

Annotationsschema zum Aufbau eines deutschsprachigen Textkorpus es im Bereich der computergestützten Identifikation von Ironie

Das vorliegende Annotationsschema dient als Leitfaden zur Annotation von Tweets mit der Zielstellung einen Textkorpus in deutscher Sprache aufzubauen, der im Bereich der computergestützten Identifikation von Ironie mit besonderem Bezug zur Sentiment-Analyse (dt. Stimmungsanalyse) eingesetzt werden kann.

Anmerkung: Annotation bezeichnet das Vorgehen zur Einordnung von Daten (im vorliegenden Fall Textdaten/Tweets) in vordefinierte Klassen.

Nachfolgend wird das Vorgehen zur Annotation im Detail beschrieben. Dieses **Vorgehen ist für jeden zu annotierenden Datensatz (Tweet) gleich und unbedingt vollständig anzuwenden.**

1. Bestimmung der Polarität

Um den Textkorpus später im Rahmen der computergestützten Identifikation von Ironie mit besonderem Bezug zur Sentiment-Analyse nutzen zu können, ist es zunächst notwendig die allgemeine Polarität des Textes zu bestimmen. Die Polarität eines Textes spiegelt sich in subjektiven Äußerungen wie Stimmungen, Gefühlen, Meinungen und Emotionen von Individuen wieder. Ironische Äußerungen sind an dieser Stelle nicht zu berücksichtigen, es ist ausdrücklich nur der **wortwörtliche Text** zu bewerten.

Klassen:

- positiv
- negativ
- neutral
- mixed

Erläuterungen:

Eine Datensatz (Tweet) ist als positiv zu bewerten, wenn dieser eine subjektiv geprägte glückliche, fröhliche, freudige etc. Aussage enthält.

Beispiele:

- „*#Spanien im #Zug mit #Interrail-Karten: eine angenehme Weise, um das Land bequem kennen zulernen, ohne selbst ein Auto fahren zu müssen 😊*“
- „*Habe heute Nachmittag eine Wurzelbehandlung beim Zahnarzt. Jippi, kann es kaum erwarten.*“
- „*Ich hab einen total super gesunden Smoothie getrunken... aus einem Bierglas.*“
- „*Was soll ich sagen? Komme einfach nicht los von dieser Landschaft... #wandern #reisen #Auvergne*“
- „*In die Twittersuche #Verspätung eingeben und #DeutscheBahn vorgeschlagen bekommen, genial :D*“
- „*Ja, Erdgas ist die Lösung! Ist teuer, emittiert enorme Mengen der Treibhausgase Methan und CO2, macht abhängig von Russland – klingt wirklich nach einer super Idee!*“

Ein Datensatz (Tweet) ist als negativ zu bewerten, wenn dieser eine subjektiv geprägte ärgerliche, ängstliche, abwertende, wütende, verletzende etc. Aussage enthält.

Beispiele:

- „*Der 15.4.2019 wird als der Tag eingehen, wo Deutschland für #Zensur stimmte und #NotreDamm aus lauter Wut das brennen anfang oder sowas.*“
- „*Irgendwie dachte ich, Ethnographien wären spannender #Studium*“
- „*No Flixbus in Sicht. Hat Verspätung 😞*“
- „*Bei aller Liebe für #Technik: Es scheint, als würde heute mehr Zeit & Geld in #machinelearning investiert statt in #Bildung. Ziel sollte sein, unsere Kinder mit Wissen & Fähigkeiten auszustatten, mit denen sie #artificialintelligence in Zukunft überlegen sind! #humancapital*“

- „na suuuper.. da kommen morgen x paar tolle filme in der glotze, bin ich wie immer an so nem abend net da.. tyypisch! 😞 “
- „wer die Geschichte nicht ehrt und aus ihr nichts lernt wird auch in der Zukunft die gleiche Fehler machen ach Moment mal so ist ja die Menschheit...hab ich total vergessen ich doofi.“

Ein Datensatz (Tweet) ist als neutral zu bewerten, wenn dieser weder eine positive noch negative subjektive Meinung enthält. Als neutral zu bewerten sind demnach beispielsweise Fakten oder rein informative Datensätze (Tweets).

Beispiele:

- „Warum Studierende seltener zu #Vorlesungen und #Seminaren gehen: #TUChemnitz-Forscher Dr. René Bochmann ordnet die Ergebnisse seiner umfassenden Untersuchung ein“
- „Heute ist #Brückentag! Ein guter Anlass, um eine #Rampe anzuschaffen, die #Barrieren überwindet“
- „Immer mehr Studenten planen nach dem Bachelor den Einstieg in die Berufswelt - Mit dem üppigsten Anfängergehalt von mehr als 50.000 Euro rechnen Chemiker, Juristen und Mediziner – Ingenieure erwarten immerhin noch knapp 45.000. #Studium“
- „Montag geht's wieder los: #Semesterbeginn, Start der #Vorlesungen, #Prüfungen und wenig Parkplätze auf dem Campus.“
- „Ein Mitarbeiter vom #LKASachsen hat auf einer Pegida-Demo ein Kamera-team angezeigt, weil er nicht gefilmt werden wollte. Sein Gesicht ist innerhalb eines Tages nun so berühmt geworden, der könnte Werbung für #calvinklein machen.“

Ein Datensatz (Tweet) ist als mixed zu bewerten, wenn er sowohl positive als auch negative subjektive Äußerungen enthält.

Beispiele:

- „Das Bett im Hotelzimmer ist toll und bequem, die Dusche aber viel zu winzig und unbequem.“
- „Ich liebe den diesjährigen Sommer, Woche um Woche furchtbares Wetter.“
- „Ich liebe es mit Migräne aufzuwachen ☹.“
- „So, die Zusammenfassung der 3. Woche #Prokrastinationsscheiße hat Osterbedingt etwas Verspätung, aber nichtsdestotrotz bin ich mit der Woche ziemlich zufrieden.“

2. Bestimmung von Ironie

Im zweiten Schritt der Annotation gilt es zu entscheiden ob ein Tweet eine ironische Äußerung enthält oder nicht. Ist eine ironische Äußerung vorhanden, ist die Form der Ironie zu bestimmen.

Klassen:

- Verbale Ironie
- Situative Ironie
- Sarkasmus
- Keine Ironie

Erläuterungen:

Verbale Ironie zeichnet sich durch eine Differenz zwischen der wörtlichen Aussage und der eigentlich intendierten Aussage aus. Die intendierte Aussage wird anhand von Kontextwissen (z.B. Allgemeinwissen zu einer Person, einem Ereignis etc.) abgeleitet.

Beispiele:

- „Habe heute Nachmittag eine Wurzelbehandlung beim Zahnarzt. Jippi, kann es kaum erwarten.“
 - „Ich liebe den diesjährigen Sommer, Woche um Woche furchtbares Wetter.“
 - „Ich liebe es mit Migräne aufzuwachen ☹.“
-

Situative Ironie (oder auch „Ironie des Schicksals“) bezieht sich auf eine Situation oder ein Ereignis, die dem Gegenteil der eigentlichen Erwartungen entspricht.

Beispiele:

- *„Ich hab einen total super gesunden Smoothie getrunken... aus einem Bierglas.“*
- *„Irgendwie dachte ich, Ethnographien wären spannender #Studium“*
- *„Ein Mitarbeiter vom #LKASachsen hat auf einer Pegida-Demo ein Kamera-team angezeigt, weil er nicht gefilmt werden wollte. Sein Gesicht ist innerhalb eines Tages nun so berühmt geworden, der könnte Werbung für #calvinklein machen.“*
- *„In die Twittersuche #Verspätung eingeben und #DeutscheBahn vorgeschlagen bekommen, genial :D“*

Sarkasmus zeichnet sich durch eine Form des Spotts bzw. der Verhöhnung aus. Dieser Spott bzw. die Verhöhnung richtet sich gegen ein bestimmtes Ziel. Dieses Ziel kann eine Person, ein Ereignis, ein Unternehmen, ein Gegenstand oder ähnliches sein.

Beispiel:

- *„Ja, Erdgas ist die Lösung! Ist teuer, emittiert enorme Mengen der Treibhausgase Methan und CO₂, macht abhängig von Russland – klingt wirklich nach einer super Idee!“*
- *„wer die Geschichte nicht ehrt und aus ihr nichts lernt wird auch in der Zukunft die gleiche Fehler machen ach Moment mal so ist ja die Menschheit...hab ich total vergessen ich doofi.“*
- *„na suuuper.. da kommen morgen x paar tolle filme in der glotze, bin ich wie immer an so nem abend net da.. tyypisch! 😞“*

Ist keine ironische Äußerung enthalten, ist die Kategorie „Keine Ironie“ auszuwählen.

Beispiele:

- „*#Spanien im #Zug mit #Interrail-Karten: eine angenehme Weise, um das Land bequem kennen zulernen, ohne selbst ein Auto fahren zu müssen 😊*“
- „*Was soll ich sagen? Komme einfach nicht los von dieser Landschaft... #wandern #reisen #Auvergne*“
- „*No Flixbus in Sicht. Hat Verspätung 😞*“
- „*Bei aller Liebe für #Technik: Es scheint, als würde heute mehr Zeit & Geld in #machinelearning investiert statt in #Bildung. Ziel sollte sein, unsere Kinder mit Wissen & Fähigkeiten auszustatten, mit denen sie #artificialintelligence in Zukunft überlegen sind! #humancapital*“
- „*Warum Studierende seltener zu #Vorlesungen und #Seminaren gehen: #TUChemnitz-Forscher Dr. René Bochmann ordnet die Ergebnisse seiner umfassenden Untersuchung ein*“
- „*Heute ist #Brückentag! Ein guter Anlass, um eine #Rampe anzuschaffen, die #Barrieren überwindet*“
- *etc.*

3. Polaritätswechsel

Das Vorhandensein von Ironie in einem Text kann zu einem Wechsel in der Polarität (vgl. Schritt 1.) führen. Ist die wörtliche Aussage z.B. positiv formuliert, die intendierte Aussage des Autors ist aber negativ, liegt ein Polaritätswechsel vor.

Klassen:

- Wechsel
- Kein Wechsel

Erläuterungen:

Ein Wechsel in der Polarität bedeutet, dass die intendierte Aussage nicht mit der wörtlichen (vgl. Schritt 1) übereinstimmt.

Beispiele:

- „Habe heute Nachmittag eine Wurzelbehandlung beim Zahnarzt. Jippi, kann es kaum erwarten.“
- „Ja, Erdgas ist die Lösung! Ist teuer, emittiert enorme Mengen der Treibhausgase Methan und CO₂, macht abhängig von Russland – klingt wirklich nach einer super Idee!“
- „Ich liebe den diesjährigen Sommer, Woche um Woche furchtbares Wetter.“
- „Ich liebe es mit Migräne aufzuwachen ☹.“

Kein Wechsel bedeutet, dass die intendierte Aussage mit der wörtlichen (vgl. Schritt 1) übereinstimmt.

Beispiele:

- „#Spanien im #Zug mit #Interrail-Karten: eine angenehme Weise, um das Land bequem kennen zulernen, ohne selbst ein Auto fahren zu müssen 😊“
- „Ich hab einen total super gesunden Smoothie getrunken... aus einem Bierglas.“
- „Was soll ich sagen? Komme einfach nicht los von dieser Landschaft... #wandern #reisen #Auvergne“
- „In die Twittersuche #Verspätung eingeben und #DeutscheBahn vorgeschlagen bekommen, genial :D“
- „Der 15.4.2019 wird als der Tag eingehen, wo Deutschland für #Zensur stimmte und #NotreDamm aus lauter Wut das brennen anfang oder sowas.“
- „Irgendwie dachte ich, Ethnographien wären spannender #Studium“
- „No Flixbus in Sicht. Hat Verspätung 😞“
- etc.

Beispiele Annotation

| Tweet | Polarität | Ironie | Polaritätswechsel |
|--|-----------|------------------|-------------------|
| „#Spanien im #Zug mit #Interrail-Karten: eine angenehme Weise, um das Land bequem kennen zulernen, ohne selbst ein Auto fahren zu müssen 😊“ | positiv | Keine Ironie | Kein Wechsel |
| „Habe heute Nachmittag eine Wurzelbehandlung beim Zahnarzt. Jippi, kann es kaum erwarten.“ | positiv | Verbale Ironie | Wechsel |
| „Ich hab einen total super gesunden Smoothie getrunken... aus einem Bierglas.“ | positiv | Situative Ironie | Kein Wechsel |
| „Was soll ich sagen? Komme einfach nicht los von dieser Landschaft... #wandern #reisen #Auvergne“ | positiv | Keine Ironie | Kein Wechsel |
| „In die Twittersuche #Verspätung eingeben und #DeutscheBahn vorgeschlagen bekommen, genial :D“ | positiv | Situative Ironie | Kein Wechsel |
| „Ja, Erdgas ist die Lösung! Ist teuer, emittiert enorme Mengen der Treibhausgase Methan und CO2, macht abhängig von Russland – klingt wirklich nach einer super Idee!“ | positiv | Sarkasmus | Wechsel |
| „Der 15.4.2019 wird als der Tag eingehen, wo Deutschland für #Zensur stimmte und #NotreDamm aus lauter Wut das brennen anfangen oder sowas.“ | negativ | Keine Ironie | Kein Wechsel |
| „Irgendwie dachte ich, Ethnographien wären spannender #Studium“ | negativ | Situative Ironie | Kein Wechsel |
| „No Flixbus in Sicht. Hat Verspätung 😞“ | negativ | Keine Ironie | Kein Wechsel |
| „Bei aller Liebe für #Technik: Es scheint, als würde heute mehr Zeit & Geld in #machinelearning investiert statt in #Bildung. Ziel sollte sein, unsere Kinder mit Wissen & Fähigkeiten auszustatten, mit denen sie #artificialintelligence in Zukunft überlegen sind! #humancapital“ | negativ | Keine Ironie | Kein Wechsel |
| „na suuuper.. da kommen morgen x paar tolle filme in der glotze, bin ich wie immer an so nem abend net da.. tyypisch! 😞“ | negativ | Sarkasmus | Kein Wechsel |
| „wer die Geschichte nicht ehrt und aus ihr nichts lernt wird auch in der Zukunft die gleichen Fehler machen ach Moment mal so ist ja die Menschheit...hab ich total vergessen ich doof.“ | negativ | Sarkasmus | Kein Wechsel |
| „Warum Studierende seltener zu #Vorlesungen und #Seminaren gehen: #TUCheumnitz-Forscher Dr. René Bochmann ordnet die Ergebnisse seiner umfassenden Untersuchung ein“ | neutral | Keine Ironie | Kein Wechsel |

| | | | |
|--|---------|------------------|--------------|
| „Heute ist #Brückentag! Ein guter Anlass, um eine #Rampe anzuschaffen, die #Barrieren überwindet“ | neutral | Keine Ironie | Kein Wechsel |
| „Immer mehr Studenten planen nach dem Bachelor den Einstieg in die Berufswelt - Mit dem üppigsten Anfängergehalt von mehr als 50.000 Euro rechnen Chemiker, Juristen und Mediziner – Ingenieure erwarten immerhin noch knapp 45.000. #Studium“ | neutral | Keine Ironie | Kein Wechsel |
| „Montag geht's wieder los: #Semesterbeginn, Start der #Vorlesungen, #Prüfungen und wenig Parkplätze auf dem Campus.“ | neutral | Keine Ironie | Kein Wechsel |
| „Ein Mitarbeiter vom #LKASachsen hat auf einer Pegida-Demo ein Kamerateam angezeigt, weil er nicht gefilmt werden wollte. Sein Gesicht ist innerhalb eines Tages nun so berühmt geworden, der könnte Werbung für #calvinklein machen.“ | neutral | Situative Ironie | Kein Wechsel |
| „Das Bett im Hotelzimmer ist toll und bequem, die Dusche aber viel zu winzig und unbequem.“ | mixed | Keine Ironie | Kein Wechsel |
| „Ich liebe den diesjährigen Sommer, Woche um Woche furchtbares Wetter.“ | mixed | Verbale Ironie | Wechsel |
| „Ich liebe es mit Migräne aufzuwachen☹.“ | mixed | Verbale Ironie | Wechsel |
| „So, die Zusammenfassung der 3. Woche #Prokrastinationsscheiße hat Osterbedingt etwas Verspätung, aber nichtsdestotrotz bin ich mit der Woche ziemlich zufrieden.“ | mixed | Keine Ironie | Kein Wechsel |