

Workshop: APIs und andere Werkzeuge für die Vernetzung zwischen Programmen und Projekten

Zeit: Donnerstag, 02.09.2021

Die Herstellung von Interoperabilität ist ein wesentlicher Anspruch innerhalb der Planung und Umsetzung von Forschungsprojekten. Gerade mit der Spezialisierung einzelner Vorhaben, welche meist ein Anwachsen von Datenbeständen und -ergebnissen zur Folge hat, stellt deren mögliche Anbindung der Ressourcen an andere Projekte eine große Herausforderung dar. Um dieser Anbindung nachzukommen, stehen eine Reihe von Werkzeugen zur Verfügung. So kann diese Anbindung auf Projektebene durch die Schaffung gut dokumentierter APIs (application programming interface) erfolgen, ebenso ist eine Orientierung hin zu übergreifenden Datenmodellen (CIDOC-CRM, CMDI etc.) oder die Verwendung von Normdateien (GND, QID etc.) feststellbar. Auch in der Variationslinguistik, die auf die Einbindung sprachlicher wie außersprachlicher Ressourcen angewiesen ist, spielen Schnittstellen eine gewichtige Rolle.

Aus diesem Grund beschäftigt sich der **Workshop „APIs und andere Werkzeuge für die Vernetzung zwischen Programmen und Projekten“ (02.09.2021)** mit dem Thema Schnittstellen im Feld variationslinguistischer Forschungsprojekte. Der Begriff der Schnittstellen soll dabei breit angelegt sein, d.h. das betrifft die Anbindung von nicht-sprachlichen Ressourcen ebenso wie Erfahrungen aus dem Bereich der Schnittstellen hin zum Menschen (UI). Ebenso bietet der Workshop Raum, Probleme wie das der informationstechnischen Fixierung der konzeptionellen/inhaltlichen Seite des Sprachsystems, zu diskutieren.

Als kommunikatives Format stehen jedem Projekt 20 Minuten für den Vortrag selbst und 40 Minuten für die Diskussion zur Verfügung. Damit soll die Möglichkeit geboten sein, im Rahmen des Workshops Offenes und Problematisches im gemeinsamen Austausch zu diskutieren. In diesem Sinne dürfen und sollen ebensolche Aspekte gerne in den Fokus der Vorträge gestellt werden, ebenso wie solche, die sich in Entwicklung befinden.

Das Orga-Team

Ludwig Maximilian Breuer (ACDH-CH/DiÖ, Wien)

Markus Jochim (IPS, Ludwig-Maximilians-Universität München)

Markus Kunzmann (Universität Wien)

Nicole Palliwoda (DMW, Siegen)

Markus Pluschkovits (DiÖ, Wien)

Manuel Raaf (IT/DH Bayerische Akademie der Wissenschaften, München)

Programm

Stand 23-08-2021

Donnerstag, 02.09.2021	
09:00-09:15	Begrüßung
09:15-10:15	Standardisierung, Interoperabilität - und FAIRness? ISO/TEI für Transkriptionen gesprochener Sprache Hanna Hedeland (Mannheim)
10:15-11:15	"Du hast da 'ne Schnittstelle" - Mit Werkzeugen zur Datenanalyse neue Zugänge schaffen Timm Lehmborg (Hamburg)
11:15-12:15	User Interface Design und APIs für Dialektwörterbücher: Wie gestaltet man moderne Web-Schnittstellen für Mensch und Maschine? Manuel Raaf (München)
12:15-14:00	Mittagspause
14:00-15:00	Das Trierer Wörterbuchnetz - technische Infrastruktur und offene Schnittstellen Thomas Burch (Trier)
15:00-16:00	Auf dem Weg zur offenen API: Überlegungen aus und für den SFB DiÖ Ludwig Maximilian Breuer/Markus Pluschkovits (Wien)
16:00-17:00	Schnittstellen und UI im Projekt VerbaAlpina Florian Zacherl & David Englmeier (München)
17:00-17:15	Verabschiedung

Abstracts

Standardisierung, Interoperabilität - und FAIRness? ISO/TEI für Transkriptionen gesprochener Sprache

Hanna Hedeland (Mannheim)

Der Beitrag stellt zunächst den TEI-basierten ISO-Standard 2462:2016 "Transcription of spoken language" vor, der in verschiedenen Kontexten als Pivotformat für Werkzeugformate in den Bereichen Aufbereitung, Auswertung und Archivierung digitaler gesprochensprachlicher Daten eingesetzt wird. Da der Standard bestimmte generische Aspekte von Transkriptionen einheitlich modelliert (aber Aspekte, die von bestimmten Theorien oder Forschungsfragen abhängen, nicht einschränkt), bedeutet bereits die Konvertierung in dieses Format in Ansätzen eine FAIRification der Daten, da die Semantik im tokenisierten TEI-Format häufig spezifischer und vor allem in einer maschinenlesbaren Form definiert wird. So weit so gut, aber nicht nur diese TEI-Variante steht noch vor großen Herausforderungen, wenn Konzepte wie FAIR-Prinzipien oder gar FAIR Digital Objects vollständig implementiert werden sollen. Im Beitrag selbst werden hierzu keine Antworten gegeben, sondern er wird hoffentlich als Anlass genommen, in einer offenen Diskussion diese und verwandten Themen zu behandeln, neben grundsätzlichen Fragen der Verständlichkeit von TEI-Daten für nachnutzende Menschen und Maschinen auch aktuelle Ansätze wie TEI+RDFa oder eben FAIR Digital Objects.

"Du hast da 'ne Schnittstelle" - Mit Werkzeugen zur Datenanalyse neue Zugänge schaffen

Timm Lehmberg (Hamburg)

Bei der Aufbereitung digitaler (zumeist textueller) Forschungsdaten in den Geisteswissenschaften sehen sich Anwenderinnen und Anwender oft subjektiv mit einer Vielzahl anspruchsvoller Vorgaben bzgl. der Standardkonformität sowie weiteren Qualitätsanforderungen konfrontiert. Nicht selten wird konstatiert, dass eigene Forschungsinteressen und Anforderungen an Datenstrukturen in den Hintergrund geraten oder gar eigene Analysen erschwert würden.

Der Beitrag versucht zu zeigen, dass das Gegenteil der Fall ist. Am Beispiel einiger neu geschaffener Analyseworkflows aus dem Akademieprojekt INEL ("Grammatiken, Korpora und Sprachtechnologie für indigene nordeurasische Sprachen") soll demonstriert werden, wie auf der Grundlage standardkonformer Daten und der somit entstandenen Schnittstellen unter Einsatz gängiger Werkzeuge der Datenanalyse sehr performante und maßgeschneiderte Zugänge zu Ressourcen geschaffen werden können.

User Interface Design und APIs für Dialektwörterbücher: Wie gestaltet man moderne Web-Schnittstellen für Mensch und Maschine?

Manuel Raaf (München)

Ansprechende und mit wenigen Klicks nutzbare Webseiten sind auch innerhalb der Geisteswissenschaften von fundamentaler Bedeutung, da Nutzer:innen durch unübersichtliche oder komplexe Darstellungen und/oder Funktionen abgeschreckt werden und die Online-Publikation nach wenigen Klicks bereits wieder verlassen. Dies gilt es jedoch dringlichst zu verhindern, damit die Forschungsergebnisse im Web genutzt und verbreitet werden und damit letztlich auch eine positive Außenwirkung der Institution, die das Projekt finanziert, erzielt wird. Auch die nicht-menschliche Abfragemöglichkeit der Datenbestände (d.h. mittels Programmierschnittstellen (APIs)) zur

eigenständigen Nachnutzung oder zur Herstellung der Interoperabilität liefert hierzu einen wichtigen und immer häufiger geforderten Beitrag.

Beides, der menschliche ebenso wie der maschinelle Zugang, ist zudem grundlegend zur Erfüllung der häufig geforderten, stark voranschreitenden FAIR-Prinzipien.

Der Vortrag behandelt daher zunächst die Frage, wie Benutzerschnittstellen am Beispiel eines Onlinewörterbuchportals gestaltet werden können, um Fachwissenschaftler:innen sowie interessierte Lai:innen gleichermaßen anzusprechen. Hierzu wird zunächst auf den Stand der Forschung zu User Experience und User Interface Design innerhalb der elektronischen Lexikographie im Speziellen sowie darüber hinaus auch im Allgemeinen eingegangen. Im Anschluss wird dargestellt, was davon im Sprachinformationssystem „Bayerns Dialekte Online“ bereits umgesetzt wurde. In dem System werden die Inhalte der drei Dialektwörterbücher „Bayerisches Wörterbuch“, „Dialektologisches Informationssystem von Bayerisch-Schwaben“ und „Fränkisches Wörterbuch“ vereint – sowohl optisch als auch funktional. Ferner wird kurz auf die Entscheidungsgrundlagen der Umsetzung eingegangen.

Darüber hinaus wird abschließend anhand eines Beispiels skizziert, wie eine einfache, aber dennoch leistungsfähige Programmierschnittstelle (API) implementiert werden kann, um hierüber sowohl der Maschine als auch dem diese programmierenden Menschen einen komfortablen und nachhaltigen Zugang zu den Daten zu ermöglichen.

Das Trierer Wörterbuchnetz – Technische Infrastruktur und offene Schnittstellen

Thomas Burch (Trier)

Wörterbuchverbünde lassen sich konstruieren, wenn Wörterbücher digitalisiert und so publiziert werden, dass die ihrer Form und Herkunft nach jeweils einander entsprechenden Stichwörter verknüpft werden, um eine schnelle, bequeme und umfassende Nutzung derartiger Beziehungen zu ermöglichen. Entscheidend ist dabei die unmittelbare Vernetzung; sie ermöglicht das Aufsuchen eines Stichwortes über alle Wörterbücher hinweg, unabhängig vom jeweils gewählten Einstieg, d.h. unabhängig davon, in welchem Wörterbuch eine Nachschlagehandlung beginnt.

Entscheidend für die Umsetzung solcher Zugriffs- und Navigationsstrukturen sind die Festlegung gemeinsamer Standards und Lösungen für lexikographische Ressourcen und die Entwicklung von technischen Strategien, Tools und Lösungen für die Extraktion, Strukturierung und Verknüpfung der Ressourcen, um deren volles Potential im Sinne von Linked Open Data zu erschließen. Die Architektur des Trierer Wörterbuchnetzes verfolgt daher den Ansatz, generische Methoden für die Integration, Präsentation und Nutzung von Daten in digitalen lexikographischen Systemen zu entwickeln und diese Methoden in einem Forschungsportal bereit zu stellen, das auch für zukünftige Forschung genutzt werden kann. Als modular konzipierte Web-Plattform verfügt das Wörterbuchnetz über eine generische Modellierung für Backend und Frontend. Entsprechende Template-Mechanismen erlauben sowohl den Import standardisierter XML-codierter Wörterbuchdaten als auch deren konfigurierbare Online-Präsentation. Nach den Prinzipien der Linked Open Data können Wörterbuchartikel und auch andere Ressourcen und deren Bestandteile über offene Schnittstellen und persistente Adressen abgerufen werden. Auch die Einbindung weiterer Inhalte wie zum Beispiel die Vernetzung mit zitierten und bereits digitalisierten Quellen wird so bidirektional realisierbar. Eine „Meta-Datenbank“ ermöglicht den wörterbuchübergreifenden Zugriff auf den Gesamtdatenbestand.

Der Beitrag stellt das Wörterbuchnetz vor und geht auf seine modulare Architektur, die Erweiterbarkeit der zugrundeliegenden Infrastruktur und die verfügbaren offenen Schnittstellen ein.

Auf dem Weg zur offenen API: Überlegungen aus und für den SFB: DiÖ

Ludwig Maximilian Breuer, Markus Pluschkovits (Wien)

Die Bereitstellung einer API ist essentiell um den Mehrwert gesammelter (linguistischer) Forschungsdaten zu gewährleisten und den Zugriff auf diese zu ermöglichen. Das Design einer solchen API will daher wohlüberlegt sein. Zu diesen Überlegungen gehören nicht nur Fragen der Standardisierung und Auswahl der Bezugssysteme (z.B. Nament Entities oder andere Referenzsysteme wie Geo-Koordinaten oder semantische Taxonomien), sondern auch die Anforderungen und Ansprüche der Endnutzer/innen. Der Spezialforschungsbereich Deutsch in Österreich: Variation – Kontakt – Perzeption (FWF F60) befindet sich gerade inmitten der Konzeptionsphase einer API, weshalb diese Überlegungen im Zentrum unseres Beitrags stehen sollen. Die internen Ansprüche an eine API ergeben sich im Projekt beispielsweise bereits anhand enger Kooperationen mit dem Citizen Science Projekt lamDiÖ und deren "Wortgut"-Onlinewörterammlung sowie dem WBÖ (Wörterbuch der bairischen Mundarten in Österreich) und seiner Online-Plattform LIÖ (Lexikalisches Informationssystem Österreich), die gemeinsam entwickelt wird. Darüber hinaus sollen aber auch andere Projekte der deutschsprachigen Variationslinguistik befragt werden, welche Ansprüche sie an eine API zum Austausch von Daten bzw. zur Abbildung der SFB-Daten auf ihren Plattformen haben. Gleichzeitig dazu stellt sich allerdings auch die Frage, ob – in Einklang mit den verschiedenen Ansprüchen an eine API – nicht andere gangbare Alternativen existieren, welche die tatsächlichen Anforderungen von Linguist/inn/en und anderen variationslinguistischen Projekten nicht besser erfüllen können.

Schnittstellen und UI im Projekt VerbaAlpina

Florian Zacherl & David Englmeier (München)

Der Vortrag beschreibt die methodischen Ansätze, die im Projekt VerbaAlpina verwendet werden, um die dort gesammelten Sprachdaten nach außen sichtbar und nutzbar zu machen sowie mit anderen Web-Portalen zu vernetzen.

Im ersten Teil wird nach einer kurzen Einführung in die Modellierung der lexikalischen Daten in einer relationalen, normalisierten Datenbank speziell auf die sogenannte Datenzugriffsschicht eingegangen, die eine persistente Schnittstelle auf die darunter liegende variablere Datenstruktur liefert. Diese wird sowohl intern für die verschiedenen Webtools als auch extern, beispielsweise für den technischen Zugriff über eine API, verwendet. In diesem Zusammenhang geben wir kurze Einblicke in die verschiedenen Aspekte, die im Bezug auf die Nachnutzbarkeit der Daten sowohl im Kontext des wissenschaftlichen Zitierens als auch für die weitere Verwendung der Rohdaten relevant sind. Insbesondere werden die Themen Versionierung, Dokumentation und Anbindung an Partnerprojekte und bestehende Normdatenbanken aufgegriffen.

Der zweite Teil des Vortrags behandelt die gestalterischen sowie technischen Grundlagen des User Interfaces von VerbaAlpina. Wir erläutern am Beispiel des Lexikons, der Interaktiven Karte sowie des Crowdsourcing Tools die Anforderungen, die sich durch verschiedene Nutzergruppen sowie Datenbasis sowohl an die technische Realisierung als auch den Aufbau der Interfaces ergeben. Abschließend resümieren wir den Vortrag mit einer Zusammenfassung gesammelter Erfahrungen und Erkenntnisse.