

ELEXIS – Eine europäische Forschungsinfrastruktur für lexikographische Daten

Wissik, Tanja

Tanja.Wissik@oeaw.ac.at
Österreichische Akademie der Wissenschaften, Österreich

Krek, Simon

simon.krek@guest.arnes.si
Institut Josef Stefan, Slowenien

Jakubicek, Milos

milos.jakubicek@sketchengine.co.uk
Lexical Computing CZ s.r.o., Tschechien

Tiberius, Carole

carole.tiberius@inl.nl
Instituut voor Nederlandse Lexicologie, Niederlande

Navigli, Roberto

navigli@di.uniroma1.it
Università degli Studi di Roma La Sapienza, Italien

McCrae, John

john@mccr.ae
National University of Ireland, Galway Irland

Tasovac, Toma

ttasovac@humanistika.org
Centar za digitalne humanističke nauke, Serbien

Varadi, Tamas

varadi@nytud.hu
Magyar Tudományos Akadémia, Ungarn

Koeva, Svetla

svetla@dcl.bas.bg
Institute for Bulgarian Language, Bulgarien

Costa, Rute

rute.costa@fcsh.unl.pt
Universidade Nova de Lisboa, Portugal

Kernerman, Ilan

ilan@kictionaries.com
K Dictionaries Ltd., Israel

Monachini, Monica

monica.monachini@ilc.cnr.it
Consiglio Nazionale delle Ricerche, Italien

Trap-Jensen, Lars

ltj@dsl.dk
Det Danske Sprog- og Litteraturselskab, Dänemark

Pedersen, Bolette S.

vnb282@ku.dk
Kobenhavns Universitet, Dänemark

Hildenbrandt, Vera

hildenbr@uni-trier.de
Universität Trier, Deutschland

Kallas, Jelena

jelena.kallas@eki.ee
Eesti Keele Instituut, Estland

Porta-Zamorano, Jordi

porta@rae.es
Real Academia Española, Spanien

Die Bedeutung von Wörterbüchern, seien es einsprachige, zweisprachige oder mehrsprachige Wörterbücher, ist in der heutigen Informationsgesellschaft nicht zu unterschätzen. Sie geben nicht nur Auskunft über Wortbedeutungen und dazugehörige Übersetzungen, sondern sind selbst Bestandteil der Kulturgüter eines Landes und sie stellen bedeutende Ressourcen für Linked Open Data und Semantic-Web-Technologien dar. Obwohl in fast jedem Land Wörterbücher erstellt wurden und werden, seien es traditionelle Wörterbücher in gedruckter Form oder lexikographische Ressourcen in digitaler Form, waren die Bestrebungen nach Kooperation auf europäischer Ebene eher limitiert. Dies führte dazu, dass bis jetzt die Synergien zwischen traditioneller Lexikographie und maschineller Sprachverarbeitung nicht optimal genutzt werden konnten. Dies soll durch das Infrastrukturprojekt ELEXIS geändert werden.

Im Rahmen des ELEXIS Projekts soll eine Infrastruktur für lexikographische Daten entwickelt werden, die auf mehreren Ebenen ansetzt und den Bereich der traditionellen Lexikographie mit dem Bereich der maschinellen Sprachverarbeitung verknüpft: Zum einen soll Kooperation und Austausch zwischen unterschiedlichen Forschungsgruppen, aber auch zwischen Forschungsgruppen und Verlagshäusern gefördert werden. Zum anderen soll an gemeinsamen Standards gearbeitet werden, um den Austausch und die Wiederverwendbarkeit von lexikographischen Daten in den unterschiedlichsten Szenarien zu fördern. Die Infrastruktur soll den Zugang zu

Methoden, Tools und Daten ermöglichen und den bis jetzt nicht so weit verbreiteten Open-Access-Gedanken fördern.

Um diese Ziele umzusetzen, hat sich ein Konsortium von 17 Partnern gebildet. Unter den Partnern befinden sich Institutionen mit Expertise in Lexikographie, in maschineller Sprachverarbeitung, in Semantic-Web-Technologien und in den digitalen Geisteswissenschaften sowie nationale Sprachinstitutionen, Verlagshäuser und Partner mit Expertise im Bereich der Standardisierung und Normung. Nachfolgend die Liste der Partner (mit den Bezeichnungen in der jeweiligen Landessprache wie im Proposal angegeben): Institut Josef Sefan (Slowenien), Lexical Computing CZ s.r.o. (Tschechien), Instituut voor Nederlandse Lexicologie (Niederlande), Università degli Studi di Roma La Sapienza (Italien), National University of Ireland, Galway (Irland), Österreichische Akademie der Wissenschaften (Österreich), Centar za digitalne humanističke nauke (Serbien), Magyar Tudományos Akadémia (Ungarn), Institute for Bulgarian Language (Bulgarien), Universidade Nova de Lisboa (Portugal), K Dictionaries Ltd (Israel), Consiglio Nazionale delle Ricerche (Italy), Det Danske Sprog- og Litteraturselskab (Dänemark), Københavns Universitet (Dänemark), Universität Trier (Germany), Eesti Keele Instituut (Estland), Real Academia Española (Spanien).

Die ELEXIS Infrastruktur wird zum einen Tools und Services zur Erstellung, Verarbeitung und Retrodigitalisierung von lexikographischen Daten und zum anderen den Zugang zu bereits existierenden lexikographischen Daten anbieten. Damit die zukünftigen Nutzer das Potential der Infrastruktur vollends ausnutzen können, sind die Entwicklung von online Trainingsmaterialien sowie die Abhaltung von Trainingsworkshops geplant. Weiters sollen GastforscherInnen-Programme den Austausch zwischen den Forschungsgruppen aktiv fördern und für Forschungsvorhaben den Zugang zu Daten ermöglichen, die aus unterschiedlichen Gründen nicht Open Access zur Verfügung gestellt werden können.

ELEXIS wird eng mit den bereits existierenden Forschungsinfrastrukturen CLARIN und DARIAH zusammenarbeiten und auf den bereits vorhandenen Infrastrukturen aufbauen und zugleich diese beiden Infrastrukturen näher zusammenbringen.

In diesem Poster werden die Grundzüge des neuen europäischen Infrastrukturprojektes beschrieben sowie die Methoden und Maßnahmen, mit denen die oben genannten Ziele erreicht werden sollen, präsentiert. Weiters wird speziell auf den Nutzen und die Vorteile der ELEXIS Infrastruktur für die DH Community eingegangen.

Bibliographie

Schreibman, Susan / Ray Siemens, John Unsworth. ed. (2004): *A Companion to Digital Humanities*. Oxford: Blackwell.

Oldman, Dominic / Doerr, Martin / Gradmann, Stefan (2016): "Zen and the Art of Linked Data: New Strategies for a Semantic Web of Humanist Knowledge", in: Susan Schreibman/Ray Siemens/John Unsworth e. (2016): *A New Companion to Digital Humanities*, 2nd Edition. 2012. Oxford: Wiley-Blackwell

Navigli Roberto / Ponzeto Simone P. (2012): "BabelNet: The Automatic Construction, Evaluation and Application of a Wide-Coverage Multilingual Semantic Network", in: *Artificial Intelligence*, 193, Elsevier, 2012, pp. 217-250.

McCrae, John / Aguado-de-Cea, Guadalupe / Buitelaar, Paul / Cimiano, Philipp / Declerck, Thierry / Gómez-Pérez, Asunción / Gracia, Jorge / Hollink, Laura / Montel-Ponsoda, Elena / Spohr, Dennis / Wanner, Tobias (2012): "Interchanging lexical resources on the Semantic Web", in: *Language Resources and Evaluation*, 46(6), pp 701-709, (2012).