



Inhalt / Content

[Import in CATMA und Stanford NER-getaggte Texten](#)

[Stanford NER Texte zum Import vorbereiten](#)

[Import in CATMA](#)

[Analysiere deine NER-Ergebnisse direkt in CATMA](#)

Ich gebe zu, dass wahrscheinlich 10% meiner Liebe zu dem Tool CATMA ([Computer Aided Textual Markup and Analysis](#)) nur daher stammt, dass es von meinen Kollegen in Hamburg entwickelt wird, die niemals müde zu werden scheinen, wenn Nutzer wie ich immer neue Funktionalitäten haben wollen. Aber so bleiben immer noch 90%, die daher rühren, dass dieses Programm inzwischen eine große Vielfalt an Nutzungsmöglichkeiten bietet, sodass ich immer wieder dazu zurückkehre. Selbst, wenn ich mir mal vornehme, einen anderen Weg zu gehen, kehre ich doch immer wieder um. So geschah es also auch als ich nach Weiterverwendungsmöglichkeiten für meine [NER](#) steht für Named Entity Recognition, eine Methode bei der eindeutig benannte Größen (Entitäten) automatisch vom Computer erkannt und im Text markiert werden. Typische Beispiele für NER-Kategorien sind Personen, Orte und Organisationen. ([Named Entity Recognition](#) Auch als NER bezeichnet, ist eine Methode bei der eindeutig benannte Größen (Entitäten) automatisch vom Computer erkannt und im Text markiert werden. Typische Beispiele für NER-Kategorien sind Personen, Orte und Organisationen. - mehr darüber [hier](#)) Ergebnisse suchte. Als jetzt im neuen CATMA-Release auch noch eine Funktion zur xml Erkennung freigeschaltet wurde, wurde das Ganze sogar noch einmal einfacher. Darum zeige ich dir heute, wie man CATMA und Stanford [NER](#) steht für Named Entity Recognition, eine Methode bei der eindeutig benannte Größen (Entitäten) automatisch vom Computer erkannt und im Text markiert werden. Typische Beispiele für NER-Kategorien sind Personen, Orte und Organisationen. zusammen nutzen kann.

Import in CATMA und Stanford **NER**-getaggte Texten

Zur [Named Entity Recognition](#) Auch als NER bezeichnet, ist eine Methode bei der eindeutig benannte Größen (Entitäten) automatisch vom Computer erkannt und im Text markiert werden. Typische Beispiele für NER-Kategorien sind Personen, Orte und Organisationen. verwende ich das Stanford [NER](#) steht für Named Entity Recognition, eine Methode bei der eindeutig benannte Größen (Entitäten) automatisch vom Computer erkannt und im Text



markiert werden. Typische Beispiele für NER-Kategorien sind Personen, Orte und Organisationen. Tool; eine Anleitung zur Installation findest du [hier](#). CATMA ist ein webbasiertes Tool, dass du nicht lokal installieren musst. Folge einfach dieser [Webadresse](#) und logge dich mit einem Google bzw. Open ID Account ein und du kannst loslegen.

Stanford **NER** steht für Named Entity Recognition, eine Methode bei der eindeutig benannte Größen (Entitäten) automatisch vom Computer erkannt und im Text markiert werden. Typische Beispiele für NER-Kategorien sind Personen, Orte und Organisationen. Texte zum Import vorbereiten

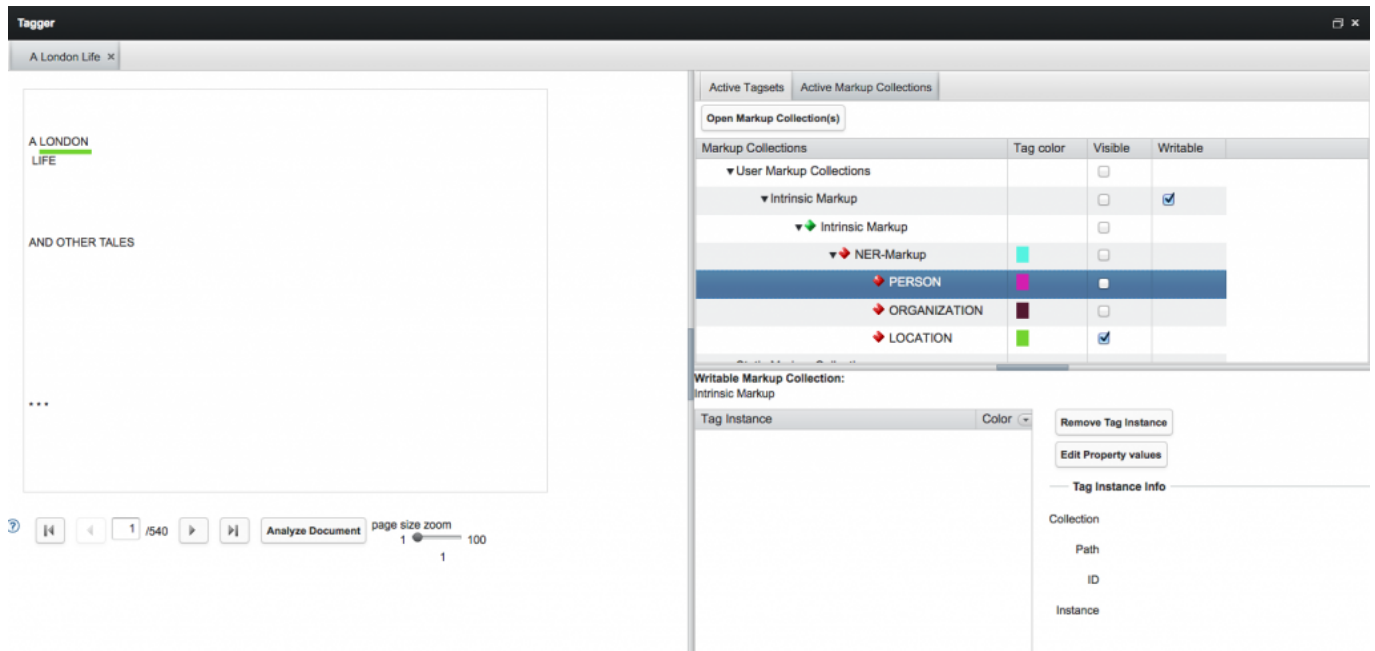
Hast du einen Text mit dem Stanford **NER** steht für Named Entity Recognition, eine Methode bei der eindeutig benannte Größen (Entitäten) automatisch vom Computer erkannt und im Text markiert werden. Typische Beispiele für NER-Kategorien sind Personen, Orte und Organisationen. Tool getaggt, so speichere ihn als getaggte Datei mit einem leicht wiederzufindenden Namen ab. Leider erhältst du damit zuerst einmal nur eine Textdatei mit einigen xml tags. Diese musst du in eine richtige xml Datei umwandeln, indem du dem Text einen entsprechenden Rahmen gibst. Dazu musst du nur einen `< head >` Tag vor Beginn des Textes einfügen und diesen am Ende wieder schließen. Zum Beispiel könntest du in die erste Zeile einfügen `< NER-Markup >` und in die letzte `< / NER-Markup >` (ohne Leerzeichen). Dann speichere die Datei so ab, dass sie eine .xml erweiterung hat. Das geht zum Beispiel mit den Programmen Wordpad oder TextWrangler, indem man einfach das txt zu einem xml machst.

Import in CATMA

Jetzt kannst du zurück zu CATMA gehen und diese xml Datei im Schritt für Schritt Upload hochladen. CATMA wird jetzt automatisch die Datei als xml erkennen, wenn alles korrekt ausgeführt wurde. Leider wird es Probleme geben, wenn die Anführungszeichen wie folgt verwendet wurden: `>...<` (das kommt leider in literarischen Texten häufig vor. Da xml diese Zeichen nicht als Text, sondern als Befehle erkennt, musst du sie suchen und mit „...“ ersetzen. Auch dieses Zeichen `&` bereitet leider Probleme und sollte durch „und“ ersetzt werden. Wenn der Upload erfolgreich war, wirst du nun eine Datei haben, die in einem



„Intrinsic Markup Collection“ genannten Bereich all deine NER steht für Named Entity Recognition, eine Methode bei der eindeutig benannte Größen (Entitäten) automatisch vom Computer erkannt und im Text markiert werden. Typische Beispiele für NER-Kategorien sind Personen, Orte und Organisationen. Tags gespeichert hat und bei Abruf als farbige Markierungen im Text anzeigt:



Analysiere deine NER-Ergebnisse direkt in CATMA

Als nächstes kannst du zu „analyze document“ gehen, um deine Daten weiter zu verarbeiten. Es wird sich ein neues Fenster öffnen, in dem du in einer Commandline reguläre Ausdrücke eingeben kannst, um deinen Text zu durchsuchen. Alternativ kannst du auch dem grafischen Klick-System folgen, dass dich über den Button „build query“ durch ein Such-System leiten wird, mit dem du Tags, Worte oder Kollokationen suchen kannst.

Wenn du in die Commandline z.B. tag="NER-Markup/LOCATION%" eingibst, so wirst du eine Liste aller Ortsnamen im Text bekommen. Wähle alle angezeigten Ergebnisse aus und du bekommst auf der rechten Seite des Fensters eine Keyword in Context Anzeige, die dir die Ortsnamen mit jeweils 5 Wörtern davor und danach anzeigt. Unterhalb dieses Fensters findest du kleine Symbole, die dir den Export als xls oder csv Tabellendokument erlauben. Die heruntergeladene Datei kannst du in Excel ansehen und weiter bearbeiten. Oder du



kannst die Tabelle weiterhin über CATMA anschauen und dir die einzelnen Ortsnamen im Text anzeigen lassen, indem du doppelt darauf klickst. Außerdem kannst du dir Visualisierungen deiner Daten erstellen lassen, indem du auf die Symbole für „distribution graph“ oder „treemap“ klickst.

The screenshot shows the CATMA interface. The 'Analyzer' tab is active, displaying a table of phrases and their frequencies. The 'Visualizer' tab is also visible, showing a line graph of occurrences.

Phrase	Frequency	Visible in Kwic
▶ Boulogne	1	<input checked="" type="checkbox"/>
▶ Japan	1	<input checked="" type="checkbox"/>
▶ Charlestown	1	<input checked="" type="checkbox"/>
▶ Marlborough	1	<input checked="" type="checkbox"/>
▶ Tishy	1	<input checked="" type="checkbox"/>
▶ Ostend	1	<input checked="" type="checkbox"/>
▶ Eton	1	<input checked="" type="checkbox"/>
▶ NEW YORK	3	<input checked="" type="checkbox"/>
▶ American	1	<input checked="" type="checkbox"/>

Total count: 74 Total frequency: 406

The Visualizer tab shows a line graph titled 'Distribution analysis for A London Life'. The x-axis represents the position in the text (0 to 100), and the y-axis represents the number of occurrences (0 to 80). The graph shows a peak of approximately 75 occurrences around position 75.

Natürlich kannst du auch - so wie ich es getan habe - deine Rohdaten im Tabellendokument für die Weiterverarbeitung aufbereiten. Hier kannst du z.B. alle vom NER steht für Named Entity Recognition, eine Methode bei der eindeutig benannte Größen (Entitäten) automatisch vom Computer erkannt und im Text markiert werden. Typische Beispiele für NER-Kategorien sind Personen, Orte und Organisationen. Tool falsch erkannten Entitäten löschen (in der deutschen Version werden das recht viele sein, im Englischen hingegen eher Ausnahmefälle). Schließlich kannst du diese Daten nutzen, um sie z.B. in ein Geolokalisationstool zu übertragen. Wie du das machen kannst, werde ich demnächst auf diesem Blog Blog ist kurz für Web-Log und steht für ein online Publikationsformat. Man kann sowohl der als auch das Blog sagen. Es gibt Blogs aller Sparten, von Linklisten über Tagebuchartige Formate bis hin zu wissenschaftlichen Blogs. Die Veröffentlichung kann schnell und unkompliziert erfolgen oder redaktionellen Standards entsprechen. In den Geisteswissenschaften etablieren sich Blogs zunehmend als Alternative zur langwierigeren wissenschaftlichen Publikation. Lebe lieber literarisch ist ein populärwissenschaftlicher Literaturblog. Kurze Podcast-Folge zum Blog-Begriff:

Schumacher, Mareike auf <https://lebelieberliterarisch.de>
Lizenz: cc-by 4.0 (Creative Commons Namensnennung)



[https://hnp9zs.podcaster.de/download/Podcast_Blog\(1\).mp3](https://hnp9zs.podcaster.de/download/Podcast_Blog(1).mp3) erklären.

Mehr literarisches Leben:

Wöchentliche Blog-News

Mit deinem Lebe-lieber-literarisch-Abonnement erhältst du jede Woche neue Artikel direkt in dein Postfach. Kein Buch-Tipp, keine Kolumne, keine digitale Fallstudie und kein Blog-Tipp wird dir je wieder entgehen!

Invalid email address

Abonnieren

Take Action Now!

Probier es einfach mal aus, du kannst dich jederzeit wieder abmelden.