außenwelt" herzustellen und damit die Reichweite des semantischen Netzes zu erweitern. Die Verknüpfungen zwischen den Knoten wurden durch die Angabe eines übergeordneten Identifikators hergestellt, was zu einer hierarchischen Struktur führte. Jeder Knoten hatte seine eigenen übergeordneten Knoten, was eine klare Organisation und ein besseres Verständnis der Beziehungen zwischen den Konzepten ermöglichte.

```
{
  "id": 2,
  "word": "Geologische",
  "parentId": 1,
  "type": "intern",
  "synonyms": [
    "Endogene",
    "tektonische"
  ],
  "wordForms": [
    "geologisch",
    "geologischer",
    "geologische",
    "geologisches"
  ]
},
```

Abb. 4.2: JSON-Kodierung für jedes Konzept

Durch die Kodierung der Daten im JSON-Format wurde eine strukturierte Darstellung erreicht, die das Finden, Vergleichen und Benutzen der Konzepte deutlich macht. Dies trägt zu einem besseren Verständnis der Themen bei und fördert das effektive Funktionieren des semantischen Netzes. Die sorgfältige Berücksichtigung der Wortformen und ihrer Bedeutungen spielte eine zentrale Rolle für die genaue Analyse und das Verständnis der Textdaten.

Schlussfolgend bestand Arbeit darin das semantische Netz zu visualisieren. Hier kam zum Einsatz VOWL (The Visual Notation for OWL). VOWL ist eine Visualisierungssprache, die speziell für Benutzer zur Darstellung von Ontologien entwickelt wurde. Sie bietet grafische Darstellungen für die verschiedenen Elemente von OWL, die in einer grafischen Struktur organisiert sind. Sie schafft eine intuitive und vollständige Darstellung, die den Benutzern hilft, Ontologien auch ohne vorherige Erfahrung zu verstehen. (S. Lohmann, S. Negru, F. Haag, T. Ertl, 2015).

Abschließend bestand die Aufgabe darin, das semantische Netz zum Thema "Naturkatastrophen" visuell darzustellen. Hierfür wurde die spezielle Visualisierungssprache VOWL (The Visual Notation for OWL) eingesetzt. VOWL wurde entwickelt, um Benutzern die Darstellung von Ontologien auf intuitive und verständliche Weise zu ermöglichen. Mit Hilfe dieser Sprache wurden die verschiedenen Elemente der OWL-Ontologie in einer strukturierten grafischen Form angeordnet (Abb. 4.3).

Durch die Generierung einer fertigen JSON-Datei, die das vollständige semantische Netz enthält, und der Verwendung der VOWL-Spezifikation¹² konnte eine Datei mit dem Namen "naturkatastrophen.json" erzeugt werden. Diese Datei ist von Bedeutung für zukünftige Browser-Anwendungen und stellt das semantische Netz grafisch dar. Dabei werden sowohl interne als auch externe Knoten deutlich differenziert (dunkelblau markiert). Durch Anklicken der Knoten können auch Synonyme angezeigt werden (Abb. 4.4, hier als "Comment" bezeichnet).

¹² <http://vowl.visualdataweb.org/v2/>

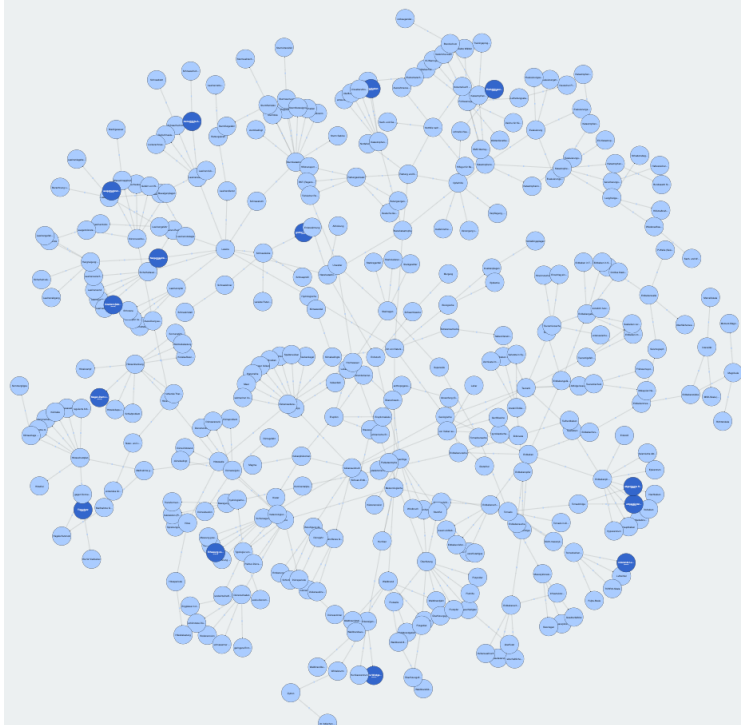


Abb. 4.3, gesamtes semantisches Netz

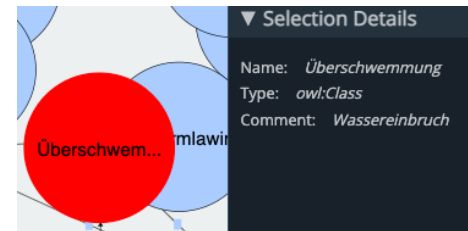


Abb. 4.4, Darstellung jedes Konzeptes in VOWL-Notation

Das Ergebnis dieser Arbeit ist nun über GitHub zugänglich, was es ermöglicht, das semantische Netz im Browser zu öffnen und zu erkunden. Dadurch wird Benutzern eine interaktive Erfahrung geboten, um die Konzepte und Beziehungen im Bereich der Naturkatastrophen besser zu verstehen.

4.4 Evaluierung des semantischen Netztes

Nachdem das semantische Netz erstellt war, stellte sich die Frage nach seiner Qualität. Um es zu bewerten, wendeten wir eine bekannte Vergleichsmethode an, die darin bestand, das Vorkommen aller Knoten des Netzes in verschiedenen Texten zu analysieren. Für dieses Verfahren wurden alle Konzepte (Knoten) des Netzwerks in einer Textdatei erfasst und drei Bücher ausgewählt, die sich mit Naturkatastrophen befassen. Mit einem einfachen Python-Code **evaluation_tool.py** fand unser Programm problemlos alle Wörter in diesen Büchern, die auch zu unserem semantischen Netzwerk passten. Hier wurden nur die Knoten (also Konzepte im Singular) bei der Analyse berücksichtigt. Eine interessante Beobachtung war, dass jedes der drei Bücher völlig unterschiedliche Ergebnisse lieferte. In einem Buch wurden nur ca. 13% der Wörter gefunden, die mit unseren Konzepten übereinstimmten, während das Ergebnis in dem anderen Buch bei 30% lag. Dies deutet auf eine unterschiedliche Verwendung der Terminologie in den verschiedenen Quellen und auf eine Mehrdeutigkeit der Beziehungen zwischen Konzepten im semantischen Web hin.

Außerdem haben wir festgestellt, dass die Ergebnisse durch eine Erweiterung der Begriffsliste verbessert werden können. Wir fügten Teilformen einiger Wörter wie "beben" oder "versichern" sowie deren Pluralformen wie "schäden" oder "wirbelstürme" hinzu. Obwohl diese Ergänzungen keinen signifikanten Einfluss auf die Gesamtstatistik hatten, verbesserten sie trotzdem die Übereinstimmung der Wörter in den Texten mit unserem Netz.

Um schließlich das semantische Netz auf der Grundlage der drei Texte zusammenzufassen, wurden alle drei Wortlisten mithilfe von dem Programm **compare_evaluation_results.py** miteinander verglichen und die Schnittmenge ermittelt, d. h. die häufigsten Wörter, die in allen drei Texten vorkommen. Die Ergebnisse waren jedoch viel schlechter als erwartet: Von 368 Knoten im Netzwerk waren nur 36 in allen drei Texten vorhanden. Dies deutet darauf hin, dass unser semantisches Netz noch verbessert und erweitert werden könnte, um die Verbindungen zwischen Konzepten im Zusammenhang mit Naturkatastrophen besser widerzuspiegeln.

Ein möglicher Grund für die niedrigen Ergebnisse ist, dass wir in unserem semantischen Netz Begriffe verwendet haben, die sich ausschließlich auf das Thema Naturkatastrophen beziehen. Das bedeutet, dass diese Begriffe bei einer normalen Suche direkt mit diesem Thema verbunden wären und eine hohe Relevanz besitzen würden.

Bei der Nutzung des semantischen Webs haben wir versucht, die Suche zu verbessern und zu verfeinern, indem nicht nur die direkten Verbindungen, sondern auch die kontextuellen Beziehungen zwischen den Begriffen in Betracht gezogen wurden. Infolgedessen wurden einige allgemeinere oder abstraktere Begriffe, die mit Naturkatastrophen in Zusammenhang stehen könnten, möglicherweise aus der Liste der Begriffe ausgeschlossen.

Diese Einschränkung des Netzes könnte zu einer unzureichenden Textabdeckung und folglich zu niedrigen Übereinstimmungsstatistiken führen. Um die Ergebnisse und Statistiken zu verbessern, kann die Aufnahme solcher allgemeiner Begriffe oder der Einsatz zusätzlicher Kontextanalyseverfahren zur Erweiterung und Verfeinerung des semantischen Netzes in Betracht gezogen werden. Dies wird die Genauigkeit und Vollständigkeit der Suche in einer Vielzahl von Texten über Naturkatastrophen verbessern.

5 Diskussion und Zusammenfassung

5.1 Bewertung der Stärken und Schwächen des semantischen Netzes und der vorgestellten Methoden

Mit Blick auf die Entwicklung und Bewertung des semantischen Webs und die vorgestellten Methoden zu seiner Erstellung ist es an der Zeit, seine Stärken und Schwächen näher zu betrachten. Im Hinblick auf die Genauigkeit der Ontologie ist es wichtig zu betonen, dass die Konzepte und deren Beziehungen basierend auf einem umfangreichen Korpus von Nachrichtenartikeln entwickelt wurden, wobei der Schwerpunkt auf den Lesern dieser Artikel lag. Diese Vorgehensweise wurde gewählt, um eine präzise und korrekte Darstellung der Konzepte zu gewährleisten, die im Zusammenhang mit Naturkatastrophen im Alltag auftreten.

Ein weiterer Aspekt ist die Vollständigkeit der Ontologie. Wir haben versucht, alle relevanten Konzepte und ihre Zusammenhänge zu erfassen. Um mögliche Lücken oder fehlende Informationen zu vermeiden, wurden zusätzliche Recherchen im Internet durchgeführt. Fast alle Konzepte wurden überprüft, analysiert und in das Netz integriert, um eine logische und verständliche Struktur zu gewährleisten.

Die Skalierbarkeit des semantischen Netzes ist ein weiterer wichtiger Faktor. Es sollte leicht erweiterbar sein und sich an neue Informationen anpassen können, um zukünftigen Anforderungen gerecht zu werden. Die Grundlage dafür kann eine vielfältige Fachliteratur sein. Es ist jedoch zu beachten, dass im Zuge der Erweiterung des Netzes Änderungen der Struktur und der Konzepte erforderlich sein können. Ein Beispiel dafür ist die Anpassung von Begriffen wie "Opfer", die je nach Kontext unterschiedliche Bezeichnungen wie "Erdbebenopfer" oder "Flutopfer" haben können.

Der Schwerpunkt liegt auf der Benutzerfreundlichkeit des semantischen Netzes, um sicherzustellen, dass die Benutzer einen einfachen Zugang und ein klares Verständnis haben. Die Konzepte und Beziehungen im Netz sind intuitiv und logisch strukturiert und bieten einen allgemeinen Überblick über das Thema "Naturkatastrophen". Informationen über Ursachen, Auswirkungen, Maßnahmen und größere Katastrophen sind leicht zugänglich.

Die Wirksamkeit des semantischen Netzes spielt ebenfalls eine wichtige Rolle. Durch eine geeignete Gruppierung der Begriffe gewährleistet das Netz eine schnelle Suche und eine effiziente Bearbeitung der Informationen.

Hinsichtlich der Zuverlässigkeit des semantischen Netzes ist anzumerken, dass der Einsatz von Experten aus dem jeweiligen Fachgebiet ideal wäre, um ein vollständiges und zu 100 % korrektes Netz zu erstellen. Da das Netz jedoch auf allgemeinen Daten beruht, kann eine absolute Zuverlässigkeit und Genauigkeit nicht erwartet werden.

Zusammenfassend lässt sich sagen, dass ein semantisches Netz leicht erweitert werden kann, um neue Konzepte, Beziehungen oder Bereiche abzudecken. Vor allem in Bezug auf die Verbindung mit anderen Bereichen, wie der Geologie, kann das Netz als Teil eines größeren Netzes funktionieren, welches Naturereignisse umfassend beschreibt.

5.2 Zusammenfassung der wichtigsten Ergebnisse und Erkenntnisse

Unser Ziel war der Aufbau eines semantischen Netzes zum Thema Naturkatastrophen. Dazu haben wir ein Korpus verwendet, das uns Informationen über aktuelle Naturkatastrophen aus verschiedenen Nachrichtenquellen lieferte. Die Informationen wurden mithilfe von Crawler-Algorithmen extrahiert, um die Textverarbeitung effizienter zu gestalten.

Wir haben verschiedene computerlinguistische Methoden zur Verarbeitung der Informationen eingesetzt. Mithilfe von NLTK-Bibliotheken haben wir Frequenzlisten erstellt, indem wir Schlüsselwörter in verschiedenen Formen berücksichtigt haben. Dabei entfernten wir nicht nur die

von NLTK vorgeschlagenen Stoppwörter, sondern filterten auch unnötige Wortarten und -formen heraus. Dies führte zu guten und genauen Ergebnissen, die uns halfen, ein relevantes Wörterbuch zu erstellen. Wir haben auch reguläre Ausdrücke verwendet, insbesondere für komplexe Wörter, wie in den Beispielen unten gezeigt. Mit Hilfe der manuellen Überprüfung konnten wir analysieren, welche Ausdrücke im Bereich Naturkatastrophen vorkommen. Dazu gehörten z. B. domänenspezifische Namen wie "Erdbeben in der Türkei", technische Begriffe wie "Nachbeben", Berufs-/Vereinsbegriffe wie "Luftrettungsstaffel", juristische Begriffe wie "Katastrophenschutz" und domänenspezifische "Entitäten" wie "Katastrophenschutzbehörde". Dieser Ansatz ermöglichte es uns, eine solide Grundlage für das semantische Netz zu schaffen.

Danach begann die wichtigste Phase unserer Arbeit — die manuelle Analyse und Bearbeitung der entfernten Wörter und Sätze. Hunderte von Zeilen wurden sorgfältig geprüft und ausgewählt. Dieser Schritt machte uns einmal mehr klar, dass wir den Korpus in dieser Phase auf keinen Fall vernachlässigen durften. Der Korpus lieferte uns detailliertere Informationen über Wortpaare und ermöglichte es uns, fast ganze Sätze von Mehrwortlexemen zu erfassen, um die genaue Bedeutung der einzelnen Konzepte zu beschreiben. Da die Sprache ein zentraler Aspekt unserer Arbeit ist, wurden auch verschiedene Formen von Wörtern berücksichtigt und für das künftige Netz vorbereitet. Wir haben auch Rechtschreibwörterbücher und Synonymwörterbücher verwendet, um umfassende Informationen zu den einzelnen Begriffen zu erhalten.

Alle diese Informationen wurden in ein computerlesbares JSON-Format übertragen. Für unsere Studie erwies sich dieses Format als einfach zu verwenden und zu verstehen. Jedes Konzept wird als JSON-Objekt dargestellt, das uns Informationen wie einen Bezeichner, Synonyme, Wortformen und einen übergeordneten Bezeichner liefert, wodurch wir ein vollständiges Bild des Konzepts erhalten. Dies ermöglicht es, das semantische Netz auf eine leicht zugängliche und visuell ansprechende Weise zu präsentieren.

Als Resultat haben wir ein semantisches Netz mit 368, davon 14 als Extern sind, Knoten erstellt, das uns ausreichende und effiziente Information über die Domäne liefert. Das Netz wird auf Basis von VOWL gebaut und kann interaktiv genutzt werden.

Insgesamt kann man sagen, dass Ontologie und Erstellung eines semantischen Netzes auf diese Art und Weise gut möglich sind. Die einzigen Schwierigkeiten kommen nur dann, wenn man als nicht Expert in diesem Thema ist und Entscheidungen wegen der Domänenzugehörigkeit treffen muss. Auch hat diese Arbeit dazu beigetragen und uns gezeigt, dass manuelle Arbeit ein der wichtigste Teil ist. Der Grund dafür ist klarer menschlicher Verstand, das die Sachen logisch und ordentlich strukturieren kann. Dies ist noch ein Hinweis dazu, dass kein Tool das besser machen kann, als menschliches Gehirn.

Generell lässt sich sagen, dass die Erstellung einer Ontologie und eines semantischen Webs auf diese Weise erfolgreich umgesetzt werden kann. Allerdings gibt es einige Schwierigkeiten, vor allem wenn eine Person kein Experte auf dem Gebiet ist und Entscheidungen über die Zugehörigkeit zu einem bestimmten Bereich treffen muss. Es hat sich gezeigt, dass manuelle Arbeit ein wichtiger Teil dieses Prozesses ist. Die menschliche Denkfähigkeit ermöglicht es uns, Konzepte auf logische und strukturierte Weise zu organisieren. Dies unterstreicht die Tatsache, dass kein Werkzeug diese Aufgabe besser erfüllen kann als das menschliche Gehirn.

6 Schlussfolgerungen

Ausblick auf zukünftige Forschungsarbeiten und Anwendungsmöglichkeiten

Die Analyse semantischer Netze zur Vorhersage und Bewertung von Naturkatastrophen ist eine echte Herausforderung für Wissenschaftler und Experten, die sich auf Naturgefahren spezialisiert haben. Semantische Netzstrukturen ermöglichen es uns, die Beziehungen zwischen Konzepten zu modellieren und Informationen auf der Grundlage ihrer semantischen Konnektivität zu analysieren. Eine der wichtigsten Anwendungen semantischer Netze im Zusammenhang mit Naturkatastrophen ist die Vorhersage von Ereignissen und die Bewertung ihrer möglichen Folgen. Semantische Netze ermöglichen die Analyse verschiedener Faktoren, wie geografische, klimatische, hydrometeorologische usw., um Signale und Muster zu erkennen, die Naturkatastrophen vorausgehen. Mit Hilfe von Netzen lassen sich beispielsweise Beziehungen zwischen den Wetterbedingungen und der Wahrscheinlichkeit bestimmter Katastrophen wie Wirbelstürme, Erdbeben oder Überschwemmungen herstellen.

Die Nutzung semantischer Netze zur Vorhersage und Bewertung der Auswirkungen von Naturkatastrophen ist ein bedeutendes Forschungsgebiet, das zunehmend die Aufmerksamkeit von Wissenschaftlern und Experten auf sich zieht. Diese Netze ermöglichen es, mehrere Faktoren wie Standort, Topografie, Bevölkerungszentren und Infrastruktur zu verknüpfen, die das Ausmaß und die Auswirkungen einer Naturkatastrophe beeinflussen können. Durch die Analyse von Informationen aus verschiedenen Quellen und die Berücksichtigung der semantischen Beziehungen zwischen diesen Faktoren können semantische Netze zur Risikobewertung und zur Entwicklung von Krisenmanagementstrategien beitragen.

Semantische Netze können verschiedene Arten von Daten integrieren, z. B. geografische Karten, Satellitenbilder, Statistiken usw. Auf diese Weise lassen sich komplexe Modelle erstellen, die Informationen aus verschiedenen Quellen kombinieren und so genauere Vorhersagen und Bewertungen von Naturkatastrophen ermöglichen.

Es gibt jedoch noch einige Herausforderungen für Computerlinguisten und Wissenschaftler im Zusammenhang mit dem Einsatz semantischer Netze bei der Vorhersage und Bewertung von Naturkatastrophen. So ist beispielsweise eine erhebliche Datenmenge zu analysieren, die eine hohe Genauigkeit und Zuverlässigkeit garantiert, und die Abstimmung des semantischen Netzes ist eine Mindestanforderung, um zuverlässige Ergebnisse zu erzielen. Darüber hinaus muss das semantische Netz ständig aktualisiert und an veränderte Bedingungen und neue Daten angepasst werden.

Der Einsatz semantischer Netze bei der Vorhersage und Bewertung von Naturkatastrophen ist daher ein vielversprechender Forschungsbereich, der das Katastrophenmanagement deutlich verbessern und Risiken verringern kann. Sie kann auch die Entwicklung semantischer Netze in verschiedenen Bereichen vorantreiben, die in Zukunft das Leben von Millionen von Menschen nicht nur verbessern, sondern auch erleichtern und zu neuen wissenschaftlichen Durchbrüchen führen können.

7 Bibliography

- H.Bouyerbou, K.Bechkoum, R.Lepage. (March 2019). *International Journal of Disaster Risk Reduction*. Abgerufen am May 2023 von Geographic ontology for major disasters: Methodology and implementation:
<https://www.sciencedirect.com/science/article/abs/pii/S221242091830476X?via%3Dihub>
- Sowa, J. b. (1991). *INTRODUCTION TO ISSUES IN KNOWLEDGE REPRESENTATION*. Abgerufen am May 2023 von Principles of Semantic Networks:
<https://www.sciencedirect.com/science/article/pii/B9781483207711500059?via%3Dihub>
- Abdel-Badeeh M.Salem, Marco Alfonse. (23-25. July 2008). *Ontology versus semantic networks for medical knowledge representation*. Abgerufen am May 2023 von ResearchGate:
https://www.researchgate.net/publication/228677121_Ontology_versus_semantic_networks_for_medical_knowledge_representation
- Рогущина, Ю. (2016). *ВИКОРИСТАННЯ ОНТОЛОГІЧНОГО АНАЛІЗУ ДЛЯ ВИЗНАЧЕННЯ КОМПЕТЕНТНОСТІ НАУКОВЦІВ В ПЕВНІЙ ПРЕДМЕТНІЙ ОБЛАСТІ*. Abgerufen am April 2023 von Проблеми програмування:
<https://core.ac.uk/download/pdf/132546583.pdf>
- ШТУЧНИЙ ІНТЕЛЕКТ В ЗАДАЧАХ КІБЕРБЕЗПЕКИ*. (26. 10 2021). Abgerufen am May 2023 von Штучний інтелект в задачах кібербезпеки (КБ-19-1, КБ-19-2):
https://learn.ztu.edu.ua/pluginfile.php/188961/mod_resource/content/1/%D0%A8%D0%86%D0%9A%D0%91_%D0%9B-8_%D0%97%D0%BD%D0%B0%D0%BD%D0%BD%D1%8F.pdf
- Dengel, A. (Hrsg.). (2012). *Semantische Technologien Grundlagen. Konzepte. Anwendungen*. Heidelberg: Spektrum Akademischer Verlag. Von SpringerLink:
<https://link.springer.com/book/10.1007/978-3-8274-2664-2> abgerufen
- СЕМАНТИЧНІ МЕРЕЖІ ПСШІ ТА ЇХ КЛАСИФІКАЦІЯ*. (2022). Abgerufen am May 2023 von Бібліотека підручників та статей Posibniki: <https://posibniki.com.ua/post-semantichni-merezhi-psshi-ta-yih-klasifikaciya>
- Nova. (4. March 2023). *Understanding Semantic Networks in Artificial Intelligence*. Abgerufen am April 2023 von AITechTrend: <https://aitechtrend.com/understanding-semantic-networks-in-artificial-intelligence/>
- M.Hsing, and A.Cherkasov. (2006). *Integration of Biological Data with Semantic Networks*. (Bentham Science Publishers Ltd.) Abgerufen am May 2023 von Current Bioinformatics:
<https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=b476898f92563fe140f63d9fad7af5d81d8f9326>
- He, P. (2014). *Counter Cyber Attacks By Semantic Networks*. Abgerufen am May 2023 von ScienceDirect:
<https://www.sciencedirect.com/science/article/pii/B978012411474600027X?via%3Dihub>
- І.Кравець. (n.d.). *НАЙБІЛЬШІ В СВІТІ КАТАСТРОФИ, СТИХІЙНІ ЛИХА І ЕПІДЕМІЇ*. Retrieved April, 2023 from <https://sno.udpu.edu.ua/index.php/naukovo-metodychna-robota/95-innovatsiini-pedahohichni-tekhnohii-v-zahalnoosvitnii-shkoli-teoriia-ta-praktyka/423-najbilshi-v-sviti-katastrofi-stikhijni-likha-i-epidemiji>
- Natural Disasters- Cause and Effects*. (kein Datum). Abgerufen am April 2023 von TargetStudy:
<https://targetstudy.com/articles/natural-disasters-cause-and-effects.html>
- Metych, M. (2023, Jan 11). *natural disaster*. Retrieved April, 2023 from Encyclopedia Britannica:
<https://www.britannica.com/science/natural-disaster>
- S.Grimm, P.Hitzler, A.Abecker. (kein Datum). *Knowledge Representation and Ontologies*. Abgerufen am May 2023 von
https://www.aifb.kit.edu/images/9/94/2007_1644_Grimm_Knowledge_Repre_2.pdf
- Андрєєв, А. (2021). *ОСОБЛИВОСТІ СТВОРЕННЯ СЕМАНТИЧНИХ МЕРЕЖ*. Abgerufen am April 2023 von AUTOMATION AND DEVELOPMENT OF ELECTRONIC DEVICES:
<https://openarchive.nure.ua/server/api/core/bitstreams/66f3c0d0-7cfe-404b-b40e-fb92b3049d2e/content>

Азаров С. І., Задунай О. С. . (26. 12 2019). *Аналіз природних катастроф та їх впливу на довкілля*. Abgerufen am April 2023 von Екологічна безпека та природокористування: http://nbuv.gov.ua/UJRN/ebpk_2019_4_8