

Lerneinheit: Netzwerkanalyse mit Gephi lehren

Mareike Schumacher  ¹

1. Universität Regensburg

forTEXT

Thema:	Netzwerkanalyse	DOI:	10.48694/fortext.3780
Jahrgang:	1	Ausgabe:	6
Erscheinungsdatum:	2024-08-30	Erstveröffentlichung:	2019-10-21 auf fortext.net
Lizenz:			open  access

Allgemeiner Hinweis: Rot dargestellte Begriffe werden im Glossar am Ende des Beitrags erläutert. Alle externen Links sind auch am Ende des Beitrags aufgeführt.

Eckdaten des Lehrmoduls

- Thema der Sitzung: Figurennetzwerke in den Dramen Karoline von Günderrodes
- Lernziele: Kenntnisse über die Methode der Netzwerkanalyse, sicherer Umgang mit Gephi, kritische Bewertung der Methode, Autorinnenkenntnis und Einblicke in das Dramenwerk einer Schriftstellerin
- Phasen: Vorstellung und Diskussion der Methode, Demonstration der Toolfunktionen (vgl. **Feature**), Gruppenarbeit, Gruppenpräsentation, Abschlussdiskussion
- Sozialformen: Vortrag, Gruppenarbeit, Diskussion
- Medien/Materialien: Alle Lernenden müssen einen Laptop haben, auf dem Gephi installiert ist; Lehrende benötigen einen Laptop und einen Beamer
- Dauer des Lehrmoduls: 2 x 90 Minuten
- Schwierigkeitsgrad des Tools: mittel

Bausteine

- Verlaufsraaster des Lehrmoduls: Aus welchen Phasen setzt sich das Lehrmodul zusammen? Dem Verlaufsplan entnehmen Sie Inhalte und Schwerpunkte.
- Anwendungsbeispiel: Anhand welcher Texte unterrichten Sie Netzwerkanalyse? Leiten Sie die Studierenden dazu an, Figurennetzwerke der Dramen Karoline von Günderrodes zu visualisieren und zu explorieren.
- Verlauf der Unterrichtseinheit(en): Wie sieht die konkrete Ausgestaltung der Phasen aus und welche Arbeitsschritte werden vorgenommen? Erfahren Sie, wie die Unterrichtseinheit strukturiert ist und welche Beispielaufgaben Sie Ihren Studierenden stellen können.
- Lösungen zu den Beispielaufgaben: Hat die Lerngruppe die Beispielaufgaben richtig gelöst? Hier erfahren Sie, wie Sie die Antworten erhalten.

Alle Materialien zu diesem Lehrmodul finden Sie auf Zenodo über [diesen Link](#).

Verlaufsrafter des Lehrmoduls

Phase	Impulse des/der Lehrenden	Erwartete Aktivität der Lernenden	Sozialform	Medien / Materialien
Vorab und Einstieg (ca. 15 Minuten)	Was ist Netzwerkanalyse? In welchen Disziplinen ist die Methode verankert?	Vorab: Methodeneintrag Netzwerkanalyse (Schumacher 2024a), Video-Fallstudie (forTEXT 2019d) zur Netzwerkanalyse, Lerneinheit Netzwerkanalyse mit Gephi (Schumacher 2024b) und Video-Tutorials (forTEXT 2019a; forTEXT 2019b; forTEXT 2019c; forTEXT 2019e) zur Netzwerkanalyse; Beteiligung an der Diskussion	Diskussion im Plenum	Beamer, Laptop
Problematisierung (ca. 15 Minuten)	Was wird in der Netzwerkanalyse betrachtet? Wie verhalten sich diese „Daten“ zum Text? Wie ist diese Komplexitätsreduktion begründet?	Beteiligung an der Diskussion; Rückbezug auf Methodeneintrag	Diskussion im Plenum	Beamer, Laptop
Erarbeitung (ca. 60 Minuten)	Vorstellung der Toolfunktionen; Betreuung der Kleingruppen	Hands-on Netzwerkanalyse in Kleingruppen; Vorbereitung einer Präsentation als Hausaufgabe	Lehrvortrag und Gruppenarbeit	Beamer, Laptop, Gephi, Korpus, Netzwerkdaten
Sicherung (ca. 60 Minuten)	Moderation der Gruppenpräsentationen	Präsentation der Gruppenarbeitsergebnisse	Präsentation	Beamer, Laptop, Gephi, Korpus, Netzwerkdaten
Reflexion & Transfer (ca. 30 Minuten)	Diskussion von Schwierigkeiten, Impulse für Transfer geben	Ergebnisse und Schwierigkeiten aus den Gruppenpräsentationen diskutieren	Diskussion im Plenum	Beamer, Laptops

Das Verlaufsrafter stellen wir Ihnen zum Download [auf Zenodo](#) zur Verfügung.

1. Anwendungsbeispiel

In zwei Seminarsitzungen werden Sie die Methode der Netzwerkanalyse (Schumacher 2024a) lehren. Sie übertragen (vgl. **Domäneadaption**) damit ein ursprünglich sozialwissenschaftliches Verfahren auf einen literaturwissenschaftlichen Anwendungsbereich. Die Studierenden probieren die Methode an einem Dramenkorporus (vgl. **Korpus**) der Schriftstellerin Karoline von Günderrode aus und wenden das Netzwerkanalysetool Gephi (Schumacher 2024c) praktisch an. Zusätzlich erlernen sie den Umgang mit der Dramen-Netzwerkdatenbank DraCor (Horstmann 2024). Schließlich erarbeiten die Studierenden in Gruppen kleinere Fallstudien, die sie dem gesamten Seminar präsentieren und zur Diskussion stellen.

2. Verlauf der Unterrichtseinheiten

2.1 Vorarbeiten

Die Studierenden sollten vorab den Methodeneintrag Netzwerkanalyse (Schumacher 2024a) und den Tooleintrag Gephi (Schumacher 2024c) gelesen sowie das Fallstudien-Video Analyse der Figurennetzwerke in Lessings Emilia Galotti (forTEXT 2019d) geschaut haben. Außerdem sollten sie mithilfe der Tutorial-Videos (forTEXT 2019a; forTEXT 2019b; forTEXT 2019c; forTEXT 2019e) Gephi installiert und die Lerneinheit Netzwerkanalyse mit Gephi (Schumacher 2024b) gemacht haben. Lassen Sie die Studierenden vorab die Netzwerkdaten im GEXF-Format zu Nikator, Magie und Schicksal und Udohla von Karoline von Günderrode herunterladen. Da in allen drei Dramen die Bedeutung von weltlicher Macht der von emotionalen Beziehungen gegenüber gestellt wird, eignen sich diese Dramen besonders für eine vielfältige interpretationsgeleitete Betrachtung literarischer Werke mittels Netzwerkanalyse. Wenn Sie mit anderen Texten arbeiten wollen, finden Sie zahlreiche für die Netzwerkanalyse aufbereitete Datensätze ebenfalls bei DraCor (Horstmann 2024). Sollten Sie Ihre Seminarsitzung nicht in einem Computerlabor abhalten, erinnern Sie Ihre Studierenden vor der Sitzung noch einmal daran, ihren eigenen Laptop mitzubringen. Eine eins-zu-eins-Ausstattung ist für das Erlernen der Funktionen am sinnvollsten. Mehr als zwei Studierende sollten auf keinen Fall an einem Laptop arbeiten. Gephi ist weder für Smartphones noch für Tablets (oder andere Touch-Geräte) geeignet.

Je nachdem, welchen Lehrstil Sie bevorzugen, können Sie für den Einstieg einige Folien zur Methode der Netzwerkanalyse vorbereiten. Um Ihnen die Vorbereitung zu erleichtern, haben wir Beispielfolien entworfen, an denen Sie sich orientieren können.

2.2 Einstieg und Problematisierung

Zu Beginn der Unterrichtseinheit sollten Sie transparent machen, was die Studierenden in der über zwei Sitzungen gehenden Einheit zur Netzwerkanalyse erwartet:

- In der ersten Sitzung wird die Methode vorgestellt und kritisch hinterfragt, kleinere Projekte für die Gruppenarbeit werden entworfen, Teams gebildet und die Arbeit an den Fallstudien begonnen.
- In der zweiten Sitzung werden die Fallstudien in Form von Ergebnispräsentationen vorgestellt. Im Anschluss werden Besonderheiten der Methode diskutiert und Erfahrungen aus der Projektarbeit ausgetauscht.

Die Netzwerkanalyse ist ein Verfahren, das in den Sozialwissenschaften entwickelt wurde, inzwischen aber in zahlreichen Disziplinen verankert ist. In einem literaturwissenschaftlichen Seminar bietet es sich an, diesen Weg noch einmal nachzuzeichnen und die Parallelen zwischen der Analyse sozialer Gruppen und literarischer Figuren hervorzuheben. Darüber hinaus ist es wichtig, die grundlegenden Parameter und deren Bedeutung zu erläutern. Machen Sie den Studierenden deutlich, dass in der Netzwerkanalyse zu Beginn immer die Reduktion auf Knoten (Nodes) und Verbindungen (Edges) steht. Wie diese Nodes und Edges definiert werden, ist aber völlig frei wählbar. Fragen Sie die Studierenden, welche Art Netzwerke (neben sozialen Gruppen) ihnen noch einfallen, und sammeln Sie diese Ideen. Heben Sie hervor, dass Netzwerkanalyse kein ausschließlich digitales Verfahren ist, sondern dass netzwerkartige Strukturen ein Denkmodell sind, das auch manuell erstellt werden kann. Als rein mentales Hilfsmittel unterstützt es z. B. dabei, einen Text schematisch zu erfassen und sich dessen Strukturen vorzustellen. Fordern Sie die Studierenden auf, an die bisher im Seminar besprochenen Primärtexte zu denken und darin Netzwerke zu finden. Dies könnten z. B. Machtstrukturen, Diskurse, familiäre Verbindungen oder auch räumliche Systeme sein.

Gehen Sie dann langsam in eine Phase kritischer Betrachtung über. Fragen Sie die Studierenden nach den Vor- und Nachteilen der Reduktion auf die zwei Parameter der Nodes und Edges. Geben Sie ihnen auch die Gelegenheit, Fragen zum Methodeneintrag Netzwerkanalyse (Schumacher 2024a) und der Fallstudie im Video (forTEXT 2019d) zu stellen. Durch die Konzentration auf Knotenpunkte und Verbindungen ist es nötig, Texte sehr genau auf diese Parameter zu untersuchen: Aus dem Text werden Netzwerkdaten (vgl. Metadaten) herausgefiltert. Diese Daten umfassen Figurennamen, Figurengender, die Anzahl der gesprochenen Worte jeder Figur und die gemeinsamen Auftritte der Figuren. Stellen Sie sicher, dass die Studierenden verstehen, dass diese Daten nur einen Teil der Informationen darstellen, die in einem literarischen Text enthalten sind. Machen Sie auch deutlich, dass das Herausfiltern von Daten immer ein interpretativer Vorgang ist und dass es so etwas wie „objektive Daten“ nicht gibt. Bei Bedarf finden Sie hierfür Anschauungsmaterial in den von uns vorbereiteten Beispielfolien.

2.3 Erarbeitung

In der Vorbereitung haben die Studierenden die Lerneinheit (Schumacher 2024b) bereits mit Hilfe der Tutorial-Videos (forTEXT 2019a; forTEXT 2019b; forTEXT 2019c; forTEXT 2019e) abgeschlossen und die grundlegenden Funktionen von Gephi (Schumacher 2024c) kennengelernt. Geben Sie ihnen nun die Gelegenheit, Fragen

dazu zu stellen und von technischen Problemen zu berichten. Beheben Sie nach Möglichkeit die technischen Probleme. Achten Sie dabei aber darauf, dass Sie nicht zu viel Zeit damit verbringen. Im Notfall müssen die Studierenden zu zweit einen Rechner nutzen, auf dem Gephi ohne Schwierigkeiten läuft. Teilen Sie das Seminar dann in sechs Gruppen ein. Die Gruppen arbeiten alle mit dem gleichen Untersuchungsgegenstand, den drei Figurennetzwerken aus Karoline von Günderrodes Dramen. Die ersten fünf Aufgaben erarbeiten alle Gruppen während der Seminarzeit. Die Aufgabe sechs hat sechs Varianten, von denen jede Gruppe eine wählen kann. Achten Sie darauf, dass jede der sechs Aufgabenvarianten nur von jeweils einer Gruppe bearbeitet wird, damit keine Doppelung entsteht.

Geben Sie den Gruppen folgende Arbeitsaufträge an die Hand:

Aufgabe 1: Schauen Sie sich die Figurennetzwerke aus Karoline von Günderrodes Dramen zunächst auf *Dracor* an. Welches ist das Drama mit der höchsten Dichte? Was bedeutet diese und wofür stehen die anderen Netzwerk-Metriken?

Aufgabe 2: Laden Sie sich nun die für Gephi aufbereiteten Netzwerkdaten für jedes Drama im GEFX-Format von Gephi herunter. Öffnen Sie die Netzwerkdaten mithilfe von Gephi. Schauen Sie sich das Datenlabor an. Welche Zusatzinformationen (außer Figuren und deren Verbindungen) finden Sie?

Aufgabe 3: Wählen Sie eines der Dramen aus, das Sie in Ihrer Projektarbeit primär untersuchen wollen. Öffnen Sie die dazugehörigen Netzwerkdaten in Gephi. Wählen Sie ein Layout für Ihr Netzwerk. Passen Sie die Farbe der Knoten nach Figurengender an. Gibt es mehr männliche, weibliche oder undefinierte Figuren? Passen Sie die Farbe der Verbindungen nach Gewichtung an. Welche Verbindung ist die stärkste?

Aufgabe 4: Erstellen Sie eine Vorschau Ihres Netzwerkes und sichern Sie diese, indem Sie die Screenshot-Funktion nutzen. Wenden Sie sich nun den anderen beiden Netzwerken zu. Laden Sie die Netzwerkdaten dazu von *DraCor* herunter. Erstellen Sie vergleichbare Netzwerke und schauen Sie sich alle drei Dramen-Netzwerke nebeneinander an. In welchem Drama kommen die meisten weiblichen Figuren vor? Welche der Figuren ist im jeweiligen Drama diejenige, die am meisten spricht?

Aufgabe 5: Begeben Sie sich nun auf die Suche nach Hintergrundinformationen zu den drei Dramen und deren Autorin. Wie sind die Dramen zeitlich einzuordnen? Wann und wo hat die Schriftstellerin Karoline von Günderrode gelebt und mit wem stand sie in Kontakt?

Aufgabe 6: 1: Lesen Sie das Drama *Nikator*. Erstellen Sie dazu mit Gephi ein Netzwerk, das die Art der Verbindung der Figuren zueinander zeigt (Macht, Verwandtschaft, Liebe,...). Geben Sie den unterschiedlichen Beziehungstypen unterschiedliche Werte auf einer Skala von eins bis fünf und tragen Sie diese Werte in die Datentabelle unter *weight* ein. Geben Sie ihrem Netzwerk ein Layout und gestalten Sie es farblich. Vergleichen Sie dieses Netzwerk mit dem während der Seminarsitzung betrachteten Netzwerk zu *Nikator*. Wo liegen Gemeinsamkeiten und Unterschiede? Welche Erfahrungen haben Sie beim Anlegen des Netzwerks gemacht?

2: Lesen Sie das Drama *Nikator*. Erstellen Sie dazu mit Gephi ein Netzwerk, das die emotionalen Verbindungen zwischen den Figuren zeigt. Entwickeln Sie eine Skala mit Werten von eins bis fünf, die die Stärke der Verbindung beziffern. Tragen Sie diese Werte in die Datentabelle unter *weight* ein. Geben Sie ihrem Netzwerk ein Layout und gestalten Sie es farblich. Vergleichen Sie dieses Netzwerk mit dem während der Seminarsitzung betrachteten Netzwerk zu *Nikator*. Wo liegen Gemeinsamkeiten und Unterschiede? Welche Erfahrungen haben Sie beim Anlegen des Netzwerks gemacht?

3: Lesen Sie das Drama *Magie und Schicksal*. Erstellen Sie dazu mit Gephi ein Netzwerk, das die Art und die Richtung der Liebe zeigt, die die Figuren miteinander verbindet. Entwickeln Sie eine Skala mit Werten von eins bis drei, die die Stärke der Liebe beziffern. Tragen Sie diese Werte in die Datentabelle unter *weight* ein. Achten Sie darauf, dass Sie auch die Richtung der Liebe berücksichtigen müssen. Bei einer beidseitigen Liebesbeziehung müssen Sie dafür zwei Kanten anlegen. Geben Sie ihrem Netzwerk ein Layout und gestalten Sie es farblich. Vergleichen Sie dieses Netzwerk mit dem während der Seminarsitzung betrachteten Netzwerk zu *Magie und Schicksal*. Wo liegen Gemeinsamkeiten und Unterschiede? Welche Erfahrungen haben Sie beim Anlegen des Netzwerks gemacht?

4: Lesen Sie das Drama *Magie und Schicksal*. Erstellen Sie dazu mit Gephi zwei Netzwerke. Im ersten Netzwerk visualisieren Sie die Beziehungen der Figuren zueinander entsprechend des Wissensstandes, die die Zuschauer*innen und auch die Figuren selbst zu Beginn des Dramas haben. Im zweiten Netzwerk zeigen Sie die tatsächlichen familiären Verhältnisse, die am Ende aufgedeckt werden. Geben Sie ihren Netzwerken Layouts und gestalten Sie sie farblich. Vergleichen Sie diese Netzwerke mit dem während der Seminarsitzung betrachteten Netzwerk zu *Magie und Schicksal*. Wo liegen Gemeinsamkeiten und Unterschiede? Welche Erfahrungen haben Sie beim Anlegen des Netzwerks gemacht?

5: Lesen Sie das Drama *Udohla*. Erstellen Sie dazu mit Gephi ein Netzwerk, das nicht nur die auftretenden, sondern alle im Drama erwähnten Figuren und ihre Beziehungen zueinander sowie ihre Religionszugehörigkeit

zeigt (fügen Sie dazu im Datenlabor eine neue Spalte „Religion“ ein). Stellen Sie die einander gegenüberstehenden familiären Gruppen im Netzwerk visuell dar. Vergleichen Sie dieses Netzwerk mit dem während der Seminarsitzung betrachteten Netzwerk zu *Udohla*. Wo liegen Gemeinsamkeiten und Unterschiede? Welche Erfahrungen haben Sie beim Anlegen des Netzwerks gemacht?

6: Lesen Sie das Drama *Udohla*. Erstellen Sie dazu mit Gephi zwei Netzwerke, die die Machtstrukturen zeigen, die die Figuren miteinander verbinden. Im ersten Netzwerk zeigen Sie das weltliche Herrschaftssystem (mit dem König als höchstem Machthaber). Im zweiten Netzwerk zeigen Sie die Machtstrukturen, die durch Liebe (familiäre oder romantische) entstehen. Geben Sie Ihren Netzwerken passende Layouts und gestalten Sie sie farblich. Vergleichen Sie diese Netzwerke mit dem während der Seminarsitzung betrachteten Netzwerk zu *Udohla*. Wo liegen Gemeinsamkeiten und Unterschiede? Welche Erfahrungen haben Sie beim Anlegen des Netzwerks gemacht?

2.4 Sicherung

Die zweite Sitzung beginnt mit den Ergebnispräsentationen. Diese finden im Plenum statt. Lassen Sie alle Gruppen direkt nacheinander präsentieren und bitten Sie die Zuhörenden, aufkommende Fragen zu notieren. Nach allen Gruppenpräsentationen sollten Sie noch etwas Zeit für Rückfragen, Anmerkungen und weiterführende Bemerkungen einplanen. Nach insgesamt etwa 60 Minuten leiten Sie langsam über in die letzte Phase dieses Lehrmoduls.

2.5 Transfer & Reflexion

In der Sicherungsphase steht zuerst die Reflexion der Erfahrungen aus der Gruppenarbeit im Mittelpunkt. Wie sind die Gruppen bei der Erarbeitung ihrer Präsentationen vorgegangen? Wie sind sie auf ihr Schwerpunktthema gekommen? Welche Besonderheiten sind ihnen bei der Erarbeitung eines eigenen Netzwerkes im Vergleich zur Arbeit mit den Netzwerkdaten von DraCor aufgefallen? Welche historischen und literaturtheoretischen Themen wurden aufgegriffen und wie wurden diese dargestellt? Wie hätte man diese Themen ohne die digitale Methode erarbeiten können? Welche anderen Themen könnten sich die Studierenden für weitere Netzwerkanalysen vorstellen?

Am Ende der Sitzung fragen Sie Ihre Studierenden, inwiefern sie Netzwerkanalyse in ihre zu schreibende Hausarbeit integrieren könnten. Fragen Sie sie außerdem, ob ihnen die Methode auch für andere Seminare bzw. ihr weiteres Studium geeignet erscheint. Betonen Sie dabei, dass die digitale Methode immer auch mit traditionelleren Methoden verbunden werden kann. Eine sozialwissenschaftliche Methode wie die Netzwerkanalyse eignet sich nicht nur für die textimmanente Analyse, sondern kann zudem und insbesondere Aufschluss über Briefnetzwerke, kreative Zirkel oder Autor*innen einer oder mehrerer Epochen geben. Wichtig ist, dass Sie zum Abschluss dieses Lehrmoduls noch einmal darauf hinweisen, dass die Netzwerkanalyse ein interpretativer Prozess ist, in dessen Rahmen die Komplexität des Textes auf bestimmte Parameter reduziert wird. Selbst bei der digitalen Netzwerkanalyse geschieht die Aufbereitung der Daten nicht auf Knopfdruck, sondern muss immer manuell durchgeführt werden. Das heißt, dass auch Projekte wie DraCor diese Daten nicht automatisch generieren (vgl. **Text Mining**), sondern in demselben Prozess erstellen, den Ihre Studierenden durchlaufen haben.

3. Lösungen zu den Beispielaufgaben

Die Lösungsvorschläge zu diesem Lehrmodul erhalten Sie auf [Zenodo](#).

Externe und weiterführende Links

- DraCor: <https://web.archive.org/save/https://dracor.org/> (Letzter Zugriff: 31.07.2024)
- Materialien zum Lehrmodul: <https://zenodo.org/records/12529966> (Letzter Zugriff: 31.07.2024)
- Netzwerkdaten zu *Nikator*: <https://web.archive.org/save/https://dracor.org/ger/guenderode-nikator> (Letzter Zugriff: 31.07.2024)
- Netzwerkdaten zu *Magie und Schicksal*: <https://web.archive.org/save/https://dracor.org/ger/guenderode-magie-und-schicksal> (Letzter Zugriff: 31.07.2024)
- Netzwerkdaten zu *Udohla*: <https://web.archive.org/save/https://dracor.org/ger/guenderode-udohla> (Letzter Zugriff: 31.07.2024)

Bibliographie

- forTEXT. 2019a. Tutorial: Gephi installieren. Netzwerkanalyse und Literaturanalyse. 8. April. doi: 10.5281/zenodo.10276970, <https://doi.org/10.5281/zenodo.10276970>.
- . 2019b. Tutorial: Gephi Daten zur Netzwerkanalyse vorbereiten. Netzwerkanalyse und Literaturanalyse. 22. April. doi: 10.5281/zenodo.10276657, <https://doi.org/10.5281/zenodo.10276657>.
- . 2019c. Tutorial: Gephi Netzwerkvisualisierung bearbeiten. Netzwerkanalyse und Literaturanalyse. 6. Mai. doi: 10.5281/zenodo.10372210, <https://doi.org/10.5281/zenodo.10372210>.
- . 2019d. Analyse der Figurennetzwerke in Lessings „Emilia Galotti“. 23. Mai. doi: 10.5281/zenodo.10221595, <https://doi.org/10.5281/zenodo.10221595>.
- . 2019e. Tutorial: Netzwerkanalyse mit Ezlinavis. Netzwerkanalyse und Literaturanalyse. 20. Juni. doi: 10.5281/zenodo.10372220, <https://youtu.be/10.5281/zenodo.10372220>.
- Horstmann, Jan. 2024. Ressourcenbeitrag: DraCor - Drama Corpora Project. Hg. von Evelyn Gius. *forTEXT* 1, Nr. 6. Netzwerkanalyse (30. August). doi: 10.48694/fortext.3785, <https://fortext.net/ressourcen/textsammlungen/dracor-drama-corpora-project>.
- Schumacher, Mareike. 2024b. Lerneinheit: Netzwerkanalyse mit Gephi. Hg. von Evelyn Gius. *forTEXT* 1, Nr. 6. Netzwerkanalyse (30. August). doi: 10.48694/fortext.3779, <https://fortext.net/routinen/lerneinheiten/netzwerkanalyse-mit-gephi>.
- . 2024a. Methodenbeitrag: Netzwerkanalyse. Hg. von Evelyn Gius. *forTEXT* 1, Nr. 6. Netzwerkanalyse (30. August). doi: 10.48694/fortext.3759, <https://fortext.net/routinen/methoden/netzwerkanalyse>.
- . 2024c. Toolbeitrag: Gephi. Hg. von Evelyn Gius. *forTEXT* 1, Nr. 6. Netzwerkanalyse (30. August). doi: 10.48694/fortext.3783, <https://fortext.net/tools/tools/gephi>.

Glossar

- Annotation** Annotation beschreibt die manuelle oder automatische Hinzufügung von Zusatzinformationen zu einem Text. Die manuelle Annotation wird händisch durchgeführt, während die (teil-)automatisierte Annotation durch **Machine-Learning-Verfahren** durchgeführt wird. Ein klassisches Beispiel ist das automatisierte **PoS-Tagging** (Part-of-Speech-Tagging), welches oftmals als Grundlage (**Preprocessing**) für weitere Analysen wie Named Entity Recognition (NER) nötig ist. Annotationen können zudem deskriptiv oder analytisch sein.
- Browser** Mit Browser ist in der Regel ein Webbrowser gemeint, also ein Computerprogramm, mit dem das Anschauen, Navigieren auf, und Interagieren mit Webseiten möglich wird. Am häufigsten genutzt werden dafür Chrome, Firefox, Safari oder der Internet Explorer.
- Data Mining** Data Mining gehört zum Fachbereich **Information Retrieval** und bezieht sich auf die systematische Anwendung computergestützter Methoden, die darauf abzielt, in vorhandenen Datenbeständen Muster, Trends oder Zusammenhänge zu erkennen. Textbasierte Formen des Data Minings sind u. a. **Text Mining**, **Web Mining** und **Opinion Mining**.
- Domäneadaptation** Domäneadaptation beschreibt die Anpassung einer in einem Fachgebiet entwickelten digitalen Methode an ein anderes Fachgebiet.
- Feature** Unter Features können Einzelfunktionen eines Tools verstanden werden, die beispielsweise komplexe Funktionen wie die Visualisierung eines Textes als **Wordcloud** ermöglichen, oder auch kleinere Funktionseinheiten wie den Abgleich einzelner Spracheigenschaften (**Properties**) mit **annotierten** Beispieltexten darstellen.
- HTML** HTML steht für *Hypertext Markup Language* und ist eine textbasierte Auszeichnungssprache zur Strukturierung elektronischer Dokumente. HTML-Dokumente werden von **Webbrowsern** dargestellt und geben die Struktur und Online-Darstellung eines Textes vor. HTML-Dateien können außerdem zusätzliche **Metainformationen** enthalten, die auf einer Webseite selbst nicht ersichtlich sind.
- Information Retrieval** Die Teildisziplin der Informatik, das Information Retrieval, beschäftigt sich mit der computergestützten Suche und Erschließung komplexer Informationen in meist unstrukturierten Datensammlungen.
- Korpus** Ein Textkorpus ist eine Sammlung von Texten. Korpora (Plural für „das Korpus“) sind typischerweise nach Textsorte, Epoche, Sprache oder Autor*in zusammengestellt.
- Lemmatisieren** Die Lemmatisierung von Textdaten gehört zu den wichtigen **Preprocessing**-Schritten in der Textverarbeitung. Dabei werden alle Wörter (**Token**) eines Textes auf ihre Grundform zurückgeführt. So werden beispielsweise Flexionsformen wie „schneller“ und „schnelle“ dem Lemma „schnell“ zugeordnet.
- Machine Learning** Machine Learning, bzw. maschinelles Lernen im Deutschen, ist ein Teilbereich der künstlichen Intelligenz. Auf Grundlage möglichst vieler (Text-)Daten erkennt und erlernt ein Computer die häufig sehr komplexen Muster und Gesetzmäßigkeiten bestimmter Phänomene. Daraufhin können die aus den Daten gewonnen Erkenntnisse verallgemeinert werden und für neue Problemlösungen oder für die Analyse von bisher unbekannten Daten verwendet werden.

Markup Language Markup Language bezeichnet eine maschinenlesbare Auszeichnungssprache, wie z. B. **HTML**, zur Formatierung und Gliederung von Texten und anderen Daten. So werden beispielsweise auch **Annotationen** durch ihre Digitalisierung oder ihre digitale Erstellung zu Markup, indem sie den Inhalt eines Dokumentes strukturieren.

Metadaten Metadaten oder Metainformationen sind strukturierte Daten, die andere Daten beschreiben. Dabei kann zwischen administrativen (z. B. Zugriffsrechte, Lizenzierung), deskriptiven (z. B. Textsorte), strukturellen (z. B. Absätze oder Kapitel eines Textes) und technischen (z. B. digitale Auflösung, Material) Metadaten unterschieden werden. Auch **Annotationen** bzw. **Markup** sind Metadaten, da sie Daten/Informationen sind, die den eigentlichen Textdaten hinzugefügt werden und Informationen über die Merkmale der beschriebenen Daten liefern.

Named Entities Eine Named Entity (NE) ist eine Entität, oft ein Eigenname, die meist in Form einer Nominalphrase zu identifizieren ist. Named Entities können beispielsweise Personen wie „Nils Holgerson“, Organisationen wie „WHO“ oder Orte wie „New York“ sein. Named Entities können durch das Verfahren der Named Entity Recognition (NER) automatisiert ermittelt werden.

Opinion Mining Unter Opinion Mining, oder Sentiment Analysis, versteht man die Analyse von Stimmungen oder Haltungen gegenüber einem Thema, durch die Analyse natürlicher Sprache. Das Opinion Mining gehört zu den Verfahren des **Text Minings**.

POS PoS steht für *Part of Speech*, oder „Wortart“ auf Deutsch. Das PoS- **Tagging** beschreibt die (automatische) Erfassung und Kennzeichnung von Wortarten in einem Text und ist ein wichtiger **Preprocessing**-Schritt, beispielsweise für die Analyse von **Named Entities**.

Preprocessing Für viele digitale Methoden müssen die zu analysierenden Texte vorab „bereinigt“ oder „vorbereitet“ werden. Für statistische Zwecke werden Texte bspw. häufig in gleich große Segmente unterteilt (*chunking*), Großbuchstaben werden in Kleinbuchstaben verwandelt oder Wörter werden **lemmatisiert**.

Property Property steht für „Eigenschaft“, „Komponente“ oder „Attribut“. In der automatischen **Annotation** dienen konkrete Wortheigenschaften wie Groß- und Kleinschreibung zur Klassifizierung von Wörtern oder Phrasen. Durch die Berücksichtigung solcher Eigenschaften in den **Features** eines Tools kann **maschinelles Lernen** bestimmter Phänomene umgesetzt werden. In der manuellen Annotation können als Properties auch Eigenschaften von **Annotationen** benannt werden.

Text Mining Das Text Mining ist eine textbasierte Form des **Data Minings**. Prozesse & Methoden, computer-gestützt und automatisch Informationen bzw. Wissen aus unstrukturierten Textdaten zu extrahieren, werden als Text Mining zusammengefasst.

Type/Token Das Begriffspaar „Type/Token“ wird grundsätzlich zur Unterscheidung von einzelnen Vorkommnissen (Token) und Typen (Types) von Wörtern oder Äußerungen in Texten genutzt. Ein Token ist also ein konkretes Exemplar eines bestimmten Typs, während ein Typ eine im Prinzip unbegrenzte Menge von Exemplaren (Token) umfasst.

Es gibt allerdings etwas divergierende Definitionen zur Type-Token-Unterscheidung. Eine präzise Definition ist daher immer erstrebenswert. Der Satz „Ein Bär ist ein Bär.“ beinhaltet beispielsweise fünf Worttoken („Ein“, „Bär“, „ist“, „ein“, „Bär“) und drei Types, nämlich: „ein“, „Bär“, „ist“. Allerdings könnten auch vier Types, „Ein“, „ein“, „Bär“ und „ist“, als solche identifiziert werden, wenn Großbuchstaben beachtet werden.

Web Mining Unter Web Mining versteht man die Anwendung von Techniken des **Data Mining** zur Extraktion von Informationen aus dem World Wide Web. Das Web Mining ist ein Teilbereich des Data Minings und zählt zu einem der wichtigsten Anwendungsgebiete für das **Text Mining**.

Wordcloud Eine *Wordcloud*, oder auch Schlagwortwolke, ist eine Form der Informationsvisualisierung, beispielsweise von Worthäufigkeiten in einem Text oder einer Textsammlung. Dabei werden unterschiedlich gewichtete Wörter, wie die häufigsten Wörter, i.d.R. größer oder auf andere Weise hervorgehoben dargestellt. Die horizontale/vertikale Ausrichtung und die Farbe der dargestellten Wörter hat meistens allerdings keinen semantischen Mehrwert.