

D5.3 Overview of Online Tutorials and Instruction Manuals

Author(s): Toma Tasovac, Carole Tiberius, Claudia Bamberg, Andrea Bellandi, Thomas Burch, Rute Costa, Matej Ďurčo, Francesca Frontini, Julia Hennemann, Kris Heylen, Miloš Jakubíček, Fahad Khan, Anne Klee, Iztok Kosem, Vojtěch Kovář, Ondřej Matuška, John McCrae, Monica Monachini, Karlheinz Mörth, Tina Munda, Valeria Quochi, Andraž Repar, Christophe Roche, Ana Salgado, Henrike Sievers, Tamás Váradi, Sandra Weyand, Anna Woldrich, Susanne Zhanial

Date: 31 January 2022



H2020-INFRAIA-2016-2017
Grant Agreement No. 731015
ELEXIS - European Lexicographic Infrastructure

D5.3 Overview of Online Tutorials and Instruction Manuals

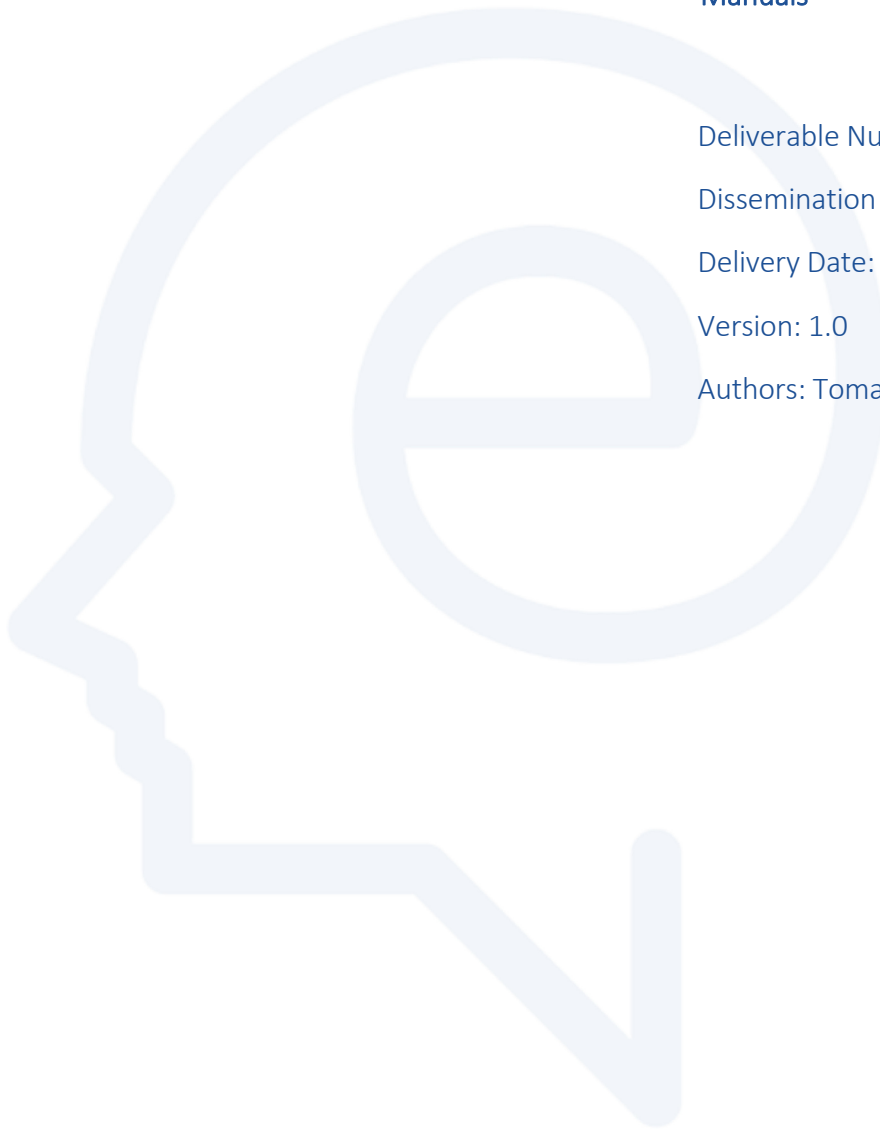
Deliverable Number: D5.3

Dissemination Level: Public

Delivery Date: 31 January 2022

Version: 1.0

Authors: Toma Tasovac, Carole Tiberius et al.



Project Acronym: ELEXIS
Project Full Title: European Lexicographic Infrastructure
Grant Agreement No.: 731015

Deliverable/Document Information

Project Acronym: ELEXIS
Project Full Title: European Lexicographic Infrastructure
Grant Agreement No.: 731015

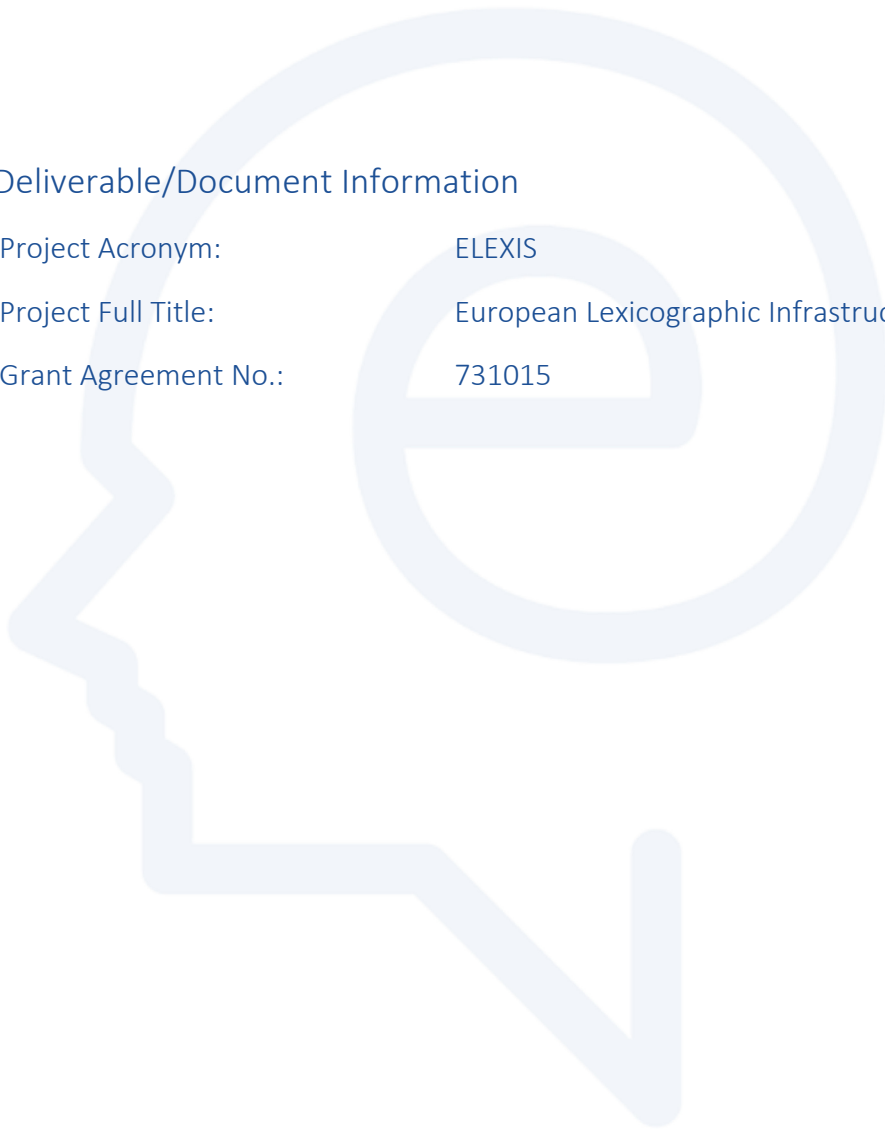


Table of Contents

INTRODUCTION	6
COURSES	8
INTRODUCTION TO DICTIONARIES	8
<i>Staging URL</i>	8
<i>Final URL</i>	8
<i>Learning Objectives</i>	8
<i>Learning Outcomes</i>	8
<i>TOC</i>	8
INTRODUCTION TO DICTIONARY USERS	9
<i>Staging URL</i>	9
<i>Final URL</i>	9
<i>Learning Objectives</i>	9
<i>Learning Outcomes</i>	9
<i>TOC</i>	10
INTRODUCTION TO CORPUS-BASED LEXICOGRAPHIC PRACTICE	10
<i>Staging URL</i>	11
<i>Final URL</i>	11
<i>Learning Objectives</i>	11
<i>Learning Outcomes</i>	11
<i>TOC</i>	11
CAPTURING, MODELING AND TRANSFORMING LEXICAL DATA: AN INTRODUCTION	12
<i>Staging URL</i>	12
<i>Final URL</i>	12
<i>Learning Objectives</i>	12
<i>Learning Outcomes</i>	12
<i>TOC</i>	12
LEX2: MASTERING ELEXIS CORPUS TOOLS FOR LEXICOGRAPHIC PURPOSES	13
<i>Staging URL</i>	14
<i>Final URL</i>	14
<i>Learning Objectives</i>	14
<i>Learning Outcomes</i>	14
<i>TOC</i>	14
LEXONOMY: MASTERING THE ELEXIS DICTIONARY WRITING SYSTEM.....	15
<i>Staging URL</i>	15
<i>Final URL</i>	15
<i>Learning Objectives</i>	15
<i>Learning Outcomes</i>	15
<i>TOC</i>	15
AUTOMATING THE PROCESS OF DICTIONARY CREATION	16
<i>Staging URL</i>	16
<i>Final URL</i>	16
<i>Learning Objectives</i>	16
<i>Learning Outcomes</i>	16
<i>TOC</i>	16
CLARIN TOOLS AND RESOURCES FOR LEXICOGRAPHIC WORK.....	17
<i>Staging URL</i>	17
<i>Final URL</i>	17
<i>Learning Objectives</i>	17

D5.3 Overview of Online Tutorials and Instruction Manuals

<i>Learning Outcomes</i>	17
<i>TOC</i>	18
STANDARDS FOR REPRESENTING LEXICOGRAPHIC DATA: AN OVERVIEW	18
<i>Staging URL</i>	18
<i>Final URL</i>	18
<i>Learning Objectives</i>	19
<i>Learning Outcomes</i>	19
MODELING DICTIONARIES IN TEI LEX-0	20
<i>Staging URL</i>	20
<i>Final URL</i>	20
<i>Learning Objectives</i>	20
<i>Learning Outcomes</i>	20
<i>TOC</i>	21
MODELING DICTIONARIES IN ONTOLEX-LEMON	21
<i>Staging URL</i>	22
<i>Final URL</i>	22
<i>Learning Objectives</i>	22
<i>Learning Outcomes</i>	22
<i>TOC</i>	22
LEX3: TRANSFORMING LEGACY DICTIONARIES USING ELEXIFIER	23
<i>Staging URL</i>	23
<i>Final URL</i>	23
<i>Learning Objectives</i>	23
<i>Learning Outcomes</i>	23
<i>TOC</i>	23
LEX3: PUBLISHING LEGACY DICTIONARIES WITH PUBLEX.....	23
<i>Staging URL</i>	24
<i>Final URL</i>	24
<i>Learning Objectives</i>	24
<i>TOC</i>	24
LEXICOGRAPHY IN THE AGE OF OPEN DATA	24
<i>Staging URL</i>	25
<i>Final URL</i>	25
<i>Learning Objectives</i>	25
<i>Learning Outcomes</i>	25
<i>TOC</i>	25
MASTERING LEX1: THE DICTIONARY MATRIX	26
<i>Staging URL</i>	26
<i>Final URL</i>	26
<i>Learning Objectives</i>	26
MASTERING OXYGEN XML EDITOR FOR DICTIONARY NERDS.....	26
<i>Staging URL</i>	26
<i>Final URL</i>	26
<i>Learning Objectives</i>	27
<i>Learning Outcomes</i>	27
<i>TOC</i>	27
EXTRACTING LEXICAL DATA: XPATH FOR DICTIONARY NERDS.....	27
<i>Staging URL</i>	28
<i>Final URL</i>	28
<i>Learning Objectives</i>	28
<i>Learning Outcomes</i>	28
<i>TOC</i>	28

TRANSFORMING LEXICAL DATA: XSLT FOR DICTIONARY NERDS	29
<i>Staging URL</i>	29
<i>Final URL</i>	29
<i>Learning Objectives</i>	29
<i>Learning Outcomes</i>	29
<i>TOC</i>	29
ELEXIS PATHFINDER TO COMPUTATIONAL LEXICOGRAPHY FOR DEVELOPERS AND COMPUTATIONAL LINGUISTS	30
<i>Staging URL</i>	30
<i>Final URL</i>	30
<i>Learning Objectives</i>	30
<i>Learning Outcomes</i>	30
<i>TOC</i>	30

Introduction

The ELEXIS Curriculum is an integrated set of training materials which contextualizes ELEXIS tools and services inside a broader, systematic pedagogic narrative. This means that the goal of the ELEXIS Curriculum is not simply to inform users about the functionalities of particular tools and services developed within the project, but to show how such tools and services are a) embedded in both lexicographic theory and practice; and b) representative of and contributing to the development of digital skills among lexicographers.

The scope and rationale of the curriculum are described in more detail in the Deliverable [D5.2 Guidelines for Producing ELEXIS Tutorials and Instruction Manuals](#). The goal of this deliverable, as stated in the project DOW, is to provide “a clear, structured overview of tutorials and instruction manuals developed within the project.”

For each course in the Curriculum we provide:

- Name of the course
- Level of difficulty
- Contributing institutions
- Authors
- Staging URL
- Final URL
- Learning objectives
- Learning outcomes
- Table of contents

The *staging URL* refers to the temporary server on which all the courses will be hosted until their final publication on [DARIAH-Campus](#). The *final URL* refers to the citable link that the courses will be published under on DARIAH-Campus. In both cases, the URL contains the unique ID of the course which is assigned by DARIAH-Campus.

DARIAH-Campus is a discovery framework and hosting platform for learning resources. It is maintained by DARIAH, which is a European Research Infrastructure Consortium (ERIC) and a European Strategy Forum on Research Infrastructures (ESFRI) Landmark. The legal status of DARIAH offers a stable framework for sustaining the ELEXIS learning resources beyond the end of ELEXIS as a funded H2020 project, while the DARIAH-Campus’ alignment with the European Open Science Cloud (EOSC) and the Social Sciences and Humanities (SSHOC) Marketplace will guarantee future discoverability and interoperability of the ELEXIS Curriculum.

D5.3 Overview of Online Tutorials and Instruction Manuals

At the time of the submission of this deliverable, all the courses are hosted on the staging server. The staging server is a temporary, mirror copy of the DARIAH-Campus platform and is being used for testing, harmonizing and mutually linking individual courses.

While some of the courses were production ready by the time of the submission of this deliverable, we decided to hold off on their publication until *all* the courses were completed. This we did for two reasons:

1. The ELEXIS Curriculum makes sense as an integrated, coherent set of learning objects only if all the courses are harmonized, mutually linked and made available to the public. Publishing a subset of courses would break the links and leave some users without the necessary contextual information (if a course has prerequisites) or the ability to pursue certain topics in more detail (if a course has follow-ups).
2. The editorial policies of DARIAH-Campus do not allow us to publish courses as drafts.

The curriculum harmonization work will be completed in the Spring 2022. The published version of the Curriculum will be presented at the [Elexis Showcase](#) event in Florence (June 7-8, 2022).

Depending on the time you're reading this deliverable, either the staging or the final URL should take you to the courses. The staging server will not be maintained indefinitely.

The screenshot shows the DARIAH-Campus website interface. At the top left is the DARIAH-CAMPUS logo. A navigation menu at the top right includes links for Home, Resources, Curricula, Topics, Sources, and Course registry. The main content area is titled "The ELEXIS Curriculum". On the left side, there is a sidebar with sections for "AUTHORS" (listing Toma Tasovac), "TOPICS" (listing LEXICOGRAPHY), and "CURRICULUM OVERVIEW" (listing various introductory and practical topics). The main text area describes the curriculum as an integrated set of training materials. Below this, there is a "Resources" section featuring a card for "Introduction to Dictionaries" with a brief description and a "Read more" link. Another card for "Introduction to Dictionary Users" is partially visible at the bottom.

The staging url of the ELEXIS Curriculum:

<https://elexis.humanistika.org/curriculum/the-elexis-curriculum>

The published url of the ELEXIS Curriculum:

<https://campus.dariah.eu/curriculum/the-elexis-curriculum>

Courses

Introduction to Dictionaries

Level: Introductory

Contributors: BCDH, FCSH-UNL

Authors: Toma Tasovac, Ana Salgado, Rute Costa

Staging URL

<https://elexis.humanistika.org/id/q0P4oqySKxkbcLhtXz49J>

Final URL

<https://campus.dariah.eu/id/q0P4oqySKxkbcLhtXz49J>

Learning Objectives

The goal of this course is to introduce a brief history of dictionaries as tools for the organization of knowledge about words and their meanings, and to analyze different ways of understanding and classifying the dictionary genre. In order to do so, the course will cover the constituent parts of a dictionary (macrostructure, microstructure and mediostructure) as well as different kinds of dictionary typologies, including those based on source and target languages (monolingual, bilingual, multilingual); types of language(s) and topic(s) covered (general language, encyclopedic, terminological); medium (print and electronic); semantic structure (onomasiological vs. semasiological dictionaries); and target audience (literate adults, language learners, language professionals). At the end of this course, students will have a fundamental understanding of the complexities of the dictionary genre as well as an appreciation of the role played by the medium in which the dictionary is compiled and consumed -- from clay tablets to computer screens.

Learning Outcomes

Upon completion of this course, students will be able to

- appreciate the complexity of the dictionary genre and its history
- analyze the structure of a dictionary entry
- categorize dictionaries based on their content and/or target audience
- understand the role played by the medium in which the dictionary is compiled and consumed

TOC

- **What is a dictionary?** *Complex object: a text, a model of language, a tool and a cultural artifact.*
- **How dictionaries came to be.** *Early history of lexicography (Middle East, Ancient China and Ancient Greece). Urra=hubullu (Summerian-Akkadian bilingual glossary): commercial and practical interests vs. literary traditions (Greece, China). Medieval glosses.*
- **Academic tradition in lexicography.** *Evidence-based lexicography (Johnson, sources) Standardization and nation-building (della Crusca, Académie française etc.) Technology, corpora.*

- **What makes a dictionary dictionary?**
 - Semantic structure: semasiology vs. onomasiology.
 - Building blocks:
 - macrostructure and the emergence of the alphabetic order
 - microstructure: understanding dictionary entries.
 - mediostructure: hypertext.
- **Dictionary typologies**
 - Target languages (monolingual, bilingual, multilingual)
 - Types of language(s) and topic(s) covered (general language, encyclopedic, terminological)
 - Standardization: prescriptive, descriptive
 - Target audience (literate adults, language learners, language professionals).
- **Carved in stone? The role of the medium** *From clay tablets to computer screens. How the material changed the way we compile, disseminate and use dictionaries.*
- **Instead of a conclusion: where next?** *This section will point in the direction of other courses in the ELEXIS Curriculum.*

Introduction to Dictionary Users

Level: Introductory

Contributors: JSI, OeAW, RILMTA

Authors: Iztok Kosem

Staging URL

<https://elexis.humanistika.org/id/xvxB5cdIPUECtRmZ0o563>

Final URL

<https://campus.dariah.eu/id/xvxB5cdIPUECtRmZ0o563>

Learning Objectives

The goal of this course is to introduce students to the important role played by dictionary usage research when developing and implementing new dictionaries. The course will address the question of how different types of target users (in terms of age, language proficiency and pre-existing skills) or different types of use (encoding, decoding, translation etc.) influence the scope of the dictionary, the lemma selection process or the very structure of a dictionary entry. At the end of this course, students will have a fundamental understanding of the ways in which user research (both commercially and academically) can contribute to the tailoring of lexicographic content. Going beyond the realm of user-centered lexicography, the course will also explore possible user contributions in the creation of content and the increasing importance of crowdsourcing in lexicography.

Learning Outcomes

Upon completion of this course, students will be able to:

- understand the role of users in dictionary-making process
- distinguish between different types of users, as well as understand some of the specifics of different user groups
- know different methods for conducting user research, including advantages and shortcomings of each of the method
- understand the concept of crowdsourcing and different techniques of using the method in dictionary compilation.

TOC

- **Dictionary users and their role in making of a dictionary:** the importance of the fact that dictionaries are made for users, that understanding users needs and habits is paramount when preparing, updating and improving dictionaries
- **Different types of dictionary user:** presenting different types of users and what that means when preparing a dictionary. It is likely that one dictionary will address the needs of different types of users, however one group should be the main target group. Differences between the groups of users across different countries.
- **Dictionary use research**
 - **Overview of literature:** more and more research available, in the past very English-centred, especially English as a foreign language research, there is also research made by publishers about the use of their dictionaries but not made available
 - **Methods:**
 - Direct:
 - questionnaire: description, advantages and shortcomings, etc.
 - interview
 - think-aloud protocol
 - eye-tracking
 - user evaluation/feedback
 - Indirect:
 - log file analysis
 - user monitoring
 - language problems (learner corpus analysis, forums...)
- **Involving users in dictionary development**
 - crowdsourcing, citizen science
 - gamification
- **Case studies:**
 - international survey
- **User research in digital era: the way forward**
 - responsive dictionaries

Introduction to Corpus-Based Lexicographic Practice

Level: Introductory

Contributors: JSI, IVDNT, LC

Authors: Carole Tiberius, Ondřej Matuška, Iztok Kosem, Vojtech Kovar

Staging URL

<https://elexis.humanistika.org/id/xvxB5cdlPUECtRmZ0o563>

Final URL

<https://campus.dariah.eu/id/xvxB5cdlPUECtRmZ0o563>

Learning Objectives

This course will explore the notion of lexicographic evidence and the limitation of subjective views on language by tracing the changes in lexicographic practice from the extensive use of manually selected citations to the employment of large language corpora. The course will introduce the fundamentals of corpus linguistics and corpus design, as well as the role played by various NLP tools such as taggers and parsers. At the end of this course, the students will have a basic understanding of how corpora can be used in dictionary writing, both for lemma selection, sense disambiguation, composing good definitions, choosing good examples and for automatic term and collocation extraction.

Learning Outcomes

Upon completion of this course, students will be able to

- understand the fundamentals of corpus linguistics and corpus design
- determine the suitability of a corpus for a particular project
- understand the role of the corpus in lexicographic practice

TOC

- Lexicographic evidence: from slips to corpora
- Corpora
 - corpus types
 - corpus design
 - corpus building
 - corpus annotation
 - corpus query systems
- Getting to know your corpus
- Corpora in lexicographic practice
 - headword lists
 - collocations
 - neologisms
 - example sentences
 - thesaurus
 - word senses in a nutshell
 - labels
 - syntactic patterns
 - translations
- Next steps

Capturing, Modeling and Transforming Lexical Data: An Introduction

Level: Introductory

Contributors: BCDH

Authors: Toma Tasovac

Staging URL

<https://elexis.humanistika.org/id/3yV0rXqfTqBFa3jXHm1HX>

Final URL

<https://campus.dariah.eu/id/3yV0rXqfTqBFa3jXHm1HX>

Learning Objectives

This course will introduce the theories, practices, and methods of digitizing legacy dictionaries for research, preservation and online distribution by focusing on the process of converting paper-based dictionaries to electronic format through image capture, text capture, data modeling and data enrichment. In addition to explaining how various OCR and HCR tools can be used to extract text from images, the course will focus on analyzing, identifying and describing lexicographic data using markup languages such as XML in order to produce semantically structured datasets that can be easily queried, shared and transformed to different outputs. The course will be of interest not only to those who are converting legacy dictionaries, but also those who want to understand the principles and modes of representing structured lexicographic data, which will be a prerequisite for more advanced coursework on TEI Lex-0, OntoLex-Lemon, XPath and XSLT.

Learning Outcomes

Upon completion of this course, students will be able to

- understand the challenges involved in converting paper-based dictionaries to digital format
- identify tools that can help with the conversion process
- appreciate the complexity of lexicographic data as structured data
- model simple dictionary entries and express them in XML

TOC

- **Introduction**
- **Digitization workflow:**
 - planing (conceptual vs. organizational phase)
 - image and text capture
 - data modeling
 - data enrichment

- publication
- maintenance
- **Capturing lexicographic data**
 - Image capture
 - digital cameras vs. scanners
 - document scanners vs. book scanners
 - image formats and optimization
 - Text capture
 - OCR
 - HRT
 - Manual input
 - single vs. double (or multiple) keying
 - Post-correction: OCR, HRT, comparison tools
 - Rich (styled) text vs. plain text
- **Modeling lexicographic data**
 - What is a model? Why do we model? Model-of vs. model-for.
 - Markup languages
 - Textual hierarchies, OHCO (ordered hierarchy of content objects) and the notion of structured data
 - Types of markup and metadata
 - Modeling with XML
 - Intro to XML as markup language
 - XML and its advantages
 - device independence
 - machine and human readability
 - separation of data and visual presentation
 - searchability. XPATH.
 - transformability. XSLT.
 - XML: Building blocks and basic rules
 - Elements and rules for element names
 - Attributes
 - Processing instructions and character entities
 - Nesting elements
 - Well-formed vs. valid XML
 - Element relationships (children, parents, siblings, ancestors etc.)
- **Where next?** *This section will point in the direction of other courses in the ELEXIS Curriculum, especially TEI, Ontolex Lemon, Mastering oXygen, XPATH and XSLT for Nerds.*

LEX2: Mastering ELEXIS Corpus Tools for Lexicographic Purposes

Level: Intermediate

Contributors: LC, IVDNT

Authors: Miloš Jakubíček, Vojtěch Kovář, Ondřej Matuška, Carole Tiberius,

Staging URL

<https://elexis.humanistika.org/id/vfyLBgtN4bMmxC4Me97ql>

Final URL

<https://campus.dariah.eu/id/vfyLBgtN4bMmxC4Me97ql>

Learning Objectives

The course will introduce corpus tools available in ELEXIS and describe various ways in which they can be exploited in lexicographic research and for compiling dictionaries. At the end of the course, students will be able to build corpora, as well as evaluate their quality and suitability for the respective lexicographic task. Students will also acquire the skills necessary to work with the tools effectively to retrieve valid linguistic information from the corpus. Practical examples will simulate the use of these tools in the lexicographic workflow.

Learning Outcomes

Upon completion of this course, students will be able to

- build a corpus in Sketch Engine
- use Sketch Engine for data analysis in the lexicographic workflow

TOC

- Building a corpus
 - from user's own data
 - from the web
- Evaluate the quality and suitability for the lexicographic task
- Using corpus tools to extract lexicographic information from corpora:
 - headword lists (including keyword/term/word lists) - video
 - frequency information
 - part of speech information
 - collocations
 - example sentences
 - word senses in a nutshell
 - syntactic patterns
 - translations
 - neologisms
 - thesaurus
 - concordance annotation tool
 - concordance advanced search, CQL
- Where to go next
 - Automating the process
 - Postediting

Lexonomy: Mastering the ELEXIS Dictionary Writing System

Level: Intermediate

Contributors: LC, IVDNT, JSI, RILMTA

Authors: Tamás Váradi, Ondřej Matuška, Carole Tiberius, Iztok Kosem, Vojtěch Kovář, Miloš Jakubiček

Staging URL

https://elexis.humanistika.org/id/Qm_2SzS_rGB-Py9YTCHYm

Final URL

https://campus.dariah.eu/id/Qm_2SzS_rGB-Py9YTCHYm

Learning Objectives

The course will explore how software tools for dictionary production (so-called dictionary writing systems, or DWS) can be used to streamline and facilitate the structural coherence and quality assurance in a dictionary project by focusing on Lexonomy, a dictionary-writing system developed as part of ELEXIS. At the end of this course the students will know how to use Lexonomy in various stages of the lexicographic workflow, from creating a dictionary, selecting a suitable data model, and setting up different configuration options, to using advanced features such as workflow monitoring, and preparing the dictionary for publication. Selected existing projects will be used as case studies.

Learning Outcomes

Upon completion of this course, students will be able to

- assess DWS in general
- create a dictionary in Lexonomy (monolingual and bilingual)
- set up different configuration options (including defining the entry structure)
- set up and use the Sketch Engine API for pulling data
- use advanced features such as workflow monitoring
- upload/download a dictionary
- prepare the dictionary for publication
- (link to other dictionaries)

TOC

- Introduction
 - What is DWS
 - Introducing Lexonomy
- Let's get started
- Creating a new dictionary
 - from scratch

- from an automatically generated draft
- Editing a dictionary
- Publishing a dictionary
 - setting up the dictionary for publishing
- A case study
- Where to go next

Automating the Process of Dictionary Creation

Level: Intermediate

Contributors: LC, IVDNT, JSI

Authors: Miloš Jakubiček, Vojtěch Kovář, Ondřej Matuška, Carole Tiberius, Iztok Kosem

Staging URL

https://elexis.humanistika.org/id/ICij_0Jcoz2uHR1u2uRhW

Final URL

https://campus.dariah.eu/id/ICij_0Jcoz2uHR1u2uRhW

Learning Objectives

Building upon the material covered in *LEX2: Mastering ELEXIS Corpus Tools for Lexicographic Purposes* and *Lexonomy: Mastering the ELEXIS Dictionary Writing System*, this course will focus specifically on the changes in dictionary production after 2000 and the increasing importance of automation and post-editing in lexicography. The course will focus on the ELEXIS One-Click Dictionary as a corpus-based dictionary-drafting tool and the way Lexonomy DWS can be used to post-edit content which has been automatically pulled from ELEXIS Corpus Tools.

Learning Outcomes

Upon completion of this course, students will be able to

- appreciate/understand the changes in the process of creating dictionaries in the past few decades
- use the OnceClick Dictionary functionality in Sketch Engine
- post-edit automatically pulled content in Lexonomy

TOC

- The rise of post editing in lexicography
 - effectivity and practical considerations
 - streamlined postediting by data type
 - management implications of the post editing process
- Dictionary entry components and their automatic extraction from corpora
 - headword list editing - selecting the right entries
 - data from a corpus without text types

- data from a corpus with text types
- word sense disambiguation
- Automatic drafting of dictionary entries
 - OneClick Dictionary
- Post-editing in a dictionary writing system
 - Lexonomy
- Case studies

CLARIN Tools and Resources for Lexicographic Work

Level: Intermediate

Contributors: CNR-ILC, OeAW

Authors: Francesca Frontini, Andrea Bellandi, Valeria Quochi, Monica Monachini (CNR-ILC); Karlheinz Mörth, Susanne Zhanial, Matej Ďurčo, Anna Woldrich (OeAW)

Staging URL

<https://elexis.humanistika.org/id/UnwYPq70Dewbn7XDEjsMM>

Final URL

<https://campus.dariah.eu/id/UnwYPq70Dewbn7XDEjsMM>

Learning Objectives

This course will present an overview of resources available from CLARIN that may be useful for the lexicographer; we refer to lexical datasets but also to textual resources such as corpora, as well as tools..

We will start from a brief outline of the CLARIN infrastructures and of its centres. Then we will show how to deposit your lexical resource in a CLARIN centre for visibility, citability, versioning and long term preservation. We will highlight links and references with other modules in the course as concerns the paradigm of Open Data and formats and standards to represent lexical data.

We will then give an overview of the resources available from the CLARIN infrastructure, by browsing the CLARIN Virtual Language Observatory and the CLARIN resource families, in particular the ones dedicated to Lexica and Dictionaries. We will illustrate the metadata used for describing these resources and the differences in terms of formats and accessibility .

We will conclude with an exploration of some of the TOOLS present in the CLARIN Switchboard which can be used for lexicographic purposes.

Learning Outcomes

Upon completion of this course, students will be able to

- appreciate importance of FAIR and Open data and the role of research infrastructures to ensure that lexical data is correctly preserved
- deposit their lexical resource on a CLARIN DSPACE type of repository (such as ILC4CLARIN), with the appropriate metadata, understand the versioning of resources, cite their resource appropriately
- search for and locate existing lexical resource within the CLARIN infrastructure (VLO, Resource Families)
- create a simple lexical resource by means of LexO, a collaborative editor of OntoLex-Lemon resources
- build a simple lexical resource (in TEI) with the <TEI>enricher

TOC

- The CLARIN infrastructure
 - What is CLARIN
 - CLARIN and FAIR, Open Data
- Depositing your lexicon / dictionary in a CLARIN centre
 - Citability
 - Versioning
 - Standards
- Searching for resources in the CLARIN infrastructure
 - On the VLO (<https://vlo.clarin.eu/>)
 - In the CLARIN Resource Families (<https://www.clarin.eu/resource-families>)
 - RF Lexical resources, RF dictionaries, other relevant textual resources
 - Some examples of browsing interfaces
- (Practical module) Using CLARIN Tools for lexicography:
 - ILC4CLARIN - Modelling Lexical Conceptual resources with LexO
(->[link to module "Modeling Dictionaries in OntoLex-Lemon"](#))
<https://ilc4clarin.ilc.cnr.it/services/lexo-info/>
 - OeAW - Creating Lexical Resources with the <TEI>Enricher

Standards for Representing Lexicographic Data: An Overview

Level: Intermediate

Contributors: FCSH-UNL

Authors: Rute Costa, Christophe Roche, Ana Salgado

Staging URL

<https://elexis.humanistika.org/id/REhOykBU7pPs5zOAENDah>

Final URL

<https://campus.dariah.eu/id/REhOykBU7pPs5zOAENDah>

Learning Objectives

This course will impress upon students the importance of shared standards for the productive cooperation among lexicographers in a multicultural and multilingual context by referring to a number of lexicographically relevant standards such as TEI, TEI Lex-0, ISO 1951, ISO 24613 (LMF), OntoLex-Lemon etc. The relevant standards will be classified according to their aims (those dealing with linguistic content; those that are used for annotation purposes; those that handle representation purposes; those addressing interoperability issues etc.).

At the end of the course, students will be able to conduct dictionary research and work with dictionaries, taking into account standards in different parts of their lexicographic workflows.

Learning Outcomes

Upon completion of this course, students will be able to

- appreciate the importance of the standards for the lexicographical work;
- understand the purpose of the diversity of the standards;
- know the different formats used for interoperability matters;
- apply the standards to the digital dictionary-making;
- understand how to model a dictionary entry in different standards;
- prepare data for operationalization needs;
- apply the acquired knowledge to concrete examples.

TOC

TABLE OF CONTENTS

- Part 1: Standards and Formats
 - Text Encoding Initiative
 - 1.1 TEI
 - 1.2 TEI Lex-0
 - 2. ISO Standards related to lexicography
 - 3. W3C Standards
- Part 2: Formal Representation of Lexical Data
 - 1. Core vocabulary to lexicographic work
 - 1.1 dictionary
 - 1.2 lexicographic article
 - 1.3 dictionary structure
 - 1.4 macrostructure
 - 1.5 microstructure
 - 1.6 mediostructure
 - 1.7 lexicon
 - 1.8 lexical unit
 - 1.9 lexicographic component
 - 1.10 headword
 - 1.11 sense
 - 1.12 part of speech

- 1.13 canonical form
- 1.14 lemma
- 1.15 sense component
- 1.16 lexicographic definition
- 1.17 label
- 1.18 note
- 2. Lexicographic article
 - 2.1 English lexicographic article: fly
 - 2.2 French lexicographic article: mouche
 - 2.3 Portuguese lexicographic article: mosca
- 3. Modelling and encoding lexicographic article
 - 3.1 OntoLex Lemon
 - 3.2 TEI Lex-0
 - 3.3 LMF
- References

Modeling Dictionaries in TEI Lex-0

Level: Intermediate

Contributors: BCDH, CNR-ILC, FCSH-UNL

Authors: Toma Tasovac

Staging URL

<https://elexis.humanistika.org/id/en5qQnz7D4o0SHU33KqiS>

Final URL

<https://campus.dariah.eu/id/en5qQnz7D4o0SHU33KqiS>

Learning Objectives

The course will focus on modeling dictionaries using TEI Lex-0, a subset of the community standard TEI (Text Encoding Initiative). The course will focus on best-practices and recommendations in view of accuracy, consistency and interoperability of lexicographic data. At the end of this course, the students will become familiar with the underlying principles and the explicit guidelines of TEI Lex-0 by learning how to encode a number of dictionary entries through step-by-step tutorials with the ultimate goal of being able to adopt TEI Lex-0 in their own work.

Learning Outcomes

Upon completion of this course, students will be able to

- understand the importance of using community-established best practices for encoding dictionaries
- recognize the differences in scope between TEI and TEI Lex-0

- efficiently use both TEI and TEI Lex-0 Guidelines
- encode moderately complex dictionary entries using TEI Lex-0
- identify channels of communication with TEI and TEI Lex-0 experts

TOC

- **Introduction: Text Encoding Initiative.** *Community standard. Not only dictionaries. Bigger picture: digital humanities, textual resources.*
- **TEI vs. TEI Lex-0**
 - What is TEI Lex-0 and why do we need it?
 - Markup expressivity vs. interoperability
 - Navigating TEI and TEI Lex-0 Guidelines
- **Encoding dictionary metadata**
 - `teiHeader`
- **Encoding dictionary components**
 - `entry`
 - why we need nested entries
 - written and spoken forms
 - lemmas
 - inflected forms
 - variants
 - orthographic
 - phonetic
 - regional
 - `mwe's`
 - grammatical information
 - `gramGrp` as a container
 - `gram` types
 - senses
 - `<sense>` as a microstructural container
 - definitions
 - examples
 - quotes
 - bibliographic references
 - translation equivalents
 - usage
 - cross-references
 - `<xr>` as a container
 - `ref`
- **From data encoding to data enrichment.** *Granularity of markup. Making explicit what is implicit vs. enriching existing data. Examples: georeferencing dialectal entries. Knowing where to stop. Limits of automation and costs of manual encoding.*
- **Where next?** *This section will point in the direction of other courses in the ELEXIS Curriculum, especially OntoLex Lemon, Mastering oXygen, XPATH and XSLT for Nerds, as well as pointing to ELEXIS tools (for instance: Elexifier).*

Modeling Dictionaries in OntoLex-Lemon

Level: *Intermediate*

Contributors: *NUIG, CNR-ILC*

Authors: *John McCrae, Fahad Khan*

Staging URL

<https://elexis.humanistika.org/id/XUpXfINpQ08j3hdNp7pSi>

Final URL

<https://campus.dariah.eu/id/XUpXfINpQ08j3hdNp7pSi>

Learning Objectives

This course describes the OntoLex-Lemon model, a recent standard for the representation of lexical information on the Web as linked data. In addition to providing a basic introduction to linked data and the Resource Description Framework (RDF), the course will cover the core model of OntoLex and how to represent basic lexical information. Additional modules of the OntoLex module for the description of syntax, term decomposition, variation & translation, metadata, lexicography, morphology and corpus information will also be described. At the end of the course, students should be able to express lexicons as linked data using the model.

Learning Outcomes

Upon completion of this course, students will be able to

- Understand how lexical data can be represented as linked data
- Describe the benefits of linked data for lexicography
- Use the OntoLex-Lemon module to model lexical information
- Design a schema based on OntoLex-Lemon to represent a specific lexical resource

TOC

- Introduction to RDF
 - RDF Data Model
 - Turtle Syntax
 - URLs
- Design of OntoLex-Lemon Model
- The OntoLex-Lemon Core Model
 - Entries
 - Forms
 - Senses
- Modules
 - Syntax and Semantics
 - Decomposition
 - Variation and Translation
 - Metadata

- Morphology
- Frequency, Attestation and Corpus Information
- Practical Modeling in OntoLex-Lemon

LEX3: Transforming Legacy Dictionaries using Elexifier

Level: Intermediate

Contributors: JSI, IVDNT

Authors: Carole Tiberius, Tina Munda, Andraž Repar

Staging URL

<https://elexis.humanistika.org/id/ltzUH-fnmqzR5O0bqqwu>

Final URL

<https://campus.dariah.eu/id/ltzUH-fnmqzR5O0bqqwu>

Learning Objectives

This course will introduce Elexifier, a cloud-based dictionary service for the conversion of legacy XML and PDF dictionaries into a shared data format based on the ELEXIS Data Model.

Learning Outcomes

Upon completion of this course, students will be able to

- convert XML or PDF dictionaries into TEI-compliant XML in line with the specification described in the ELEXIS data model.

TOC

- Introduction:
 - Elexifier
 - ELEXIS Data Model
- Let's get started
- Transforming an XML Dictionary
- Transforming a PDF Dictionary
- Hands-on exercise
- Next steps

LEX3: Publishing Legacy Dictionaries with Publex

Level: Intermediate

Contributors: UT

Authors: Anne Klee, Thomas Burch, Claudia Bamberg, Julia Hennemann, Henrike Sievers, Sandra Weyand

Staging URL

<https://elexis.humanistika.org/id/IbZ4arshkA06y5Uem0UsT>

Final URL

<https://campus.dariah.eu/id/IbZ4arshkA06y5Uem0UsT>

Learning Objectives

This course will introduce the Legacy Dictionary Viewer Publex, a generic, modular dictionary publication tool for retrodigitized dictionaries.

TOC

- **Introduction: What is Publex?** *Functions and target user*
- **Let's get started**
 - Registration and profile management
 - How to create a dictionary and import data (*data requirements*)
 - File Management (*How to update dictionary data*)
 - Dictionary Metadata Management
- **Configuration and preview** *Defining formatting rules (font and layout) for XML-elements*
- **Dictionary Viewer** [Format: text + video] *Components and utilisation of the Dictionary Viewer*
- **Publication of a dictionary** [Format: text + video]
- **It's getting technical - for programmers and developers** [Format: text]
 - Technical infrastructure used to build Publex
 - How to install Publex on your own server
- **Hands-on** *The user receives test data and tasks to run through the workflow once by himself*

Lexicography in the Age of Open Data

Level: Intermediate

Contributors: CNR-ILC, NUIG

Authors: Fahad Khan, John McCrae

Staging URL

https://elexis.humanistika.org/id/n7FC_IX46pwwJQN50Dtjs

Final URL

https://campus.dariah.eu/id/n7FC_IX46pwwJQN50Dtjs

Learning Objectives

Technology should not necessarily be seen as the ultimate challenge facing lexicography today: social, cultural and legal obstacles often stand in the way of collaboration and knowledge sharing. This course will explore the principles of open access, open data, FAIR principles and open science as they apply to lexicography including the specific challenges posed by intellectual property rights and copyright issues in the context of lexicographic work.

Learning Outcomes

Upon completion of this course, students will be able to

- Understand the FAIR principles and why they matter to the promotion of Open Science and Open Access,
- Understand how the FAIR principles can apply to lexicographic resources both from the user and creator's point of view,
- Understand how standards and infrastructures can help in the creation and maintenance of FAIR resources,
- Understand the main initiatives, projects and organisations which promote FAIR
- Have a basic grasp of the main IPR licences and when to use them

TOC

- **Brief introduction to Open Science**
 - Introduction to initiatives that are currently promoting or supporting Open Access. Discussion of potential of infrastructures like ELEXIS as well as infrastructures such as CLARIN to help make the relevant data, resources and tools (e.g., lexica, corpora, NLP tools) openly available to lexicographers and others interested in dictionaries
 - Discussion of importance of using standards such TEI, Ontolex-lemon, LMF, forthcoming ELEXIS standard) for producing reusable and interoperable resources
- **An introduction to the FAIR principles for lexicographers**
 - Why does FAIR matter?
 - Description of the FAIR Guidelines (with special emphasis on those principles relevant to lexicographic data and to showing how they can be implemented through existing models and vocabularies such as TEI, LMF, Ontolex). i.e., How to make your (lexicographic data) FAIR
 - The importance of Data Management Plans and how to create them using readily available tools
- **Finding out more: Initiatives, projects, and organisations which promote FAIR**
 - An introduction to initiatives, projects and organisations which promote FAIR and from which lexicographers can find out more about FAIR
- **Intellectual Property Rights and Copyright Issues**
 - A brief introduction to IPR issues
 - An introduction to the different kinds of licenses and when to use them

Mastering LEX1: The Dictionary Matrix

Level: Advanced

Contributors: JSI, NUIG, FCSH-UNL

Authors: Iztok Kosem

Staging URL

https://elexis.humanistika.org/id/EwnzIOPkgATuEVHML_94G

Final URL

https://campus.dariah.eu/id/EwnzIOPkgATuEVHML_94G

Learning Objectives

This course will introduce the concept and the ELEXIS implementation of the dictionary matrix, a universal repository of linked senses, and other types of lexical information found in existing lexicographic resources. Students who complete the course will become a) familiar with the techniques used to compile the dictionary matrix; b) capable of searching the dictionary matrix using the GUI (with input fields, dropdowns etc.) and/or using a SPARQL endpoint; and c) aware of the steps needed to contribute to the matrix by linking monolingual dictionaries at a sense level using the tools provided by LEX1; or extending this to multilingual linking using BabelNet.

Mastering oXygen XML Editor for Dictionary Nerds

Level: Advanced

Contributor: BCDH

Authors: Toma Tasovac

Staging URL

https://elexis.humanistika.org/id/CPYfBfo_5Rz1d8tow9Gfu

Final URL

https://campus.dariah.eu/id/CPYfBfo_5Rz1d8tow9Gfu

Learning Objectives

This course will help users learn how to use oXygen XML, a versatile, professional-grade XML editor to edit, validate, query and transform lexicographic data. The course will focus on a range of practical and time-saving functionalities and features, such as the schema-based tool tips, document and code templates, customizable shortcuts as well as different editing modes (text, grid and CSS-styled Author) which can be used to adapt the display of dictionaries to one's particular editing needs. At the end of this course, the students will have become accustomed to the oXygen editing environment and ready to work productively and efficiently in it.

Learning Outcomes

Upon completion of this course, students will be able to:

- customize their working environment in oXygen to improve lexicographic workflows
- understand, set up and exploit oXygen *projects*
- create individual shortcuts and use code templates to speed up encoding practice
- use create styles for Author Mode in oXygen

TOC

- **Introduction**
- **Projects**
 - **Creating new projects**
 - **Managing projects**
 - Project view
 - Adding resources to projects
 - Moving and renaming resources
 - **Batch validation and transformation**
 - **Sharing projects**
- **Author mode**
 - **Using author mode.** Including how to “surround with code”, edit attributes etc.
 - **Modifying CSS stylesheets**
- **Shortcuts and templates**
 - **Default shortcuts**
 - **Customizing shortcuts**
 - **Code templates**

Extracting Lexical Data: XPath for Dictionary Nerds

Level: Advanced

Contributor: BCDH

Authors: Toma Tasovac

Staging URL

<https://elexis.humanistika.org/id/XUDsnt3PCB0WTEr-6EM7w>

Final URL

<https://campus.dariah.eu/id/XUDsnt3PCB0WTEr-6EM7w>

Learning Objectives

This course will cover the fundamentals of XPath (XML Path Language), a standard query language for selecting nodes from XML documents. After explaining the basic syntax of XPath (axes, nodes and predicates), the course will guide the students through a number of real-life dictionary-specific scenarios (for instance: how do you select only those entries whose translation equivalents are missing the xml:lang attribute? how do you select only those entries whose etymological information contains etymons from Latin?) in order to help them hone their skills. At the end of the course, students will be able to write their own XPath expressions in order to navigate around XML-encoded dictionaries and select only those bits of data that they are interested in.

Learning Outcomes

Upon completion of this course, students will be able to

- recognize the advantages of XPath over plain-text search in lexicographic contexts
- understand the basic syntax of XPath as a query language
- apply XPath to real-life scenarios for querying dictionary data
- write their own XPath expressions

TOC

- **Introduction: what is XPath?**
 - Prerequisites
 - Another language: why, oh why?
 - What do I need to work with XPath
- **XPath in oXygen**
 - XPath input field
 - XPath Builder
 - Launching XPath Builder
 - Executing XPath from the Builder
 - XPath expressions
 - Selecting nodes
 - Selecting predicates
 - Selecting unknown nodes
 - Selecting more than one path
 - Axes
- **Exercises**

Transforming Lexical Data: XSLT for Dictionary Nerds

Level: Advanced

Contributor: BCDH

Authors: Toma Tasovac

Staging URL

<https://elexis.humanistika.org/id/DPF60ESE5Bf1jmjtPXzWI>

Final URL

<https://campus.dariah.eu/id/DPF60ESE5Bf1jmjtPXzWI>

Learning Objectives

The course builds upon *Extracting Lexical Data: XPath for Dictionary Nerds* and introduces the basics of XSL Transformations (XSLT), a standard language for transforming XML documents. After explaining the basic syntax and processing model of XSLT (stylesheet declarations, templates, pattern matching etc.), the course will guide students through a number of real-life dictionary-specific scenarios (renaming, adding or removing elements and attributes, rearranging and sorting elements, performing tests, hiding and showing portions of the dictionary content etc.) in order to help them improve their skills. At the end of this course, students will be able to write their own XSLT stylesheets to transform lexicographic data.

Learning Outcomes

Upon completion of this course, students will be able to

- understand the basic syntax and processing model of XSLT
- assess different use-case scenarios for XPath and XSLT
- write their own XSLT stylesheets to transform lexicographic data

TOC

- Introduction
 - What is XSLT?
 - Prerequisites
 - XPATH vs. XSLT
- XSLT syntax and processing model
 - Stylesheet declarations
 - Templates
 - Pattern matching
- Processing dictionary data
 - Manipulating elements
 - Manipulating attributes
 - Counting dictionary components

- Performing tests
- Hiding and showing dictionary components
- Gathering HTML views of dictionary content
- Exercises

ELEXIS Pathfinder to Computational Lexicography for Developers and Computational Linguists

Level: Advanced

Contributors: IVDNT

Authors: Carole Tiberius, Kris Heylen

Staging URL

<https://elexis.humanistika.org/id/VL3saRXqj1b2la911h1yK>

Final URL

<https://campus.dariah.eu/id/VL3saRXqj1b2la911h1yK>

Learning Objectives

This learning resource provides software developers and computational linguists with an overview of the typical computational processing tasks and software tools in the lexicographic workflow. The resource introduces the most widely used custom developed tools for corpus-based lexicography as well as their functionality. Additionally, we give pointers to trending R&D topics in computational lexicography ranging from automatic extraction of good examples and word-sense disambiguation to machine learning. The course is aimed at developers and computational linguists who are new to lexicography and who quickly want to familiarise themselves with the most important literature and/or available learning resources which can bring them up to speed in the fast-changing field and prepare them to collaborate in a multidisciplinary team of lexicographers, computational linguists and software developers.

Learning Outcomes

Upon completion of this course, students will be able to

- assess different tools used in the lexicographic workflow
- understand the typical functions and custom features of lexicographic software tools
- compare and evaluate different approaches to automatic extraction of lexicographic data (e.g. headword lists, collocations and example sentences)
- find the most important literature and/or available learning resources in computational lexicography

TOC

- Software development in a lexicographic team

D5.3 Overview of Online Tutorials and Instruction Manuals

- Overview of tools in the lexicographic workflow
- Typical functions and custom features of lexicographic tools
- Automatic extraction of lexicographically relevant information from corpora
- Pointers to other resources