|  |  |
| --- | --- |
| **Glossar** | |
| **Begriff** | **Inhalt** |
| **Crowd Sourcing** | Der Begriff Crowdsourcing tauchte erstmals 2006 in einem Artikel von Jeff Howe im US-amerikanischen Magazin Wired auf. Ziel ist, die Masse der Internetnutzer dazu aufzurufen, sich aktiv an Arbeits- und Kreativprozessen von Unternehmen zu beteiligen. Mithilfe der Intelligenz der Masse und der Arbeitskraft von vielen gelingt es, Projekte zu bewältigen oder neue Lösungsmöglichkeiten zu gewinnen. Quellehttps://www.textbroker.de/crowdsourcing 2022 10 14 |
| **Disambiguierung** | Die Auflösung von Mehrdeutigkeiten bei der Verarbeitung natürlicher [Sprache](https://www.spektrum.de/lexikon/neurowissenschaft/sprache/12159). Im Gegensatz zu den meisten in der Mathematik und der Informatik angewandten formalen Sprachen weisen natürliche Sprachen vielfach Mehrdeutigkeiten oder Unbestimmtheiten auf. Diese können sich auf einzelne Wörter (z.B. "Mann" im Sinne von "Ehemann", "Mann" als Gegensatz zu "Frau", "Mann" als Gegensatz zu "Junge"), Satzteile oder ganze Sätze ("Der Mann sah das Kind mit dem Fernglas"), oder auch über Satzgrenzen hinweg (Mehrdeutigkeiten bei Verwendung von Relativpronomina wie "dieser", "der") erstrecken. Um die Bedeutung einer sprachlichen Äußerung erschließen zu können, ist die Disambiguierung deshalb eine der wichtigsten kognitiven Leistungen beim Verstehen natürlicher Sprache und zugleich eines der größten Probleme in der maschinellen Sprachverarbeitung. Quelle: <https://www.spektrum.de/lexikon/neurowissenschaft/disambiguierung/2888> 2022 10 14 |
| **Linked Open Data** | **Linked Open Data** (LOD) bezeichnet frei verfügbare Daten, die im Internet über einen Uniform Resource Identifier (URI) eindeutig identifiziert und adressiert sind und ebenfalls per URI auf andere Daten verweisen. Die miteinander verknüpften Daten ergeben ein weltweites Netz, das “[»Web of Data«](http://linkeddatabook.com/editions/1.0/)“. Um eine möglichst einfache automatische Nutzung der Daten unabhängig von ihrem konkreten Ursprung zu gewährleisten, sollten diese offenen Daten gemäß internationaler [»W3C Standards«](http://www.w3.org/) und nach den “Linked Open Data” (LOD) Richtlinien im Internet vernetzt werden. Quelle: https://openall.info/formate/linked-opendata/ 2022 10 14 |
| **manuelle Annotation** | This invovles getting human annotators to manually perform POS annotation. It is a particularly laborious process and because of that, manual annotation is very rarely performed in today’s day and age.  For this process to be carried out well, more than one annotator is required and attention must be paid to annotator agreement. This is usually facilitated by the use of a specialized annotation software which does not assign POS tags but detects any inconsistencies between annotators. When the software detects that there is a word (a token) that has been assigned different tags by different annotators, the annotators would need to find a resolution on how to annotate the word or they may even decide to expand the tagset to accommodate the new situation.  In current times, manual annotation is mostly used to annotate a small corpus that will be used as training data for the development of a new automatic POS tagger.  Performing manual annotation on modern multi-billion-word corpora isn’t really feasible, which is why automatic tagging is used instead. Quelle:<https://www.engati.com/glossary/part-of-speech-tagging> 2022 10 14 |
| **Open Annotation Modell** | The Open Annotation Core Data Model specifies an interoperable framework for creating associations between related resources, annotations, using a methodology that conforms to the Architecture of the World Wide Web. Open Annotations can easily be shared between platforms, with sufficient richness of expression to satisfy complex requirements while remaining simple enough to also allow for the most common use cases, such as attaching a piece of text to a single web resource.  An Annotation is considered to be a set of connected resources, typically including a body and target, where the body is somehow about the target. The full model supports additional functionality, enabling semantic annotations, embedding content, selecting segments of resources, choosing the appropriate representation of a resource and providing styling hints for consuming clients. Quelle: <http://www.openannotation.org/spec/core/> 2022 10 14 |
| **Part of Speach (POS) Tagging** | If we talk about Part-of-Speech (PoS) tagging, then it may be defined as the process of assigning one of the parts of speech to the given word. It is generally called POS tagging. In simple words, we can say that POS tagging is a task of labeling each word in a sentence with its appropriate part of speech. We already know that parts of speech include nouns, verbs, adverbs, adjectives, pronouns, conjunction, and their sub-categories.  POS tagging is often also known as annotation or POS annotation. The annotation can be performed manually or automatically. <https://www.engati.com/glossary/part-of-speech-tagging> 2022 10 14 |
| **prozedurales Makrup** | Hierbei handelt es sich um ein darstellendes Datenbeschreibungsformat, welches für die korrekte Ausgabe eines digitalen Dokuments am Ausgabegerät, wie zum Beispiel einem Bildschirm oder Drucker, verantwortlich ist. Im Gegensatz zu deskriptiven können prozedurale Auszeichnungssprachen mit gewissen prozeduralen Elementen angereichert werden. Unter prozeduralen Elementen sind Aufgaben bzw. Tasks zu verstehen, die im Rahmen des digitalen Dokuments eine bestimmte Funktion ausüben. Quelle <https://www.seobility.net/de/wiki/Markup> 2022 10 14 |
| **Semantic Web** | Semantic Web bezeichnet eine Technologie, bei der die Daten einer konventionellen Webseite um [**strukturierte Daten**](https://www.advidera.com/glossar/structured-data-markup/) angereichert werden, um die Bedeutung der Information für Maschinen leichter auswertbar zu machen. Auf diese Weise sollen Maschinen die Bedeutung von Inhalten verstehen können. Das Semantic Web wird auch als Web 3.0 bezeichnet. Quelle: <https://www.advidera.com/glossar/semantic-web/> 2022 10 14 |
| **semantische Annotation** | Den Prozess des Hinzufügens von semantischen Meta-Daten zu Dokumenten, die den Inhalt eines Dokuments in maschinen-verarbeitbarer Form beschreiben, nennt man Semantische Annotation (engl. Semantic Annotation). Dabei kann es sich bei den Dokumenten um Informationsquellen wie z.B. Web-Seiten, Bilder, Audio oder Video Dateien handeln. Quelle: <https://www.merlin.uzh.ch/contributionDocument/download/2165> |
| **Textsorten** | Der Begriff ‚Textsorte‘ ist in der Forschung nicht einheitlich definiert. Drei mögliche Definitionen sollen im Folgenden näher betrachtet  als werden. Zunächst soll Brinkers Worterläuterung genannt werden:  „Textsorten sollen […] komplexe Muster sprachlicher Kommunikation verstanden werden, die innerhalb einer Sprachgemeinschaft im Laufe der historisch-gesellschaftlichen Entwicklung aufgrund kommunikativer Bedürfnisse entstanden sind.“  Diese Definition rückt die kommunikative Funktion von Textsorten in den Mittelpunkt, womit ihre fundamentale Bedeutung für die kommunikative Praxis deutlich wird.  Als Bedingung für die Zuordnung zu einer Textsorte nennt Brinker außerdem das Vorhandensein komplexer Muster. Auf diesen Aspekt stützt sich vorrangig die Definition von Linke: Textsorten sind „Gruppen von Texten, die sich durch bestimmte Bündel von Merkmalen auszeichnen“. Auch Ulla Fix bespricht den Begriff ‚Textsorte‘ und kommt bezüglich der Worterläuterung zu einem ähnlichen Ergebnis: „Unter einer Textsorte ist […] eine Klasse von Texten zu verstehen, die einem gemeinsamen Textmuster folgen“. Als Textmuster ist hier laut Fix „eine Anweisung mit prototypischen Elementen und Freiräumen“ zu verstehen, die kennzeichnend für eine bestimmte Textsorte ist.  Alle drei Begriffserklärungen weisen den Aspekt der Muster auf, nach denen ein Text kategorisiert werden kann. Quelle: <https://www.grin.com/document/118538> 2022 10 14 |

Drei Favoriten:

Dandelion <https://dandelion.eu/semantic-text/entity-extraction-demo/>

Theodor Fontane: Notizbücher https://fontane-nb.dariah.eu/index.html

Edition meiner Wahl: Religionsgeschichte der Gandenpa: Abschnitt über die Abtfolge des Klosters Ganden

<https://brugnara.at/TibetanTEI-main/content/TEI%20Ganden%20History_BP_excersise%20WS2021_2021-10-03_Albert_1.xml>