

WHITEPAPER

Text als Daten im öffentlichen und sozialen Sektor

Von unstrukturierten Daten zu Erkenntnissen und
Auswirkungen.

elinor
text as data

Entwickelt von:

&effect

Gefördert durch:



Bundesministerium
für Wirtschaft
und Klimaschutz

aufgrund eines Beschlusses
des Deutschen Bundestages

Einführung

90% aller Daten, die in Organisationen vorkommen, sind unstrukturiert! Der Großteil dieser Daten sind Texte. Diese Textdaten kommen vor allem in Excel-Tabellen oder Datenbanken vor und werden als Datenquelle für die quantitative Analyse oft übersehen.

Unstrukturierte Textdaten umgeben uns tagtäglich:

- Wir lesen Medienartikel.
- Wir überprüfen, was Twitter über bestimmte Themen postet.
- Wir lesen Berichte.
- Wir bearbeiten Formulare.
- Wir legen Dokumente ab.



Wir sind überzeugt, dass die Analyse von Textdaten im öffentlichen und sozialen Sektor Organisationen helfen kann, gesellschaftliche und ökologische Probleme besser zu lösen. Aus diesem Grund haben wir sechs Datenquellen, vier Anwendungsfälle und vier Wirkungsbereiche zusammengestellt, bei denen Textanalysen zum Einsatz kommen. Die konkreten Anwendungsfälle, können Ihnen helfen, den Wert von Textdaten in ihrem Kontext besser zu verstehen.

Mit **elinor – text as data** bieten wir eine Lösung, welche die unstrukturierten Textdaten Ihrer Organisation in wertvolle Erkenntnisse verwandelt, damit Sie etwas bewirken können. Mehr Informationen zu elinor finden Sie unter elinor.and-effect.com.

Nützliche Quellen für unstrukturierte Daten

Um das Potenzial der Textanalyse besser zu verstehen, werden in diesem Kapitel die möglichen Datenquellen identifiziert. Im Folgenden finden Sie einen Überblick über die Datenquellen, welche wir in den letzten Jahren im Rahmen unserer Forschung und Projekte genutzt haben.

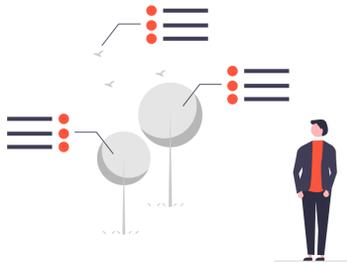
1) Medienartikel sind eine Fundgrube für Informationen, welche weit in der Vergangenheit zurückgehen und die uns Einblicke geben:

- wie Menschen ein Ereignis wahrgenommen haben, als es geschehen ist.
- in verschiedene Standpunkte zu einem Thema.
- in historische Entwicklungen und um diese im Laufe der Zeit nachvollziehen zu können.
- Themen im Kontext der Zeit zu untersuchen, in der sie stattfanden.



Die Verwendung von Zeitungsartikeln als Datenquellen ist in der sozialwissenschaftlichen Forschung weit verbreitet. Jetzt erkennen immer mehr Organisationen das Potenzial: 1) die Analyse von Entwicklungen in ihrem Bereich, 2) die Anreicherung von Daten über Themen oder 3) die Analyse der Wahrnehmung ihrer Organisation in den Medien.

2) Social Media Posts bieten fast in Echtzeit einen Einblick in Geschehnisse, welche gerade passieren. Über jedes große Ereignis wird mit Tausenden von Tweets, Posts und Bildern berichtet. Ähnlich wie bei Zeitungsartikeln können soziale Medien eine wichtige Rolle spielen, um den Kontext von Organisationen besser zu verstehen.



Social Media gibt uns aktuelle Einblicke:

- wie Menschen auf Ereignisse oder Themen reagieren.
- wie Menschen auf Ihre Organisation reagieren oder diese sehen.
- wie sich die Reaktionen im Laufe der Zeit entwickeln.

Der Umgang mit Posts in den sozialen Medien muss aus ethischen und datenschutzrechtlichen kritisch evaluiert werden. Auch dann ist bei der Interpretation von Social Media Daten Vorsicht geboten. Dennoch gibt es ein enormes Potenzial, wenn man sich des Kontextes bewusst ist.

3) Online-Stellenausschreibungen sind auf den bekannten Jobportale wie Indeed, StepStone oder LinkedIn jederzeit abrufbar. Mit Millionen von neuen Jobangeboten, die monatlich von Arbeitgebern veröffentlicht werden, sind Stellenanzeigen eine interessante Datenquelle, um Trends auf dem Arbeitsmarkt zu untersuchen.

Eine Online-Stellenanzeige gibt Aufschluss darüber:

- welche Jobs gefragt sind.
- welche Kompetenzen und Fertigkeiten auf dem aktuellen Arbeitsmarkt gesucht werden.
- welche Qualifikationen erforderlich sind, um auf dem Arbeitsmarkt erfolgreich zu sein.



Da sich die Stellenanzeigen zunehmend in die digitale Welt verlagern, sind die Daten bereits online verfügbar.

4) Berichte und Reports enthalten vor allem im öffentlichen Sektor sehr viele Informationen. Dazu gehören Profile, Projektzusammenfassungen, regelmäßige Briefings, Jahresabschlussberichte usw. Durch die Kombination oder automatische Verarbeitung dieser Berichte können Muster aufgedeckt und einzigartige Einblicke gewonnen werden.

Die Berichte konzentrieren sich häufig auf Folgendes:

- frühere Performanz.
- Pläne und Maßnahmen.
- Zusammenfassung von Problemen und Entwicklungen.



Durch systematisches Sammeln von Berichtsdaten können Sie mit uns einen einzigartigen Datensatz erstellen, der in Verbindung mit Ihren Zielfragen und -themen einen enormen Wert für strukturierte Analysen hat.

5) Qualitative Umfragedaten lassen sich zum Teil mit Hilfe einfacher Statistiken zusammenfassen. In einigen Umfragen werden jedoch detailliertere Informationen in offenen Fragen ergründet. Die Strukturierung dieser Erhebungen kann die Daten einer Studie weiter anreichern und Informationen über:



- Meinungen
- Erfahrungen
- Erzählungen und
- Berichte geben.

6) Beschwerde-/Anforderungsformulare werden oft von Organisationen angefordert, um Feedback zu sammeln oder Beteiligungen abzufragen. Manchmal kommen Beschwerden oder Anfragen in Form von Briefen, welche eine weitere Quelle wertvoller Textdaten sind und im Optimalfall digitalisiert werden.

Die Analyse dieser Formulare kann die Effizienz der Bearbeitung enorm verbessern durch:

- Kategorisieren von Anfragen
- Bereiche oder Themen mit Verbesserungspotenzial



Durch das bessere Verständnis von Beschwerden und Anfragen können Sie die Bearbeitungszeiten verkürzen und die Bedürfnisse sowie Wünsche Ihrer Interessengruppen besser verstehen und entsprechend anpassen.

Die Liste der potenziellen Datenquellen ist hier noch nicht zu Ende. In der Sozialwissenschaft analysieren Forscher zum Beispiel Reden oder Parteiprogramme. Organisationen könnten Benutzerfeedback oder Bewertungen untersuchen. Möglicherweise verfügt Ihre Organisation bereits über Tools und Anwendungen, die große Mengen an Textdaten erzeugen.

Welche Informationen für Sie wertvoll sind und für die Analyse verwendet werden können, hängt vom Kontext und den Zielen Ihrer Organisation ab. Im nächsten Kapitel wird untersucht, was Textdaten für Ihre Organisation leisten können.

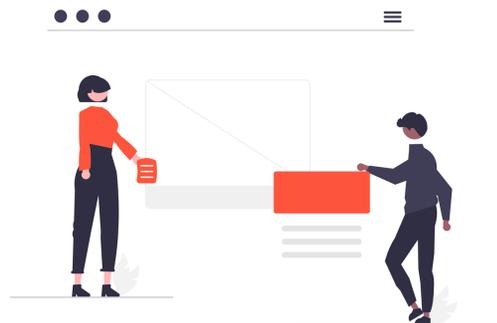
Was können wir mit strukturierten Textdaten tun?

Nachdem wir festgestellt haben, welche Datenquellen sich für die Textanalyse eignen, können wir Methoden und Ergebnisse erarbeiten, die für Ihre Organisation von Nutzen sind.

1) Verstehen der öffentlichen Meinung: Mit Daten aus traditionellen und neuen Medien kann Ihnen die Textmodellierung dabei helfen, zu verstehen, welche Themen im Zusammenhang mit Ihrer Organisation wichtig sind.

Es gibt drei wesentliche Möglichkeiten, wie wir diese Daten nutzen können:

- Durch die Klassifizierung von Mediendokumenten nach Kategorien lässt sich z. B. feststellen, über welche Themen in welchen Zusammenhängen gesprochen wird.
- Durch die Extraktion von benannten Entitäten aus Texten können wir feststellen, welcher Inhalt von wem in welchem Kontext gesprochen wurde.
- Durch die Bewertung des allgemeinen Tons (positiv, neutral, negativ, etc.) eines Medienartikels oder -beitrags kann die Stimmung gegenüber einer Entität, einem Thema oder einem Bereich eingeschätzt werden.



Die Anwendung dieser Techniken können vor allem in der Interessenvertretung hilfreich sein. In einem von uns durchgeführten Projekt haben wir im Vereinten Königreich untersucht, wie in den verschiedenen Medien die Stimmung gegenüber Pflegekräften ist. Das Ergebnis zeigte, dass im Vergleich zu medizinischen und sozialen Berufen die Pflegekräfte unterrepräsentiert sind. Auch in den Medien werden Pflegekräfte häufig im Zusammenhang mit negativen Ereignissen (z.B. Verbrechen oder Skandalen) erwähnt und nicht mit der systemrelevanten Arbeit, die sie leisten oder der mangelnden Unterstützung für bessere Arbeitsbedingungen.

2) Trends verstehen: Die vorher genannten Datenquellen können zur Beobachtung von Trends verwendet werden. Nach der Klassifizierung oder Extraktion von bestimmten Konzepten aus den Daten können Sie mit unserer Unterstützung Veränderungen im Laufe der Zeit beobachten:

- Ändert sich die Anzahl der Erwähnungen im Laufe der Zeit?
- Ändert sich der Kontext der Erwähnungen?
- Gibt es neue Themen, die auftauchen?
- Gibt es bestimmte Ereignisse, die den Trend umkehren?



Die ermittelten Daten können für eine evidenzbasierte Entscheidungsfindung genutzt werden. In einem weiteren Projekt analysieren wir Millionen von Online-Stellenanzeigen. Durch effiziente Textanalysen ist es möglich, die Veränderungen in den verschiedenen Berufen, in unterschiedlichen Sektoren bezüglich der Fähigkeiten und Qualifikationen festzustellen. Die gesammelten Informationen können direkt in eine datenbasierte Politikgestaltung am Arbeitsmarkt einfließen.

3) Schnellere Bearbeitung von Dokumenten: In Organisationen fallen oft große Mengen an Dokumenten an, welche manuell sortiert und bearbeitet werden müssen. Die Dokumente bestehen unter anderem aus Anträgen, Beschwerden oder Fragen. Durch die Strukturierung dieser Daten kann die Bearbeitung beschleunigt werden, so dass die Mitarbeiter mehr Zeit haben, sich um den Inhalt der Dokumente zu kümmern:



- Worum geht es in einem Dokument, und wer soll es bearbeiten?
- Gab es in der Vergangenheit bereits ähnliche Anfragen?
- Welche Daten können wir extrahieren, die uns bei der Bearbeitung der Dokumente helfen?

Die Automatisierung der Bearbeitung von den Dokumenten unterstützt Ihre Mitarbeiter:innen bei der manuellen Arbeit. Durch bereits bestehende Techniken und Algorithmen können Dokumente automatisiert sortiert werden. Im sozialen und öffentlichen Sektor findet noch sehr viel durch manuelle Arbeit statt. Mit einer Automatisierung soll der Bearbeitungsprozess, vor allem im Posteingang erleichtert und unterstützt, jedoch nicht ersetzt werden.

Auch hier können wir bereits ein passendes Beispiel für ein bereits durchgeführtes Projekt anhängen: Hierbei geht es um ein Programm, welches Sozialarbeiter:innen hilft Fachfragen zu beliebigen Themen zustellen. Durch einen Textalgorithmus werden die Begriffe automatisch zur Taxonomie hinzugefügt. In der Datenbank der Sozialarbeiter:innen können dann die passenden Ansprechpartner:innen gesehen und kontaktiert werden.

4) Erweiterung bestehender Datenbanken: Organisationen nutzen oft strukturierte Datenbanken für einzelne Tätigkeitsbereiche; z.B. für eigene Programme, Events oder für die gesamte Institution sein. Gelegentlich ist das manuelle Befüllen und Aktualisieren die Kerntätigkeit der Organisation, was häufig ein sehr aufwändiger und fehleranfälliger Prozess ist.

Die automatische Anreicherung der Datenbank mit Textdaten kann diesen Prozess entweder beschleunigen oder die Daten mit strukturierten Informationen aus dem Text weiter anreichern:



- Daten können mit Medienerwähnungen angereichert werden.
- Ereignisse, an denen eine Organisation beteiligt ist, können Datenbank ergänzen.
- Themen oder Kategorien können Entitäten erweitern.

Für die sozialwissenschaftliche Forschung ist das automatische Anreichern von Datenbanken bedeutsam, somit ist es möglich, z.B. die Umweltauswirkungen von Unternehmen sichtbar zu machen oder einen Index über die transparente Unternehmensdaten zu erstellen.

In welchen Bereichen wirken Textdaten?

Nachdem Sie nun einen Einblick in die verschiedenen Anwendungsgebiete und Datenquellen bekommen haben, folgen nun die vier Wirkungsbereiche, in denen wir bereits Textdaten erfolgreich genutzt haben.

1) Arbeitsmarkt:

Mit der Unterstützung von innovativen Methoden und Datenquellen, ist möglich eine Orientierung zu geben, welche Kompetenzen zukünftig auf dem Arbeitsmarkt gefragt sind und welche Trends sich über einen gewissen Zeitraum entwickeln.

2) Nachhaltige Finanzmärkte:

Der Finanzsektor kann durch die Erhöhung der Transparenz nachhaltiger und verantwortungsvoller gestaltet werden. Die Textdaten ermöglichen die Unterstützung bei der Transformation zu einer grüneren Wirtschaft.

3) Öffentliche Verwaltung:

In der öffentlichen Verwaltung gibt es viele Schnittstellen in der Bevölkerung. Mit der Unterstützung von Textdaten können vor allem die bestehenden Herausforderungen noch besser verstanden und vorhandene Prozesse optimiert werden, um gemeinwohlorientierte Programme effektiver und effizienter zu gestalten.

4) Sozialwissenschaftliche Forschung:

In der Sozialwissenschaft sind Textdaten seit langem ein Teil der quantitativen Forschung. Durch innovative Methoden und Modellen ist es möglich, die größer werdende Mengen an Textdaten noch besser zu analysieren.

Kontaktieren Sie uns für eine kostenlose Beratung zu Textdaten und individuellen Datenlösungen!

Johannes Müller

Head of Data Science

johannes.mueller@and-effect.com

+49 151 24010386

Über elinor – text as data:

elinor strukturiert Ihre Textdokumente mithilfe einer Domain-Ontologie - einer digitalen Karte Ihres Fachwissens. Durch die Annotation von Beispielen in Texten kann diese verwendet werden, um Erkenntnisse aus Textdaten zu gewinnen und leistungsstarke Modelle für maschinelles Lernen zu erstellen.

Über &effect:

Unser Ziel ist es, Data Science zu einem integralen Bestandteil der Entscheidungsfindung im öffentlichen und sozialen Sektor zu machen. Dafür entwickeln wir wirkungsorientierte Datenprodukte an der Schnittstelle zwischen Sozialwissenschaften, Data Science und Softwareentwicklung.

Herausgeber:

&effect data solutions GmbH
Pasteurstraße 34
10407 Berlin

Gefördert durch:



Bundesministerium
für Wirtschaft
und Klimaschutz

aufgrund eines Beschlusses
des Deutschen Bundestages