

TeLeMaCo—A tool for the dissemination of teaching and learning materials

Hannah Kermes and **Jörg Knappen** and **José Manuel Martínez** and **Elke Teich** and **Mihaela Vela**
Fachrichtung 4.6, Universität des Saarlandes, Campus A2.2, D-66123 Saarbrücken, Germany
{h.kermes, j.knappen, j.martinez, e.teich, m.vela}
@mx.uni-saarland.de

Abstract

This paper presents TeLeMaCo, a collaborative portal for training and teaching materials relevant in linguistics and digital humanities hosted at the CLARIN-D centre at Saarland University in Saarbrücken. The portal is easy to use both for casual users who search for teaching and training material and for community members who want to contribute descriptions of their materials. We collect structured metadata of the described resources to provide advanced search and to integrate them in the wider CLARIN framework of resources.

1 Introduction

For language resources and tools, there is a growing number of repositories, and there are platforms to search the metadata of many collections at once, like the Virtual Language Observatory (VLO) (Van Uytvanck et al., 2012). Also there are web services and chaining tools such as, e.g., WebLicht (Hinrichs et al., 2010) that provide facilities for text and speech processing and support processing pipelines for a wide variety of scientific tasks.

However, the documentation and teaching materials remain scattered over many places, including institutional web pages, YouTube channels, or software repositories like sourceforge¹ or Bitbucket². A common interface to access and search those teaching and learning materials was lacking when we started the design of our service.

We developed the **Teaching and Learning Material Collection** (TeLeMaCo³) to overcome this situation. Our approach is community driven as we collect descriptions of relevant materials contributed from all over the world in our service.

TeLeMaCo offers search and access to a wide range of teaching and learning materials, including the following

- technical documentation (e.g., quick starts, tutorials, or full manuals),
- learning material for self study (e.g., YouTube videos and screen casts),
- short teaching modules (2–4 hours) that can be integrated in existing courses,
- full courses covering a broader spectrum of language resources and tools or focusing on specific topics of application of language resources and tools,
- reference materials (e.g., specialised dictionaries).

TeLeMaCo brings together materials stored at different institutions and locations through a unified interface and it provides access to materials published in different languages (currently mostly German

This work is licensed under a Creative Commons Page numbers and proceedings footer are added by the organisers. Licence details: <http://creativecommons.org/licenses/by/4.0/>

¹<http://sourceforge.net/>

²<https://bitbucket.org/>

³<http://fedora.clarin-d.uni-saarland.de/hub/>

and English, but also French). For some tools, e.g., EXMARaLDA (Schmidt and Wörner, 2009), there is already a comprehensive coverage of the available materials.

In Section 2 of this paper we present some ideas behind TeLeMaCo. In Section 3 we describe TeLeMaCo as seen from a casual user searching for learning and teaching materials. In Section 4 we explain how to add materials to the TeLeMaCo service. Section 5 is about the metadata we store and in Section 6 we talk about some additional benefits of TeLeMaCo. We conclude with an evaluation of TeLeMaCo (Section 7) and a comparison to three other services for the dissemination of teaching and learning materials (Section 8).

2 Ideas behind TeLeMaCo

2.1 A social network

The idea behind TeLeMaCo is that the user community can easily contribute descriptions of teaching and learning materials. We see it as a collaborative and ongoing effort, not as the work of a small dedicated project which is finished at some fixed date.

TeLeMaCo allows the users to choose their own tags for the teaching and learning materials, creating some kind of folksonomy (Vander Wal, 2007). We expect some alignment of the chosen keywords because the user can easily see already existing keywords and we have implemented the autosuggest feature of tags in the service.

TeLeMaCo is not a wiki; every contributor to TeLeMaCo “owns” the descriptions he/she has added and he/she’s the only one (except for the administrators) who can change them.

Another kind of user interaction is given by the feedback system we have implemented. Any user can give feedback on the usefulness of the material (and of its description) by answering four simple questions (see Figure 2). TeLeMaCo displays the aggregated feedback score.

We also allow the user to report on quality issues like stale links or inappropriate content to the portal administrator.

2.2 Implementation considerations

TeLeMaCo is implemented using the well-established and sustainable LAMP (Linux, Apache httpd, mysql, Python) software stack with Django as a web framework. All tasks are automated to minimise administrator interaction with the system.

All local dependencies are stored in a small configuration file. TeLeMaCo can be migrated to another machine with minimal effort.

We have designed the URLs of the description pages (see Figure 1 for an example) to be plain bookmarkable URLs, no forms are involved. The same holds for other primary pages, like the browsing interface, the *what’s new* page, or the (currently experimental) *metadata* pages. These pages are easy to index by search engines and get good rankings.

2.3 Embedding into the CLARIN infrastructure

TeLeMaCo is a complementary part of the CLARIN infrastructure. It does not copy features of existing services like WebLicht or the VLO; it supplements them. It uses CMDI aware metadata, and an integration with the VLO using an OAI provider is planned. In addition, TeLeMaCo is integrated into the Helpdesk of CLARIN-D.

3 Searching and browsing TeLeMaCo

TeLeMaCo offers several facilities to search for materials: Simple and advanced search as well as browsing keywords and authors.

Simple search does a substring query over author, title, keyword and description, this way optimising the recall.

Advanced search allows querying specific fields, e.g., to get only materials in a certain language.

TeLeMaCo: TAToM: ...

fedora.clarin-d.uni-saarland.de/hub/resource/313

The Linguistic Teaching Resources Hub

CLARIN-D

Image © Paul Watson, Licence CC BY-NC-SA 2.0

Home | What's new? | Browse | Login / Create Account | Advanced search | search Search

TAToM: Text Analysis with Topic Models for the Humanities and Social Sciences

* Allen Riddell *

Keywords: [topic model](#), [NLTK](#), [python](#), [chunking](#), [tokenization](#), [MALLET](#)

<https://de.dariah.eu/tatom-intro>

This tutorial explains basic techniques of text analysis from the very beginning (starting with the introduction of the necessary software and pointing to tutorials on it) and in great detail.

Table of contents:

- Preliminaries & Getting started
- Working with text
- Preprocessing
- Feature selection: finding distinctive words
- Topic modeling with MALLET
- Topic modeling in Python
- Visualizing topic models
- Classification, Machine Learning, and Logistic Regression
- Case Study: Racine's early and late tragedies

Feedback

Sorry, there is no feedback available. Be the first one to provide feedback!

[Give feedback](#)

Resource details

Institution: DARIAH-DE
 Year of publication: 2014
 Language: english
 Type: Tutorial
 Audience:
 Level: basic
 Prerequisites: none
 Media: text/html
 Objective:
 Licence: CC-BY 4.0
 Access: open
 Creation date: Thursday, 31 July 2014 14:00:04
 Last modified: Thursday, 31 July 2014 14:00:04
 BibTeX type: @misc
 BibTeX entry:

Figure 1: Display page of a sample material in TeLeMaCo

Was the described material relevant for me?— No, not at all ... Yes, very much
 I think, the level of the described material is ... —Basic ... Expert
 I reached the objectives given by studying the material.— No, not at all ... Completely
 I think, the prerequisites are—Grossly wrong ... Accurately stated

Figure 2: The four feedback questions and their answer ranges

Alignieren, Aligning, Analysis, ANNIS, annotation, annotation management, annotation of speech data, annotation panel, Annotieren, AntConc, Artificial Intelligence, Äußerungsliste, automatic segmentation, bash, Bayes, character encoding, chunking, CLARIN-D, COCA, collocation, Coma, combinatorics, command line, Concordance, corpus, corpus analysis, corpus annotation, corpus linguistics, Corpus Manager, corpus search, corpus workbench, corrections, COSMAS II, CQPweb, creation of speech data, data mining, Demo Corpus, DeReKo, dictionary, digital editing, Dublin Core, entropy, errors, estimator, etree, EXAKT, EXMARaLDA, Fehler, formant analysis, fraktur, GATE, geolocation, Gesprächslinguistik, glossary, HIAT, hidden Markov model, historical text, Importieren, Importing, infrastructure, intensity analysis, Introduction, IPA, Java, JavaScript, Korpus erstellen, Korpuslinguistik, Korrigieren, language resources, latent Dirichlet allocation, LaTeX, LaTeX2e, LDA, lemmatization, lexicography, lexicon, linguistic resources, linguistics, Linux, lxml, Machine Learning, Maintenance, MALLET, manual annotation, markup language, MAUS, Merging, metadata, Metadaten, MMAX2, MOSES, named entity recognition, NER, NLP, NLTK, OAI-PMH, OCR, Open Archives Initiative, parser, part of speech, Partitur Editor, personal name, phonetic alphabet, PHP, pitch analysis, Praat, python, quality management, R, RDF, Regular Expressions, Reguläre Ausdrücke, relation, resource description framework, Rhetorical Structure Theory, RST, scientific writing, sed, segmentation, segmentation errors, Segmentieren, Segmentierung, sentence splitting, sound, spectral analysis, Splitting, Spoken Language, Stanford, statistical analysis, statistical machine translation, statistics, Statistik, Struktur, STTS, Stuttgart, Support, tagger, tagging, tagset, teaching, TEI, teilen, TeLeMaCo, test theory, text encoding, Textile, Textlinguistik, TIGERSearch, TIGERXML, tokenization, tools, topic model, Transcription, Transkription, treebank, Tübingen, typesetting, typography, UAM corpus tool, UNIX, utterance list, verbinden, Video, video annotation, Visualisierung, W3C, WebAnno, WebLicht, WebMAUS, Weka, Windows, word cloud, written data, XML, XPath,

Figure 3: Available keywords in the browse interface (as of 2015-03-09)

The browse page⁴ allows the user to select a keyword or an author. A list of known keywords is found in Figure 3.

4 Adding material to TeLeMaCo

Adding material to TeLeMaCo is easy and should not take longer than five minutes. The following fields can be filled (only two of them—marked with a star—are compulsory):

Title* The title of the material. This field must be filled.

Keywords Some keywords describing the resource. The user can chose the keywords freely. There can be any number of keywords.

Author The author(s) of the material. Note, that you can add materials to the portal that other people have created.

Language The language in which the material is written. We offer currently a maintained list of admissible languages.

Institution The institution that makes the material available. The portal uses this information to create a $\text{BIB}\text{T}\text{E}\text{X}$ entry for the material.

Year The year when the material was published.

Objective What can be learned from the material?

Audience Who are the intended users for the described material?

⁴<http://fedoara.clarin-d.uni-saarland.de/hub/browse/>

input field	Dublin Core term	notes
access	accessRights	
BIBTEX entry	bibliographicCitation	generated by TeLeMaCo
author	contributor	
institution	creator	
media	format	recommendation: MIME type
	identifier	generated by TeLeMaCo
language	language	
url	relation	
licence	rights	
keywords	subject	
title	title	
type	type	we don't use the DCMI Type Vocabulary

Table 1: Correspondence between input fields and Dublin Core terms.

Prerequisites What is needed to make use of the material? It is possible to link to other descriptions in the portal using the Textile⁵ markup language.

Level The level of the material, described by a closed vocabulary containing the values “not specified”, “basic”, “intermediate”, “advanced”, and “expert”.

Type The type or genre of the material. The values can be freely chosen by the user, popular choices include quickstart and tutorial.

Media The media of the material, given as a MIME type.

URL* The location where the material can be accessed. This field must be filled.

Licence The licence under which the material can be reused.

Access There are only two admissible values: “open” and “academic”. We do not support commercial items.

Description A free text describing the material.

BIBTEX type A type for creating the BIBTEX entry for the material, usually misc. It is possible to manually post-edit the generated BIBTEX entry.

5 Metadata

Most of the input fields of TeLeMaCo have a natural correspondence to Dublin Core terms (DCMI Usage Board, 2012), as detailed in Table 1. Some more Dublin Core concepts are automatically filled by TeLeMaCo; this includes bibliographicCitation and identifier. For the Dublin Core term type, DCMI suggests a coarse grained controlled vocabulary. We don't follow this suggestion. Instead, we allow the users to enter a type at their deliberation, frequently used types include *Tutorial* and *Quickstart*.

For the pedagogical metadata we consider a mapping to the concepts of the Learning Resource Metadata Initiative (LRMI) (Association of Educational Publishers and Creative Commons, 2014).

We have a preliminary implementation⁶ of CMDI (Broeder et al., 2011) metadata for the materials registered in TeLeMaCo. We plan to make these metadata harvestable via an OAI-PHM interface.

⁵<http://txstyle.org/>

⁶See, e.g., <https://fedora.clarin-d.uni-saarland.de/hub/cmd/313/>

6 Additional benefits

The contents of TeLeMaCo are crawled and indexed by the big search engines (Google, MSN, Yahoo, Yandex, Baidu). This has two effects to materials added to TeLeMaCo: Some users will find the TeLeMaCo display page in the search engine of their choice and go on to the wanted material, and the page rank of the original page is boosted (leading users directly from the search engine to the material).

The display page of a teaching or learning material⁷ has internal weblinks to the authors and keywords weaving a web of related resources. This allows the user to navigate to other materials for the same tool, the same task, or by the same author.

7 Evaluation

After a phase of internal testing within the CLARIN-D project, TeLeMaCo went public in September 2013 and was announced at GSCL 2013 (Amoia et al., 2013). Since then, a steady trickle of descriptions has been added to TeLeMaCo, now holding a total of 145 materials. Most of the contributions still come from members of the CLARIN-D project, but we start seeing submissions from other places, too.

We see in the logs that users from all over the world start consulting TeLeMaCo for teaching and learning materials. It was a surprise for us to see calls to a specific description directly without prior searching or browsing. These hits are coming from users being directed to TeLeMaCo by a search engine.

The assignment of keywords by the contributors works reasonably well as can be seen in Figure 3. The autosuggest feature helps in selecting already existing keywords. Some contributors have chosen German keywords for materials in German language. We currently do not attempt to merge the different languages.

Since September 2014 TeLeMaCo is listed in the large directory at LINSE (*Linguistik-Server Essen*)⁸. LINSE is a German language portal to all kinds of materials and services around linguistics.

8 Comparison with other services

We are aware of two collections of teaching resources for the documentation of endangered languages. The E-MELD School of Best Practices⁹ is a project supported by the LINGUIST list. Resources are added by the project team, and although there was little activity since 2007 the project is still alive. There are short descriptions of the materials in free text format and there is a search function.

The Resource Network for Language Documentation (RNLD)¹⁰ hosts a more up-to-date list of resources for language documentation. Materials are described in free text format. They provide a full text search over the whole website.

The project DARIAH-DE¹¹ has launched a service called *Schulungsmaterial-Sammlung*¹² in July 2014. The target group are Digital Humanities. The interface to this service is in German, materials are added by the staff members only. The materials have short textual descriptions and come with the following annotations: institution, media, title, tools, didactic type, discipline, language, date, keywords (up to three) chosen from the closed TaDiRAH (Perkins et al., 2014) vocabulary, and licence. There is a full text search over all the fields available.

9 Conclusion

We think that TeLeMaCo fills a gap in the existing ecosystem of language infrastructures by providing easy and quick access to teaching and learning materials. Descriptions of materials can be provided by everyone after registration at the service, avoiding a bottleneck in extending the service.

⁷e.g., <https://fedora.clarin-d.uni-saarland.de/hub/resource/313/>

⁸<http://www.linse.uni-due.de>

⁹<http://emeld.org/school/index.html>

¹⁰<http://www.rnld.org/resources>

¹¹<http://www.dariah-de.eu>

¹²<https://de.dariah.eu/schulungsmaterial>

TeLeMaCo provides structured metadata of the resources that can be further integrated in the CLARIN infrastructure.

Both tools and available documentations benefit from being added to TeLeMaCo. They are not only findable through TeLeMaCo itself, but also their visibility in search engines is improved.

References

- Marilisa Amoia, Hannah Kermes, Jörg Knappen, José Manuel Martínez Martínez, Elke Teich, and Mihaela Vela. 2013. TeLeMaCo—a collaborative repository for training and teaching materials in linguistics. In *Proceedings of the International Conference of the German Society for Computational Linguistics and Language Technology (GSCL 2013)*, Darmstadt, Germany.
- Association of Educational Publishers and Creative Commons. 2014. LRMI Specification Version 1.1. <http://dublincore.org/dcx/lrmi-terms/1.1/2014-10-24/>.
- Daan Broeder, Oliver Schonefeld, Thorsten Trippel, Dieter Van Uytvanck, and Andreas Witt. 2011. A pragmatic approach to XML interoperability—the component metadata infrastructure (CMDI). In *Balisage: The Markup Conference 2011*, volume 7.
- DCMI Usage Board. 2012. DCMI Metadata Terms. <http://dublincore.org/documents/2012/06/14/dcmi-terms/>.
- Marie Hinrichs, Thomas Zastrow, and Erhard Hinrichs. 2010. Weblicht: Web-based LRT services in a distributed escience infrastructure. In *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)*, Valetta, Malta.
- Jody Perkins, Quinn Dombrowski, Luise Borek, and Christof Schöch. 2014. Building bridges to the future of a distributed network: From DiRT categories to TaDiRAH, a methods taxonomy for digital humanities. In *International Conference on Dublin Core and Metadata Applications 2014*, pages 181–183.
- Thomas Schmidt and Kai Wörner. 2009. EXMARaLDA – Creating, analysing and sharing spoken language corpora for pragmatic research. *Pragmatics*, 19(4):565–582.
- Dieter Van Uytvanck, Herman Stehouwer, and Lari Lampen. 2012. Semantic metadata mapping in practice: the virtual language observatory. In *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC 2012)*, Istanbul, Turkey.
- Thomas Vander Wal. 2007. Folksonomy Coinage and Definition. <http://www.vanderwal.net/folksonomy.html>.