

neonion – Kollaboratives, semantisches Annotieren von Dokumenten als Mehrwert für das Forschen in den Geisteswissenschaften und der Informatik

Claudia Müller-Birn¹, Florian Schmaltz², Tina Klüwer¹, Juliane Stiller²

¹Freie Universität Berlin, Institut für Informatik, Human-Centered Computing

²Max-Planck-Institut fuer Wissenschaftsgeschichte, Forschungsprogramm Geschichte der Max-Planck-Gesellschaft

neonion ist eine Webanwendung, die es Benutzer_innen erlaubt, Wörter und Textteile in Dokumenten oder Dokumente selbst semantisch zu annotieren. Das Ziel bei der Softwareentwicklung ist es dabei insbesondere, den Prozess der Erstellung der semantischen Annotationen so intuitiv zu gestalten, dass die Komplexität des zugrundeliegenden Datenmodells vor den Nutzer_innen weitestgehend verborgen werden kann. Mit neonion sollen somit vor allem Wissenschaftler_innen angesprochen werden, die nicht mit semantischen Technologien wie RDF, Ontologien oder SPARQL vertraut sind, aber trotzdem von den Vorteilen dieser Technologien profitieren wollen. neonion ermöglicht es Dokumente gemeinschaftlich zu annotieren, d.h. als Mensch-Mensch oder Mensch-Maschine-Kollaboration. Annotationen können privat sein, in einer Gruppe gemeinsam erstellt oder sogar öffentlich sichtbar gemacht werden. So kann das bei der Annotation erstellte Wissen auch Teil des *Webs of Data* werden.

Derzeit wird neonion gemeinsam mit den Nutzer_innen aus dem Bereich der Geschichtswissenschaften entwickelt. In einem Pilotprojekt für das Forschungsprogramm „Geschichte der Max-Planck-Gesellschaft, 1948-2002“, welches am Max-Planck-Institut für Wissenschaftsgeschichte in Berlin angesiedelt ist, finden regelmäßige Treffen mit den potentiellen Nutzer_innen der Software statt.

Parallel zur Entwicklung der Software wird eine Studie durchgeführt, in der Wissenschaftler_innen interviewt werden und ihre Benutzung von neonion experimentell beobachtet wird. Erste Ergebnisse der Interviews zeigen, dass die dem Annotationsprozess zugrundeliegenden mentalen Modelle sehr unterschiedlich ausfallen, d.h. dass die Befragten ein individuelles Verständnis zum Konzept der Annotation besitzen. So wurde der verwendete Annotationsinhalt (z.B. Kategorie, Freitext, Tag) durch das Ziel der Auswertung (z.B. Klassifizierung, Übersetzung) beeinflusst. Die Befragten sehen semantische Annotationen dann als nützlich an, wenn Nutzer_innen ihre Annotationen gemeinschaftlich erstellen und verwenden sowie diese maschinell weiter verarbeiten können. In den Interviews hat sich gezeigt, dass Wissenschaftler_innen um die Nützlichkeit des semantischen Ansatzes wissen, aber sehr unsicher bei der eigentlichen Anwendung sind. Die fehlende Erfahrung in diesem Bereich führt letztlich wieder zur Verwendung von Werkzeugen wie beispielsweise Textverarbeitungsprogrammen oder PDF-Software, obwohl die Nachteile, beispielsweise bezüglich der Weiterverwendung der Annotationen, bekannt sind. Erste Usability-Studien mit neonion haben gezeigt, dass Nutzer_innen sich sehr gut in der Software zurechtfinden und Annotationen schnell und intuitiv durchführen können. Durch die enge Zusammenarbeit von Anwender_innen und Entwickler_innen können Möglichkeiten zur Verbesserung des Interaktionsdesigns frühzeitig im Entwicklungsprozess erkannt und adressiert werden.

Wie bereits dargelegt, ist ein Ziel von neonion die kollaborative Annotation von Dokumenten nicht nur für menschliche Interaktionen, sondern auch Mensch-Maschine-Interaktionen zu unterstützen. Dazu werden zwei Ebenen der manuellen semantischen Annotation unterschieden, die im Folgenden kurz anhand der Personenannotation erläutert werden:

(1) Auf der *Konzeptebene* annotiert der Nutzer_innen ausgewählte Wörter oder Textteile mit vorher festgelegten Begriffen (sog. Konzepte). Zum Beispiel sollen alle in historischen Dokumenten genannten Personen auf ihr gemeinsames Vorkommen (z.B. bezüglich Zeit und Ort) hin untersucht werden. Daher werden von den Nutzer_innen Namen, z.B. „Feodor Lynen“ mit dem Konzept „Person“ verbunden und damit im Dokument eine Annotation erzeugt. Die resultierende RDF-basierte Beschreibung der Instanz enthält die ausgewählten Namen vom Typ Person.

(2) Benutzer_innen können nicht nur die ausgewählten Namen auf ein Konzept in einer Ontologie beziehen, sondern diese Instanzen auch mit einer direkt identifizierbaren Ressource im Web verknüpfen. Wir bezeichnen dies als Annotation auf der *Referenzebene*. So könnte die lokal annotierte Person „Feodor Lynen“ mit dem Wikidata-Eintrag Q44597 zu Feodor Lynen referenziert werden.

Die Annotationen können nun um weiteres Wissen angereichert werden. Derzeit nutzt neonion Wikidata und dies erlaubt Nutzer_innen ebenfalls auf Daten aus der VIAF (Virtual International Authority File) oder der GND (Gemeinsame Normdatei) zuzugreifen. Indem also weitere Informationen aus der Linked Open Data Cloud einbezogen werden, kann auf Basis einer einfachen Annotation ein komplexes Wissensnetzwerk erzeugt werden. Solche Wissensnetzwerke können dann auch zur weiteren Analyse visualisiert werden.

Diese manuellen Annotationsebenen werden durch automatische Annotatoren, derzeit durch einen Named Entity Recognizer, erweitert. Während der manuellen Annotation von Dokumenten werden den Nutzer_innen Vorschläge des automatischen Annotators präsentiert. Die Nutzer_innen können diese annehmen, ablehnen oder editieren. Über diesen Interaktionsprozess werden zukünftig angebotene Empfehlungen schrittweise verbessert. Das entwickelte Mensch-Maschine-Interaktionskonzept wird derzeit evaluiert.

neonion soll als ein Beispiel für eine gelungene Zusammenarbeit zwischen den Geisteswissenschaften und der Informatik dienen, da die Forschung in beiden Fachdisziplinen mit diesem Projekt vorangetrieben werden kann.