

Guidelines Sentiment Analysis DB

Stand 06.02.2017

Guidelines Annotation Sentiment Analysis

Annotation beschreibt den Vorgang der Markierung und Klassifizierung von Elementen in Texten. In diesen Guidelines wird die Annotation von Kundenkommentaren zum Themenkomplex Deutsche Bahn (DB) aus verschiedenen Quellen mit Hilfe des Annotationswerkzeugs *WebAnno* erklärt.

Zugang zu WebAnno

Der Zugang zu *WebAnno* erfolgt über

<https://clarin.ukp.informatik.tu-darmstadt.de/webanno-testing/login.html?0>

mit den persönlich ausgehändigten Zugangsdaten. Ein Klick auf *Annotation* öffnet die Übersicht der zu annotierenden Daten. Wähle eine Datei aus und klicke auf *Open*.

Grundsätzliches

Ziel der Annotationen ist es, Trainingsdaten für Sentiment Analysis bereitzustellen, um es der DB künftig zu ermöglichen, Stimmungen und Meinungen ihrer Kunden möglichst genau und auf firmeninterne Unterkategorien bezogen auszuwerten. Nicht von Belang sind hierbei von der DB eigens geschaltete Anzeigen und Erwähnungen von anderen Bahnunternehmen wie beispielsweise aus dem Ausland.

Annotationsebenen (Layer)

Die Annotation ist in drei Stufen untergliedert: *Relevanz*, *Sentiment* und *Relation*. *Sentiment* untergliedert sich in zwei Features: *Polarity* und *Category*.

Eine zusätzliche Ebene, genannt *Comment* eröffnet die Möglichkeit, bei Bedarf auf Sonder- und Problemfälle bezüglich des jeweiligen Abschnittes hinzuweisen. Diese können dann in der darauffolgenden Sitzung besprochen werden.

Die drei Annotationsebenen (Layer) werden im Folgenden erklärt.

Relevance

Auf der Annotationsebene *Relevance* soll die inhaltliche Relevanz des Kommentars im Hinblick auf den Themenbereich DB bewertet werden. Besteht kein Bezug zum genannten Themenbereich oder handelt es sich offensichtlich um von der DB selbst geschaltete Anzeigen, werden die Hashtags am Anfang des Dokuments mit *irr* annotiert. Wird die DB zwar erwähnt, ist aber nicht eigentliches Thema des Beitrags, werden die genannten Hashtags mit *no sent* annotiert. Zumeist stellt in solchen Fällen „Bahn fahren“ den Hintergrund für das eigentliche Geschehen dar. *No sent* wird auch gewählt, wenn der Beitrag kein DB-bezogenes Stimmungsbild transportiert. Aufpassen sollten die Annotatoren, wenn das „eigentliche Geschehen“ das Verhalten anderer Fahrgäste beschreibt. In solchen Fällen wäre die Äußerung der Kategorie „Atmosphäre“ zuzuordnen (s. **DB Sentiment**).

Annotation	
1	(Irrelevant) ID1234 Bahn will einen # Planungsverantwortlicher # Ingenieur (w/m) in # Brunswick , bewirb dich ! # jobs # stelle http://t.co/z1NWspVA0R

Sentiment

Auf der Annotationsebene *Sentiment* wird die Stimmung des Kommentars beschrieben.

Polarity

Drei Einordnungen sind möglich: *pos* für positive Stimmung, *neg* für negative Stimmung und *neut* für neutrales Feedback. Mit Polarity-Tags werden Wörter ausgezeichnet, die für die positive, negative oder neutrale Wirkung des Textabschnitts „verantwortlich“ sind.

Annotation	
79	##### Schöner neuer Welcome-Screen für ICE-WLAN , @Telekom_hilft @DB_Bahn ! Bemerkenswert : Wettervorhersage für den Zug ? ;-) https://t.co/c0TAacHpBZ

Hier: „schöner“, „neuer“

Category

Das Feature *Category* bietet die differenziertesten Auswahlmöglichkeiten. Zwischen 17 verschiedenen Kategorien, die jeweils Teilbereiche um das Themenfeld DB abdecken, kann hier gewählt werden.

Annotation	
79	##### Schöner neuer Welcome-Screen für ICE-WLAN , @Telekom_hilft @DB_Bahn ! Bemerkenswert : Wettervorhersage für den Zug ? ;-) https://t.co/c0TAacHpBZ

Hier: „Welcome-Screen“, „ICE-WLAN“

Die folgende Tabelle enthält eine Beschreibung aller möglichen Kategorien. Überkategorien sind fettgedruckt, Unterkategorien nicht. Wenn keine der beschriebenen Kategorien zu passen scheint, ist das eine Unterkategorie von *fehlende Kategorie* zu wählen.

Kategorie	Beschreibung
Allgemein	Allgemeine Aussage über die DB, die keiner anderen Kategorie zugeordnet werden kann.
Atmosphäre	Die Kategorie „Atmosphäre“ gruppiert Feedback zu Themen wie zum Beispiel zur angenehmen und entspannten Fahrt zum Zielort, zum Lärmpegel, zu Störgeräuschen durch den Zug oder Belästigung durch andere Fahrgäste, zur Möglichkeit der Konzentration oder der Interaktion mit anderen Reisenden.

Lautstärke	Die Unterkategorie „Lautstärke“ der Kategorie „Atmosphäre“ gruppiert Feedback zu Störgeräuschen durch den Zug, Lautstärke der Durchsagen oder Belästigung durch andere Fahrgäste.
Beleuchtung	Die Unterkategorie „Beleuchtung“ beinhaltet Feedback zum Thema Licht, Beleuchtung etc. Z. B. Licht ist zu grell, zu hell, zu kaltes Licht.
Fahrgefühl	Die Kategorie „Fahrgefühl“ gruppiert Feedback zu Themen wie Wackeln oder Schaukeln des Zuges und (un)ruhige Fahrt.
Temperatur	Die Kategorie „Temperatur“ gruppiert Feedback zu Themen wie zum Beispiel zur Funktionsfähigkeit der Klimaanlage, der Belüftung, Zugluft oder zur wahrgenommenen Temperatur.
Sauberkeit allgemein	Die Unterkategorie „Sauberkeit allgemein“ gruppiert Feedback zu Themen wie zum Beispiel zur Sauberkeit des Zuges außen und innen (außer Sitzplatz, Toilette und Bordbistro), zu Beschädigungen (bspw. Graffiti), zur Sauberkeit im Fahrgastraum, in den Gängen und den Eingangsbereichen oder zur Benutzbarkeit von Abfallbehältern.
Geruch	Die Unterkategorie „Geruch“ gruppiert Feedback zum Thema Geruch, Gestank und (stickige) Luft.
Sonstiges	Die Unterkategorie „Sonstiges“ der Kategorie „Atmosphäre“ gruppiert sonstiges Feedback zur Atmosphäre.
Connectivity	Die Kategorie „Connectivity“ gruppiert Feedback zu Themen wie zum Beispiel zu Vorhandensein, Geschwindigkeit und Stabilität des WLANs und der mobilen Internetverbindung, zu Info- und Entertainmentangeboten (ICE Portal) oder zur Verfügbarkeit und Qualität des Handyempfangs.
WLAN/Internet	Alles zum Thema WLAN/Internet. Sowohl „Angebot vorhanden“, als auch Verfügbarkeit des WLAN/Internet-Angebots, Empfang, Nutzbarkeit des Internets via WLAN, insbesondere in Bezug auf Geschwindigkeit und Verbindungsstabilität. Auch das Thema „kostenloses WLAN“, zum kostenlosen Internet der 1.Klasse als auch Feedback zur Preisgestaltung des WLAN-Angebots in der 2.Klasse findet sich hier wieder.
Telefonie/Handyempfang	Die Unterkategorie „Telefonie/Handyempfang“ der Kategorie „Connectivity“ gruppiert Feedback zur Nutzbarkeit der Telefonie während der Zugfahrt, insbesondere in Bezug auf Verfügbarkeit und Qualität des Handyempfangs.
ICE Portal	Die Unterkategorie „ICE Portal“ gruppiert Feedback zum ICE Portal (Info- und Entertainmentportal), wie Verfügbarkeit, technische Probleme, Ladezeit und Inhalt.
Sonstiges	Die Unterkategorie „Sonstiges“ der Kategorie „Connectivity“ gruppiert sonstiges Feedback zur Connectivity.
Design	Die Kategorie „Design“ gruppiert Feedback zu Themen wie zum Beispiel der optischen Gestaltung, Innengestaltung, wie Teppich sowie zum allgemeinen Aussehen des Zuges. Bei Formulierungen wie „neuer Zug“ soll die Kategorie nur gewählt werden, wenn die Formulierung alleinsteht, nicht wenn weiter spezifiziert wird wie in „neuer Zug, trotzdem unpünktlich...“.
Gastronomisches Angebot	Die Kategorie „Gastronomisches Angebot“ gruppiert Feedback

	zu Themen wie zum Beispiel zum Gastroservice und dessen Verfügbarkeit, zur Auswahl und Qualität von Speisen und Getränken auf der Speisekarte und deren Preise, zur Sauberkeit und zum Service im Bordbistro, zur Wartezeit oder den Erfahrungen während des Aufenthalts im Bordbistro sowie zum Service am Platz.
Verfügbarkeit Bordbistro/-restaurant	Die Unterkategorie „Verfügbarkeit Bordbistro/-restaurant“ der Kategorie „Gastronomisches Angebot“ gruppiert Feedback zur Verfügbarkeit des Gastroservices, wie Ausfall der Bordgastronomie.
Verfügbarkeit angebotener Produkte	Die Unterkategorie „Verfügbarkeit angebotener Produkte“ der Kategorie „Gastronomisches Angebot“ gruppiert Feedback zur Verfügbarkeit der angebotenen Speisen und Getränke, wie fehlende, nicht nachgelieferte Produkte und kleine technische Defekte.
Vielfalt/Auswahl	Die Unterkategorie „Vielfalt/Auswahl“ der Kategorie „Gastronomisches Angebot“ gruppiert Feedback zur Auswahl der angebotenen Speisen und Getränke.
Preise	Die Unterkategorie „Preise“ der Kategorie „Gastronomisches Angebot“ gruppiert Feedback zu den Preisen der angebotenen Speisen und Getränke.
Gastronomiebetreuung	Die Unterkategorie „Gastronomiebetreuung“ der Kategorie „Gastronomisches Angebot“ gruppiert sowohl Feedback zur Freundlichkeit, Höflichkeit des Gastronomiepersonals als auch zur Kompetenz und Auskunftsfähigkeit.
Sonstiges	Die Unterkategorie „Sonstiges“ der Kategorie „Gastronomisches Angebot“ gruppiert sonstiges Feedback zum gastronomischen Angebot, wie z. B. Ambiente im Bordbistro/-restaurant.
Informationen	Die Kategorie „Information“ gruppiert Feedback zu Themen wie zum Beispiel zu Inhalt, Zeitpunkt und Verständlichkeit der Fahrgastinformationen in Durchsagen, durch das Personal oder auf Anzeigetafeln, sowohl im Regelfall als auch im Störfall, im Zug oder am Bahnhof. Es geht um Informationen zu Verspätungen, dem Grund der Verspätung, Gleisänderungen, Bord Bistro oder Informationen über Anschlussmöglichkeiten.
DB App und Website	Die Kategorie „DB App und Website“ gruppiert Feedback zur Nutzung der DB App und der DB Homepage.
Informationen DB App und Website	Die Unterkategorie „Informationen DB App und Website“ der Kategorie „DB App und Website“ gruppiert Feedback über die Möglichkeiten, sich via DB App und DB Website zu informieren, erhaltene Informationen und Fehlinformationen.
Störungen DB App und Website	Die Unterkategorie „Störungen DB App und Website“ der Kategorie „DB App und Website“ gruppiert Feedback über ordnungsgemäßes Funktionieren und technische Störungen der DB App und der DB Website.
Service/Kundenbetreuung	Die Kategorie „Service/Kundenbetreuung“ gruppiert Feedback zu Themen wie zum Beispiel zur Freundlichkeit, Kompetenz (Auskunftsfähigkeit) und Verfügbarkeit des Zugpersonals oder zur Verfügbarkeit der angebotenen Zeitungen.
Zugbetreuung	Die Unterkategorie „Zugbetreuung“ der Kategorie „Service / Kundenbetreuung“ gruppiert sowohl Feedback zur Freundlichkeit, Höflichkeit des Zugpersonals als auch zur

	Kompetenz und Auskunftsfähigkeit.
Am-Platz-Service/1. Klasse-Service	Die Unterkategorie „Am-Platz-Service/1. Klasse-Service“ der Kategorie „Service/Kundenbetreuung“ gruppiert alles zum 1. Klasse Service inkl. Am-Platz-Service (bereits im gastronomischen Angebot), Themen wie Wartezeit und Regelmäßigkeit des Am-Platz-Services (nicht die Verfügbarkeit des Gastroservices im Allgemeinen). Zudem werden Zeitungsservice, wie Verfügbarkeit und Auswahl der am Platz angebotenen Zeitungen sowie Goodies erfasst.
Sonstiges	Die Unterkategorie „Sonstiges“ der Kategorie „Service“ gruppiert sonstiges Feedback zum Service.
Komfort/Ausstattung	Beinhaltet allgemeines Feedback zu den Themen Komfort, wie Komfort im Zug, komfortables Reisen, Treppen, allgemeine Geräumigkeit und der Ausstattung insbesondere am Sitzplatz wie zum Beispiel reservierten und nicht reservierten Sitzplätzen, Bahn.Comfort-Sitzplätzen, der Reservierungsanzeige, Beinfreiheit, Tischen, Bestuhlung, Steckdosen oder Gepäckablagen. (Nicht wenn „enger Sitzplatzabstand“ - nicht wenn konkreter Bezug zum Sitz.) Außerdem wird Feedback zur Kennzeichnung zu Bereichen wie den Ruhe- oder Familienbereichen erfasst.
Sitzkomfort	Die Unterkategorie „Sitzkomfort“ der Kategorie „Komfort/Ausstattung“ gruppiert Feedback zum konkreten Sitzkomfort und zu Fußstützen, Beinfreiheit und Sitzplatzabstand.
Funktionsfähigkeit Sitz und Sitzverstellbarkeit	Die Unterkategorie „Funktionsfähigkeit Sitz und Sitzverstellbarkeit“ der Kategorie „Komfort/ Ausstattung“ gruppiert Feedback zur Verstellbarkeit der Sitze, Rückenlehne und Kopfstütze und die Funktionsfähigkeit.
Reservierung	Die Unterkategorie „Reservierung“ der Kategorie „Komfort/Ausstattung“ gruppiert Feedback zu Reservierungsthemen wie Belegung des reservierten Sitzplatzes, Reservierungsanzeige auch Störungen/Ausfall und Doppeltvergabe von Sitzplatzreservierungen.
Steckdosen	Die Unterkategorie „Steckdosen“ der Kategorie „Komfort/Ausstattung“ gruppiert Feedback zur Verfügbarkeit, Beschaffenheit, Funktionsfähigkeit und Anzahl der Steckdosen am Sitzplatz.
Kleiderhaken	Die Unterkategorie „Kleiderhaken“ der Kategorie „Komfort/Ausstattung“ gruppiert Feedback zu Kleiderhaken, Aufhängen der Jacken etc.
Sauberkeit Sitzplatz	Die Unterkategorie „Sauberkeit Sitzplatz“ der Kategorie „Komfort/Ausstattung“ gruppiert Feedback zu Feedback zum Thema Sauberkeit am Sitzplatz, wie Krümel, unaufgeräumter Sitzplatz, Sitzplatz liegt voll mit Zeitungen/Müll, voller Mülleimer am Platz etc.
Sonstiges	Die Unterkategorie „Sonstiges“ der Kategorie „Komfort/Ausstattung“ gruppiert sonstiges Feedback zum Komfort und der Ausstattung im Zug, z.B. Verfügbarkeit von bahn.comfort Sitzplätzen, Kennzeichnung von (Ruhe-)Bereichen, oder Feedback zu den Treppen (IC2).
Gepäck	Die Kategorie „Gepäck“ gruppiert Feedback zur Verfügbarkeit

	von ausreichend viel Stauraum für Gepäck in der Nähe des Sitzplatzes.
Auslastung und Platzangebot	Die Kategorie „Auslastung/Platzangebot“ gruppiert sowohl Feedback zur Verfügbarkeit von Sitzplätzen als auch zu ausreichender/ungenügender Wagenzahl, inkl. Zugteil/Wagen vorhanden, als auch Feedback zur Verfügbarkeit und Auffindbarkeit des vom Kunden gewählten Zugteils/Wagen entsprechend der Reservierung.
Ticketkauf	Die Kategorie „Ticketkauf“ gruppiert Feedback zu Themen wie zum Beispiel zum Aufwand des Ticketkaufs, zu den Vertriebskanälen (online, Reisecenter, Automat, im Zug), zu den angebotenen Preisinformationen, zum Preis-Leistungs-Verhältnis der gemachten Reise oder zum Prozess der Ticketkontrolle.
Toiletten	Die Kategorie „Toiletten“ gruppiert Feedback zu Themen wie zum Beispiel Funktionsfähigkeit, Sauberkeit, Geruch oder Hygiene der Toiletten und zur funktionalen Gestaltung der Toilette, sowie Verfügbarkeit des Verbrauchsmaterials wie Seife, Papierhandtücher oder Toilettenpapier.
Funktionsfähigkeit Toiletten	Die Unterkategorie „Funktionsfähigkeit Toiletten“ der Kategorie „Toiletten“ gruppiert Feedback zur Funktionsfähigkeit der Toilette, insbesondere in Bezug auf WC-Spülung, Waschbeckenfunktion (inkl. Spender) der Inneneinrichtung und gesperrte Toiletten.
Sauberkeit Toilette	Die Unterkategorie „Sauberkeit Toilette“ der Kategorie „Toiletten“ gruppiert Feedback zur Sauberkeit des WCs, Waschtisch, Boden und Inneneinrichtung.
Geruch Toilette	"Geruch Toilette“ gruppiert Feedback zum Geruch in der Toilette. Wenn es im Fahrgastraum nach Toilette riecht, fällt dies nicht in „Geruch Toilette“.
Verfügbarkeit Verbrauchsmaterial	Die Unterkategorie „Verfügbarkeit Verbrauchsmaterial“ der Kategorie „Toiletten“ gruppiert Feedback zu Verbrauchsmaterial wie Papierhandtücher, Toilettenpapier, Flüssigseife und Desinfektionsmittel.
Sonstiges	Die Unterkategorie „Sonstiges“ der Kategorie „Toiletten“ gruppiert sonstiges Feedback zu den Toiletten.
Zugfahrt	Die Kategorie „Zugfahrt“ gruppiert Feedback zu Themen wie zum Beispiel Verspätungen aller Art oder der Pünktlichkeit bei Erreichung des Zielortes sowie Anchlusserreichung und ob Anschlüsse fraglich sind oder verpasst wurden.
Pünktlichkeit	Die Kategorie „Pünktlichkeit“ gruppiert Feedback zu Themen wie zum Beispiel Verspätungen aller Art oder der Pünktlichkeit bei Erreichung des Zielortes.
Anchlusserreichung	Die Kategorie „Anschluss“ gruppiert Feedback zu Themen wie zum Beispiel das Erreichen oder Verpassen vorgesehener Anschlüsse oder zur Verfügungstellung einer adäquaten Unterkunft. Wenn Pünktlichkeit und Anschluss gleichzeitig genannt werden, wird nur Kategorie „Anschluss“ gewählt.
Technische Schäden / Störungen am Zug	Die Unterkategorie „Technische Schäden/Störungen am Zug“ gruppiert Feedback zu Schäden am Zug und technische

	<p>Probleme.</p> <p>Ausgenommen sind Schäden des Bordbistros und der Toilette sowie Störungen der Displays und Reservierungsanzeigen.</p>
Wagenreihung	<p>Die Kategorie „Wagenreihung“ gruppiert Feedback zu Themen wie zum Beispiel zur Reihung und Orientierung der Züge.</p> <p>Kein Feedback zur Information über die Wagenreihung.</p>
Fahrtzeit / Schnelligkeit	<p>Feedback zur Fahrtzeit und Schnelligkeit des Zuges, betrifft Geschwindigkeit, kurze Fahrtzeit, x Stunden von hier nach da, etc.</p>
Streckennetz	<p>Die Unterkategorie „Streckennetz“ der Kategorie „Zugfahrt“ umfasst Feedback zum Streckennetz, Anbindung einzelner Orte und zur Taktung der Züge.</p>
Sonstige Unregelmäßigkeiten	<p>Die Unterkategorie „Sonstige Unregelmäßigkeiten“ gruppiert Feedback zum Thema „Sonstige Unregelmäßigkeiten“, wie Bereitstellung und außerplanmäßiger Stillstand des Zuges.</p> <p>Ausgenommen sind die Themen Pünktlichkeit/Verspätung, Anschluss, technische Störungen/Schäden am Zug und Wagenreihung.</p>
Reisen mit Kindern	<p>Die Kategorie „Reisen mit Kindern“ gruppiert Feedback zum Kleinkindbereich, wie Gestaltung, Platzierung, Platz und generellen Themen beim Reisen mit Kindern.</p>
Image	<p>Die Kategorie „Image“ gruppiert Feedback zum Bild der DB in der Öffentlichkeit.</p>
Sponsoring	<p>Die Unterkategorie „Sponsoring“ der Kategorie „Image“ gruppiert Feedback zum Thema Sponsoring.</p>
Marketing	<p>Die Unterkategorie „Marketing“ der Kategorie „Image“ gruppiert Feedback zu Marketing-Kampagnen wie DB-Werbung auf Plakatwänden, Werbespots und Aktionen zur Kundenzufriedenheit.</p>
QR-Code	<p>Die Kategorie „QR-Code“ gruppiert Nennungen zum QR Code, der den Kunden Möglichkeiten eröffnet, sich zu beschweren und Meinungen zu äußern.</p>
Barrierefreiheit	<p>Die Kategorie „Barrierefreiheit“ gruppiert Feedback zur barrierefreien Fortbewegung in den Zügen und am Bahnhof.</p>
Sicherheit	<p>Die Kategorie „Sicherheit“ gruppiert Feedback zur Sicherheit im Zug und am Bahnhof.</p>
Unklare Kategorie	<p>Die Kategorie „Unklare Kategorie“ gruppiert Feedback, das nicht zugeordnet werden kann, weil entweder keine passende Kategorie existiert oder das Dokument ungünstig „abgeschnitten“ wurde, sodass dem Leser die nötigen Informationen fehlen, um das Dokument zuordnen zu können.</p>
Nicht vorhandene Kategorie	<p>Die Unterkategorie „Nicht vorhandene Kategorie“ der Kategorie „Unklare Kategorie“ gruppiert Feedback, für das keine passende Kategorie existiert.</p>
Nicht identifizierbare Kategorie	<p>Die Unterkategorie „Nicht identifizierbare Kategorie“ der Kategorie „Unklare Kategorie“ gruppiert Feedback, für das keine passende Kategorie existiert.</p>
No sent (im Layer „Relevance“)	<p>Dieses Tag ist für die Annotation von Beiträgen gedacht, die kein DB-bezogenes Sentiment transportieren. Dies trifft hauptsächlich auf Dokumente zu, in denen die DB zwar erwähnt wird, aber nicht eigentliches Thema des Beitrags ist. Zumeist stellt in solchen Fällen „Bahn fahren“ den Hintergrund für das</p>

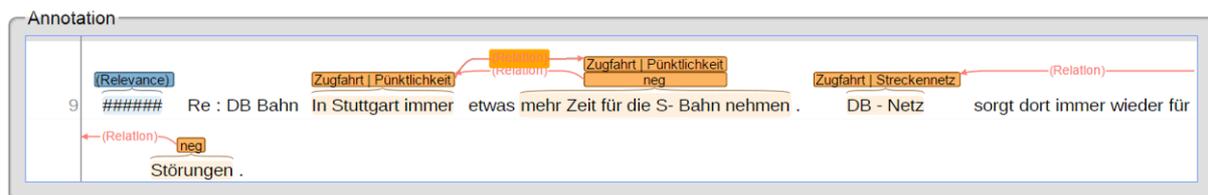
	eigentliche Geschehen dar. Aufpassen sollten die Annotatoren, wenn das „eigentliche Geschehen“ das Verhalten anderer Fahrgäste beschreibt. In solchen Fällen wäre die Äußerung der Kategorie „Atmosphäre“ zuzuordnen.
Irr (im Layer „Relevance“)	„Irr“ gruppiert irrelevante Dokumente, die die DB nicht thematisieren, als auch von der DB geschaltete Werbe- und Stellenanzeigen.

Relation

Inhaltlich zusammenhängende Token, die mit einem *Category*-Tag ausgezeichnet sind, werden mit Relationsbögen verbunden. Hierfür wird der Mauszeiger gehalten von einem zum anderen *Category*-Tag gezogen. Inhaltlich zusammenhängende Token derselben *Category* werden von links nach rechts verbunden. Um kenntlich zu machen, auf welche *Category* sich ein *Polarity*-Tag bezieht, wird ein Relationsbogen vom *Polarity*- zum *Category*-Tag gezogen. Befindet sich das *Category*-Tag in einer Kette aus mehreren, durch Relationsbögen verbundenen *Category*-Tags, so wird **nur ein** Relationsbogen zum **ersten** *Category*-Tag gezogen.



Sind einem *Polarity*-Tag, das einen Bezug zu anderen Token aufweist, ein oder mehrere *Category*-Tags zugeordnet, so muss darauf geachtet werden, *Polarity*- und *Category*-Tags zu „stapeln“, da es sonst nicht möglich ist, die Regeln zur Ausrichtung von Relationsbögen einzuhalten. Hierfür wird das Token, nachdem das erste Tag ausgewählt wurde, erneut per Doppelklick aktiviert, bevor schließlich das zweite Tag ausgewählt wird.

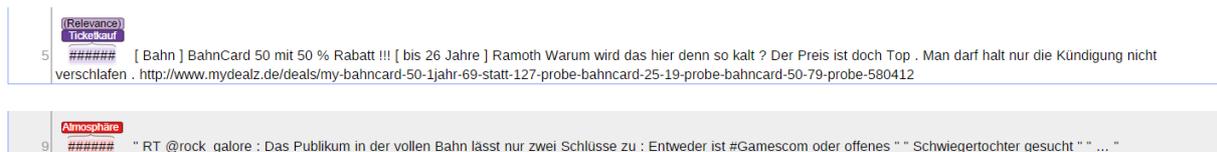


Annotationen

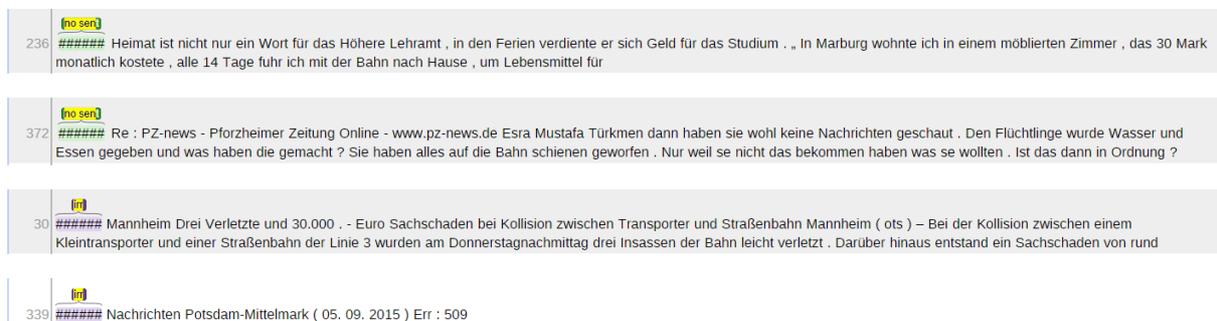
Es werden zwei Annotationsarten unterschieden, die im Folgenden *Kategorie-Annotationen* und *feine Annotationen* genannt werden.

Kategorie-Annotation

Der Begriff *Kategorie-Annotation* bezeichnet die Annotation eines Dokuments mit einer DB Kategorie. Um ein Dokument entsprechend auszuzeichnen, erfolgt ein Doppelklick auf die Hashtags am Anfang des jeweiligen Dokuments. Die *Actions*-Box auf der rechten Seite des Bildschirms ist nun aktiviert. Im Layer *Sentiment* wird nun eine DB Überkategorie ausgewählt. Existieren für eine Überkategorie eine oder mehrere Unterkategorien, öffnet sich nach der Auswahl einer Überkategorie ein zusätzliches Dropdown-Menü mit Unterkategorien. Zur jeweiligen Überkategorie passende Unterkategorien sind fettgedruckt. Die *Relevance*-Markierung muss nicht gelöscht werden. Gehört ein Dokument mehreren Kategorien an, werden alle passenden Kategorien ausgewählt.



Handelt es sich bei dem Dokument um irrelevantes Material oder wird die DB nur als Hintergrund-Information erwähnt, wird zwischen den Kategorien *irr* und *no sent* gewählt. Definiert sind diese im Abschnitt **Category**. Werden die Kategorien *irr* oder *no sent* gewählt, muss die vorannotierte Markierung mit *Relevance* gelöscht werden. Dies erfolgt über einen Doppelklick auf *Relevance*, in der *Actions*-Box auf der rechten Seite kann nun im Layer *Relevance* eine Kategorie ausgewählt werden.



Die Kategorien *irr*, *no sent* und *unklare Kategorie* im Vergleich

Mit den Kategorien *irr*, *no sent* und *unklare Kategorie* werden Dokumente markiert, die keiner DB-Kategorie zugeordnet werden können und sich auch nicht auf die Bahn als Ganzes beziehen. Im Folgenden werden die Unterschiede zwischen den drei Annotationsmöglichkeiten eingehender erläutert.

Irr

irr steht für „irrelevant“. In diese Kategorie fallen Dokumente, die nichts mit der DB zu tun haben, Stellenanzeigen der DB und Werbung von der DB selbst. Datenschnipsel, aus denen nicht hervorgeht, ob die DB überhaupt thematisiert wird, sind ebenfalls mit *irr* zu annotieren.

no sent

no sent steht für „no sentiment“ und ist für Dokumente gedacht, in denen der Themenkomplex DB nur Hintergrundinformation darstellt, als auch für Dokumente, die zwar die DB zum Thema haben, aber weder positives noch negatives Feedback transportieren. Häufig ist dies der Fall bei informativen Texten und Fragen. Der Unterschied zu *neut* besteht darin, dass kein neutrales, sondern überhaupt kein Sentiment transportiert wird. Da es durchaus vorkommen kann, dass ein Dokument zwar keine Polarität transportiert, jedoch DB-Kategorien behandelt, kann *no sent* mit DB-Kategorien kombiniert werden. *no sent* wird dann auf die Hashtags am Anfang des Dokuments gesetzt, die DB-Kategorie auf das jeweilige Signalwort.

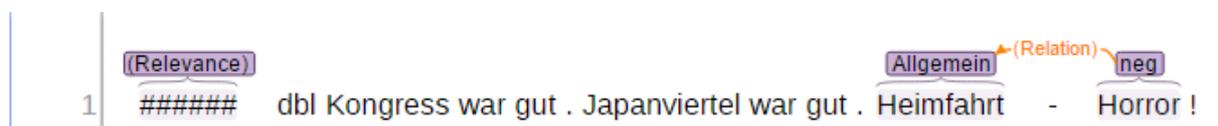
Unklare Kategorie

Unklare Kategorie enthält zwei Untergruppen: *nicht identifizierbare Kategorie* für Datenschnipsel, die die DB zwar thematisieren, jedoch keine Rückschlüsse auf die Kategorie zulassen, und *nicht vorhandene Kategorie* Dokumente, deren Kategorie klar zu erkennen, aber nicht in der Liste der zu wählenden Kategorien zu finden ist.



Feine Annotation

Feine Annotation bezeichnet die detaillierte Annotation eines Dokuments mit *Polarity*- und *Category*-Werten sowie deren Verhältnis zueinander. Die DB Kategorie wird dem Wort oder der Wortgruppe im Text zugeordnet, die die Zuordnung zur jeweiligen Kategorie ermöglicht. Der *Polarity*-Wert wird dem Wort oder der Wortgruppe zugeordnet, die das positive, negative oder neutrale Sentiment erzeugt.



Die annotierten Abschnitte müssen linguistisch nicht alleine stehen können, dürfen für sich gelesen also durchaus „unsinnig“ erscheinen. Allein der inhaltliche Wert eines Wortes in Bezug auf die Annotationsebenen ist von Interesse.



Die Kategorien *irr*, *no sent* und *nicht identifizierbare Kategorie* werden nicht fein annotiert, sondern stets als *Kategorie-Annotation* behandelt.

60 **no sent** ##### Polizeischutz für G7-Treffen in Südbayern und Bilderberg-Konferenz in Telfs REGION . Kein Fleckchen Wald , Wiese , Berg und schon gar nicht Wege und Straßen oder die Bahnstrecke bleiben in diesen Tagen von der Polizei unbeobachtet , selbst der Luftraum wird überwacht : Bahn fre

26 **tr** ##### Nächstes Mal #Bibelteilen : Mo , 8. 6. , 19:30 , Frauenkirche Baden (Endstation Badner Bahn) . Freu mich aufs gemeinsame Bibellesen ! :) CC @kjwien

Annotation von Ironie

Ironie wird nur dann annotiert, wenn sie sprachlich klar festzumachen ist und einer Kategorie zuzuordnen ist. Häufig ist das der Fall, wenn Wörter mit positiver oder negativer Bedeutung neben Abtönungspartikeln stehen.

Beispiel:

"Das hat die DB ja mal wieder toll hingekriegt!" [ja mal wieder toll hingekriegt]

Besteht nur eine geringe Möglichkeit, dass die Äußerung ernst gemeint war (Beispiel: "DB, ich liebe dich!"), wird diese nicht mit *Polarity*-Werten versehen.

Annotation von Vorschlägen

Vorschläge werden meist neutral vorgetragen und sollten dann auch als solche annotiert werden. Die Tokens, aus denen hervorgeht, dass es sich um einen Vorschlag handelt („sollte“, „könnte“, „wäre besser“, etc.) werden mit dem *Polarity*-Wert *neut* markiert und durch *Relation*-Bögen mit dem *Category*-Tag verbunden, auf das sie sich beziehen.

Gelnhausen gebrochen , und zwischen Gelnhausen und Wächtersbach führen eigene kurzläufige RBs . Und genau deshalb sollte man den sinnvollen **neut** **pos** Zugfahrt | Streckennetz Laufweg der RB

Frankfurt-Wächtersbach auch auf neue RBs Frankfurt-Heigenbrücken , ohne lange Aufenthalte in Hanau oder Aschaffenburg , übertragen . Dazu natürlich modernisierte 76-cm-Bahnsteige **neut**

Annotation von Vorschlägen, die negatives Sentiment transportieren

Auch bei Vorschlägen, die negatives Sentiment transportieren, greift das oben beschriebene Schema. Allerdings muss das erste Glied der Kette dann aus einem negativen Polaritätswert bestehen. Besteht das erste Glied wie im vorliegenden Fall nicht aus einem negativen Polaritätswert (hier: „kann da bitte“ ist „neut“), so wird das Tag, das das negative Sentiment des Dokuments bewirkt, per Relationsbogen mit dem ersten Tag verbunden (hier: „schafft das nicht“ mit „kann da bitte“).

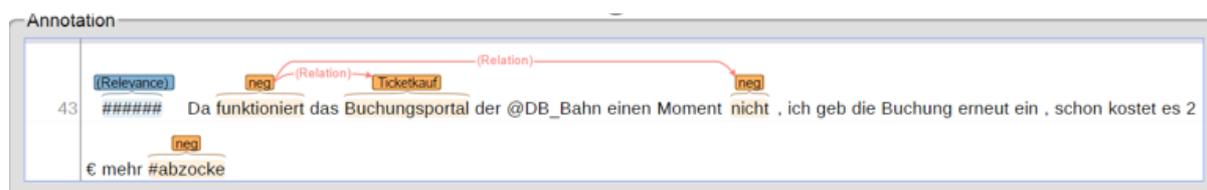
78 **(Relevance)** ##### @DB_Bahn - ICE 597 - Wagen 4 - kann da bitte das **neut** Connectivity | WLAN und Internet **pos** Connectivity | WLAN und Internet mit dem Internet verbunden werden ? Danke . Das **neg** Service und Kundenbetreuung Personal am Boden schafft das nicht .

Annotation von U-Bahnen, S-Bahnen und Straßenbahnen

S-Bahnen gehören in einigen deutschen Städten zur DB, weshalb Dokumente, die diese behandeln, annotiert werden. Anders verhält es sich mit Straßenbahnen und U-Bahnen. Dokumente, die sich auf diese beziehen, werden mit *irr* ausgezeichnet.

Annotation von Verneinungen

Verneinungen mit dem Partikel „nicht“ werden wie in folgendem Beispiel annotiert.



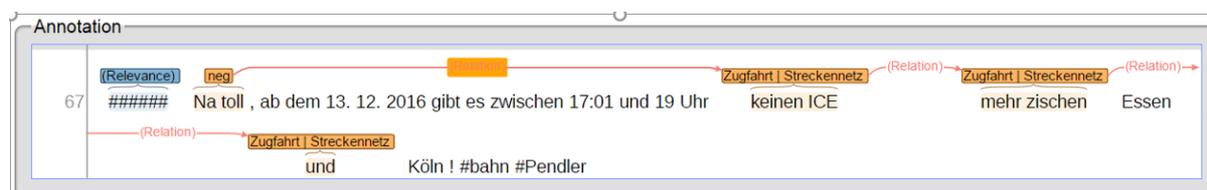
Sowohl „nicht“ als auch das Wort, auf das es sich bezieht, werden markiert, mit einem Relationsbogen verbunden und mit dem passenden (gemeinsamen) *Polarity*-Wert versehen.

Annotation von „DB“, „Deutsche Bahn“, etc.

Als relevant geltende Dokumente haben immer etwas mit der DB zu tun. Es ist daher in den meisten Fällen nicht nötig, den Eigennamen zu markieren. Ausnahmen sind die Kategorie „Allgemein“, Komposita („DB-Navigator“) und Konstruktionen, in denen die namentliche Erwähnung der DB zur Kategorisierung notwendige Information transportiert.

Annotation von Eigennamen (NE) und Zahlen

Eigennamen und Zahlen werden nach Möglichkeit ausgespart, es sei denn, sie haben eine inhaltliche Bedeutung, die über die bloße Minutenzahl, bzw. die Ortsangabe, hinausgeht. Ein Beispiel für eine Zahl, die im DB-Zusammenhang eine zusätzliche Bedeutung hat, sind die häufig erwähnten „85 %“, mit denen auf das Ziel der DB, 85 % Pünktlichkeit zu gewährleisten, angespielt wird.



Annotation von indirektem Feedback

Als indirektes Kundenfeedback gilt Feedback, das nicht direkt vom Kunden ausgesprochen wird, sondern durch andere Personen, beispielsweise DB-Mitarbeiter, transportiert wird

(Beispiel: „Die Kunden beschwerten sich über das Verhalten unserer Fahrkartenkontrolleure.“). Dieses indirekte Feedback wird wie direktes Feedback annotiert.

Annotation von Emoticons

Emoticons, also Smileys im weitesten Sinne, werden mit ihren jeweiligen Polaritäten annotiert.



Annotation von langen und komplexen Dokumenten

Lange und komplexe Dokumente, bei denen man leicht den Überblick verliert, lassen sich leichter bearbeiten, indem man als erstes alle Polaritätswörter markiert, da sich diese leichter identifizieren lassen. Im zweiten Schritt werden dann die Wörter, auf die sie sich beziehen, markiert und mit der entsprechenden Kategorie versehen. Ist keine passende DB- Kategorie ausfindig zu machen, ist entweder die Kategorie „nicht identifizierbare Kategorie“ oder „nicht vorhandene Kategorie“ zu wählen.

Es geht nicht darum, vollständige Geschichten nachzuerzählen. Wichtig ist, dass Polaritäten und Kategorien erkannt werden.



Doppelte Dokumente

Durch Retweets etc. kann es vorkommen, dass sich Dokumente innerhalb einer Datei inhaltlich gleichen. Die Dopplungen werden beim Annotieren nicht ausgelassen, sondern wiederholt annotiert.

Vergleiche

Bei der Annotation von Vergleichen wie „schöner als...“, „langsamer als“, etc. ist jeweils nur das Adjektiv und der Vergleichspartikel zu markieren.



Annotationen der Kategorie „Atmosphäre“

Dokumente der Kategorie „Atmosphäre“ sind in aller Regel so aufgebaut, dass ein Mitfahrer („der Typ in der S-Bahn“, „die Assis neben mir“, ...) erwähnt wird, auf den sich mit positiv oder negativ konnotierten Beschreibungen bezogen wird. Das „Atmosphäre“-Tag sollte stets „auf“ der beschriebenen Person („Assis“, „Typ neben mir“, etc.) angebracht werden. Meist handelt es sich hierbei auch um das Subjekt.



Umfang der Span-Annotationen

Es dürfen jeweils nur diejenigen Tokens markiert werden, die für die Zuordnung zu einer Kategorie einer Polarität benötigt werden. Das impliziert, dass Artikel, Pronomina und Partikel (auch Gradpartikel) nicht umschlossen werden sollten. Auch Adverbien beziehen sich häufig nur auf die Intensität einer Polarität und sind dann ebenfalls nicht zu annotieren. Selbiges gilt meist auch für Modal- und Hilfsverben.

Praktisch wird aus „so schön“ „schön“, aus „gar nicht gut“ wird „nicht gut“, aus „immerhin nicht kaputt“ wird „nicht kaputt“, aus „sehr schön“ wird „schön“, aus „wird nicht besser“ wird „nicht besser“, aus „gelingt mal wieder gar nicht“ wird „gelingt nicht“, aus „hat damit keine Erfahrung“ wird „keine Erfahrung“.

Eine gute Hilfe ist es, sich während des Annotierens vorzustellen, man würde ein Wörterbuch mit positiv und negativ konnotierten Begriffen sowie Kategoriewörtern füllen.



Hier: „werden“ wird nicht benötigt.



Hier: „so“ wird nicht benötigt.

Token mit Polarity- als auch Category-Tags

Ist ein Wort sowohl mit Category-Tags als auch mit einem Polarity-Tag auszuzeichnen, so müssen diese in separaten Tags annotiert werden. Hierfür ist das betroffene Token nach der Auswahl des ersten Tags abermals zu markieren (Doppelklick oder mit gedrückter Maustaste über das Token fahren). Anschließend kann das gewünschte zweite Tag ausgewählt werden.



Annotation von Satzzeichen

Satzzeichen werden bei der Annotation ausgespart.

Annotation von Zugausfällen

Nicht bereitgestellte Züge sind der Kategorie „Sonstige Unregelmäßigkeiten“ zuzuordnen.

Annotation fremdsprachlichen Materials

Fremdsprachliches Material wird bei der Annotation ausgespart.

Überlappende Span-Annotationen

Umfasst ein Category-Tag ein Token, das mit einem Polarity-Tag annotiert wurde, so ist kein zusätzlicher Relationsbogen vom Polarity-Tag zum Category-Tag zu ziehen.

Annotation von „Oberleitungsstörungen“

„Oberleitungsstörungen“ fallen in die Kategorie „Sonstige Unregelmäßigkeiten“.

Annotation von „Deutsche Bahn“, „Bahn“ und „DB“

Dürfen nur in Eigennamen wie „DB Navigator“ mitmarkiert werden.

Annotation von „dass“

„dass“ sollte bei der Annotation ausgespart werden.

Annotation von „weder..noch“

„Weder...noch“ ist als Verneinung zu annotieren. Dabei werden sowohl „weder“ als auch „noch“ mit einem „neg“ versehen. Anschließend wird das erste Element der Kette mit allen Category-Tags verbunden, auf die sich die Verneinung bezieht.

Reaktionen auf Antworten seitens der Bahn („Danke!“)

Hin und wieder finden sich kurze Textschnipsel, die aus Gesprächsverläufen mit der DB stammen. Häufig bestehen diese aus nicht mehr als einem knappen „Danke“. Dieses „Danke“ ist zumeist nicht eindeutig einer Polarität zuzuordnen (vgl.: „Saftladen! Danke!“). Dokumente, die keine zusätzlichen negativ oder positiv besetzten Abschnitte aufweisen, müssen daher als „no sent“ getaggt werden.

Annotation von Dokumenten, deren Polarität nicht erkennbar ist

Dokumente, deren Polarität nicht klar erkennbar ist, werden nur mit Category-Tags versehen. Das Tag „Relevance“ wird zu „no sent“ geändert.

Annotation von Kategorien, die per Definition bereits negatives Sentiment transportieren

Kategorien, die per Definition bereits negatives Sentiment transportieren („Sonstige Unregelmäßigkeiten“, „technische Störungen“, etc.) werden stets mit einem Polarity-Tag versehen

Annotation von Umleitungen

Umleitungen haben einen Einfluss auf das Streckennetz und sollten daher „Streckennetz“ zugeordnet werden.

Annotation von Ersatzzügen

Ersatzzüge sind, sofern sie wie geplant fahren, **nicht** als „neg“ zu annotieren.

Annotation von Gründen

Gründe für Ausfälle u. ä. werden nur dann mitannotiert, wenn sie selbst eine DB Kategorie repräsentieren. Fällt ein Zug aus, weil ein technischer Defekt vorlag, wird dieser technische Defekt folglich annotiert. Fällt ein Zug aus, weil er von einem Auto gerammt wurde, wird das Auto bei der Annotation jedoch nicht berücksichtigt.

Bahnhöfe

Die Erwähnung eines deutschen Bahnhofs genügt, um eine Einordnung in „no sent“ zu legitimieren.

Annotation von Fernbus-Vergleichen etc.

Bei Vergleichen, die im Schema „Fernbus schneller als Bahn“ gehalten sind, sollte „Fernbus schneller“ als „neg“ und „schneller“ je nach Kontext mit dem entsprechenden Kategorietag versehen werden.

Streiks

Streiks fallen in die Kategorie „Unklare Kategorie | nicht vorhandene Kategorie“.

The screenshot shows an annotation interface with a text area containing the following text: "38 ##### Re : Initiative Neue Soziale Marktwirtschaft (INSM) @ Rudolf Stainer : Wenn Arbeitgeber ihre Interessen durch zu setzen versuchen ist das OK , Gewerkschaften dürfen das Ihrer Meinung nach nicht ? Übrigens waren die Bahnstreiks völlig überflüssig . Hätte die Bahn ihre Forderungen gar nicht erst gestellt bzw. früher aufgegeben , wäre überhaupt kein Schaden entstanden . Das haben die Bahnmanager zu verantworten - und hat man sie dafür zur Verantwortung gezogen ? Und was die Geldverwendung angeht . Wenn ich etwas kaufe was aus der Metall- oder Elektroindustrie kommt (auch in Teilen) bin ich gezwungen damit z. B. die Initiative Neue Soziale Marktwirtschaft (INSM) zu unterstützen . Weil ein Teil meines Geldes für die INSM verwendet wird . Wio ist da Freiwilligkeit , Auswahl , Wettbewerb und Transparenz ? Und jetzt kommen Sie mir nicht auch damit , dass man ja nichts kaufen muss wo Metall oder Eletronik drin steckt .". Annotations include a blue box labeled "(Relevance)" above the text, and a yellow box labeled "unklare Kategorie | nicht vorhandene Kategorie" with "neg" below it, pointing to the word "Bahnstreiks". Another yellow box labeled "neg" points to the word "überflüssig".

„zu + Adj.“

„zu“ wird jeweils mitannotiert, wenn die negative Konnotation sonst nicht eindeutig ist (Bsp.: „zu schnell“).

Personenschäden

Personenschäden sind als „Sonstige Unregelmäßigkeiten | neg“ zu annotieren.

Annotation

	(Relevance)		neg	
67	#####	" RT @PSAIKODINO : durchsage eben in der bahn wegen eines personenschadens : " gleich , wenn die gleisen ENDLICH gereinigt sind , kann die fahrt w ... "	Sonstige Unregelmäßigkeiten	

Weichenstörung

Weichenstörungen sind als „Sonstige Unregelmäßigkeiten | neg“ zu annotieren.

Personen im Gleis

„Personen im Gleis“ sind als „Sonstige Unregelmäßigkeiten | neg“ zu annotieren.

Zug fahren

Allgemeine Äußerungen über sind der Überkategorie „Zugfahrt“ zuzuordnen. Eine Unterkategorie muss nicht gewählt werden.

Beendete Bahnstreiks, etc.

Das Beenden von Tätigkeiten ist zunächst weder positiv noch negativ konnotiert. Die Polarität des Ausdrucks ist abhängig davon, ob „etwas Positives“ oder „etwas Negatives“ beendet wird. Bei der Annotation entsprechender Stellen wird der vollständige Ausdruck einschließlich der Verneinung („beendet“, „gestoppt“, etc.) vom Polaritätstag umschlossen.

Annotation

	(Relevance)		pos	pos	
82	#####	Einigung bei der Bahn : Lokführer beenden ihren Streik . Matthias #Platzeck und @bodoramelow sollen Lösung suchen . http://t.co/eyuNTmta1r	unklare Kategorie	nicht vorhandene Kategorie	

Sperrungen

Sperrungen sind als „Sonstige Unregelmäßigkeiten“ auszuzeichnen.

Umleitungen

Umleitungen sind der Kategorie „Streckennetz“ zuzuordnen.

Schienenersatzverkehr

Schienenersatzverkehr stellt eine „Sonstige Unregelmäßigkeit dar.

Entgleisungen

Entgleisungen sind der Kategorie „Sonstige Unregelmäßigkeiten“ zuzuordnen.

Funklöcher

Funklöcher beschreiben Lücken im Funknetz, das dazu dient, Lokführern die Kommunikation untereinander zu ermöglichen. Funklöcher sind daher nicht der Kategorie „Handyempfang und Telefonie“ sondern der Kategorie „Sicherheit“ zuzuordnen.

Schlichtung (eines Tarifstreits)

Schlichtungen sind bei der Annotation auszusparen, es sei denn, es wird mit zusätzlichen polaritätstragenden Ausdrücken auf sie eingegangen.