

# TEI-based Interactive Critical Editions\*

Simon Schiff<sup>1</sup>, Sylvia Melzer<sup>2</sup>, and Ralf Möller<sup>1</sup>

<sup>1</sup>University of Lübeck, Institute of Information Systems ,  
Ratzeburger Allee 160, 23562 Lübeck, Germany ,  
{schiff,moeller}@ifis.uni-luebeck.de  
<sup>2</sup>Universität Hamburg, Centre for the Study of Manuscript  
Cultures , Warburgstraße 26, 20354 Hamburg, Germany ,  
{sylvia.melzer,eva.wilden}@uni-hamburg.de

## Abstract

A critical edition is the reconstitution of a text based on a survey of the available witnesses (manuscripts and quotations), resulting in a text and all its attested variants. It is usually created with high effort by scholars in the humanities, possibly separated by chronological or geographical boundaries, over several years. During the editing process, scholars in the humanities prefer to work with any tools and documents in any format they are familiar with. Independently of any boundary or tool in use, one primary goal usually is to produce a traditional print edition. However, working towards such a print edition is very time-consuming, and we argue that there is potential for reducing the time required without the need to forgo preferred tools and methods. Our contribution consists of providing different interactive views of the data as an additional offer to perform analyses. We enable scholars in the humanities to easily transform documents into the well-known TEI format and to store them in a database to create services on demand. The database management system Heurist, a tool for the humanities, allows for searching, automatic linking parts of the documents, interaction with the documents, and exporting the documents into a printable version of the edition. Our approach is validated by direct collaboration with scholars who are currently working on a critical edition. Our solution allows these scholars to efficiently work on a critical edition, independent of chronological or geographical borders, while they still use their preferred tools and document formats.

---

\*The research for the extended abstract was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – EXC 2176 'Understanding Written Artefacts: Material, Interaction and Transmission in Manuscript Cultures', project no. 390893796.

# 1 Introduction

A critical edition is the reconstruction of a text including all witnesses. It is completely as possible and usually created by humanities scholars over several years. They work with their preferred tools and document formats, possibly separated by chronological and geographical borders. Usually, the goal is to have at the end an edition in a printable format, which is preferred by most of the readers. Due to the different tools, formats and borders, some tasks involved in the creation of the critical edition are time-consuming and cost-intensive. For instance, tasks such as transcribing texts from pictures of manuscripts are very time-consuming if humanities scholars have to merge their transcribed texts of any format, produced by different tools, into one that could be printed. However, there is a growing interest in having access to digital editions, in contrast to a printed edition. By digital editions, we do not refer to, for instance, scans of printed editions. Instead, we refer to digital editions that follow some principles, such as the FAIR (**F**indable, **A**ccessible, **I**nteroperable, **R**euse) principles. Creating a digital output in addition to the print format output is additional costly work that must be done either manually by humanities researchers or by an IT expert who processes the documents semi-automatically.

Our objective is to enable humanities scholars to work with their preferred tools and document formats across chronological and geographical borders and yet to achieve their desired results. We aim to achieve that by providing a web application as an add-on in addition to the tools that are already in use. The web application should only be an add-on, as we aim to support humanities scholars without trying to change their preferred tools and document formats in use. The web application allows humanities scholars to transform documents, in any format, into the well-known text encoding initiative (TEI) format. Documents transformed into the TEI format are machine-readable and thus are combined optionally and then exported into any desired format to be printed or published online. In addition, a database is created on demand, that allows for simple search queries as well as classification of an object by assignment of several independent terms (faceted searches) to analyze data. A database would not only represent the data for one text, but would allow the data to be analyzed as a whole in large quantities in less time.

The database management tool Heurist [9] offers a management for access rights, an environment for sharing and collaborating online with other users, and modification of the database scheme without programming.

We give in Section 2 a detailed overview of critical editions, including their creation, use and present why it is challenging but useful to publish them online, following the FAIR principles. The related work in Section 3 gives an insight into current research and tools. In Section 4, we present the work by Eva Wilden who is a leading expert in classical Tamil literature and poetics, and currently working on the critical edition of a Tamil poetic anthology on the basis of 18 manuscripts. The preferred tool for producing a critical edition in a printable format is Microsoft Word. We worked together across disciplines for testing our developed web application using Microsoft Word DOCX documents as input.

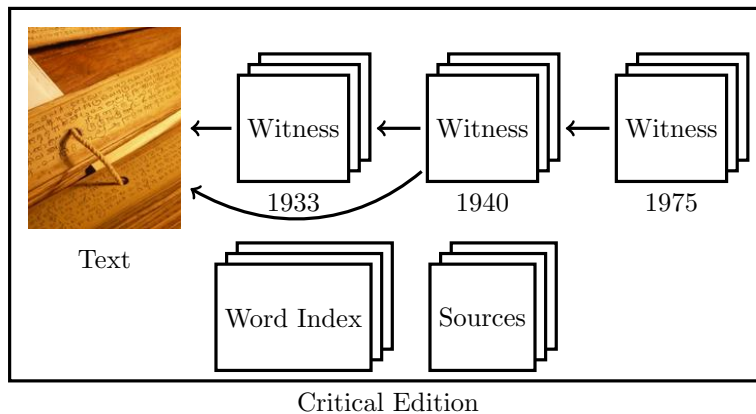


Figure 1: Contents of a Critical Edition

We present our web application step by step in the sections following.

First, in Section 5, we present how to enable humanities researchers to transform documents from their preferred format into TEI without the help of an IT expert. Documents encoded in TEI are designed to be machine-readable and not human-readable, therefore we present in Section 6 a viewer, part of our web application, for documents that stem from our transformation of documents into TEI. Depending on the context, one possibly needs a database, which we create on demand from the TEI documents, for a more extensive data analysis. The databasing on demand process is described in Section 7. TEI documents visualized by our web application are combined with linked sources such as images, audio, or other websites. As a use case, we present in Section 8 our annotation system, which allows humanities scholars to annotate images that are linked and visualized by our viewer. In addition, the annotation system provides a login for access control and a chat below each annotation for remarks. Finally, we present the use of our web application in Section 9, briefly summarize our paper in Section 10, and give an outlook for future work. Among other things, we aim to extend our web application such that it could be adapted by humanities scholars, without the help of an IT expert.

## 2 Preliminaries

A critical edition is a scholarly attempt at reconstructing the transmission history of a text by following up the various textual witnesses over the centuries, and thus reproducing the trajectory the text made through various manuscript and print versions into the modern days, as depicted in Figure 1.

The task comprises the collection and decipherment of a set of sources in manuscript and print as completely as possible, a comparison of the differences between these sources, and their documentation. This is achieved by the con-

stitution of a core text and a critical apparatus that records the variants and distinguishes what is meaningful from what is lost, damaged or simply erroneous. Additionally, critical editions contain word indexes with morphological and semantic analysis of the words and their occurrences in the texts. The advantage of this type of edition is that its author presents a hypothesis about the constitution of the text while at the same time giving to his or her potential readers all the elements necessary to revise or even reject it and form conclusions of their own.

Readers often prefer critical editions in printed format. However, there is a growing interest of critical editions in a digital format [1]. Having a critical edition in digital format alone is not sufficient and one should follow specific principles such as the FAIR principles.

Editions are easier to find by using web search engines if they are digitally available online. An author of an edition that is easy to find has a higher chance of being cited than others. Depending on the format of the edition and the viewer that visualizes the contents of the edition the implementation of a faceted search engine helps to satisfy the information needs of various readers. Following links by clicking on them to find linked elements is not as time-consuming as going through a book by hand. One might need to pay a lot to access a complete edition in printed form and only some parts, such as pictures of manuscripts, are the reason for high prices. If an edition is digitized, one has a more fine-grained control as to which parts can be accessed by whom. In particular, specific parts can be made publicly available. Results are combined with other results automatically on demand, if the edition is in a machine-processable format (cf. Section 7). In addition, editions in such a format are reused for various kind of research purposes, such as in the field of artificial intelligence (AI) [2].

Following the FAIR principles during the creation of a critical edition is a challenging task, as even the creation of a critical edition in any format is a cost-intensive process lasting several years. Additionally, many humanities scholars who work on a critical edition are possibly separated by chronological and geographical boundaries and with preferences for different tools and document formats. We argue that the solution is not to intervene in the processes, tools, and document formats humanities scholars prefer to work with. Instead, we provide a web application that allows humanities scholars to upload their documents of any format and transform them to TEI, which is a common and standardized format in the digital humanities. As depicted in Figure 2, we create Antlr4 [12] grammars for the generation of parsers for parsing different document formats, such as Microsoft Word DOCX, Markdown, or TXT, where parts of the documents are written in a specific syntax we know beforehand. Antlr4 is a tool that generates the source code for a parser from a grammar. The source code provides an application programming interface (API) to access parts of the texts, depending on the grammar. Parsed documents are transformed into TEI and stored in a repository. The documents can be linked with external sources such as images, audio, or websites and exported into any format, visualized with a viewer, accessed using for instance RESTful or a webservice, and

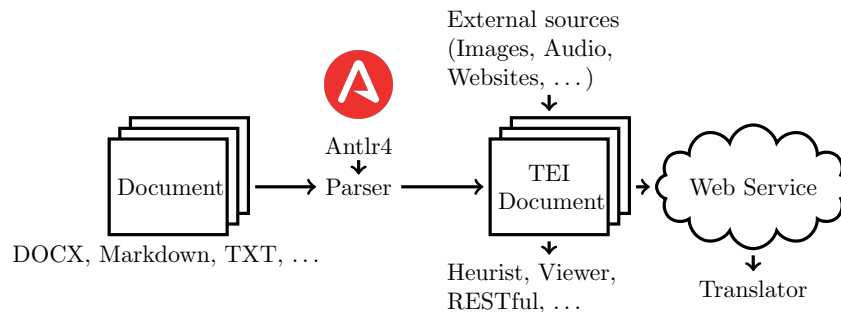


Figure 2: Processing of Word Documents

exported into a Heurist database instance. RESTful provides an API for tools to access data over the web.

Heurist [9] is an open-source web-based database management system for the humanities and allows researchers without prior IT knowledge to develop data models, store and search their data, and publish it on an automatically-generated website. A further highlight of Heurist is the possibility of displaying the data according to the WYSIWYG (What You See Is What You Get) principle which plays an important role in the humanities. The combination of, on the one hand, storing the data and, on the other hand, representing it in a humanities-friendly way with a powerful search function for extensive data analysis with large amounts of data argues for the use of a database. If the data is already available, it can be manually inserted into a database. However, with an increasing amount of data, it is advisable to carry out this time-consuming work with the help of a computer. Heurist has a CSV import function that could be used to transform data into the CSV format. If the data is in TEI, the EpiDoc extensible stylesheet language transformations (XSLT) stylesheets [4] can be used to convert the TEI files to CSV. When using these stylesheets, the text in the CSV file is stored in HTML encoded form. These stylesheets also can be easily customized. Importing such a CSV file displays the text as in HTML encoded on the web page generated with Heurist. The transformation process is either manual or automatic. The automatic process is called databasing on demand.

### 3 Related Work

In the field of manuscript research, the amount of data has grown significantly, however at present it cannot be taken into account by search engines in part because the data is available only in printed form. The evaluation is therefore mostly done manually or with a high resource effort by implementing customized algorithms. However, existing algorithms and tools can easily be made to use more powerful, customized applications that can be derived from existing ap-

proaches through the systematic harmonization of processes and tools.

For proteomics researchers, the creation of a custom database was made as easy as possible by developing a so-called Database on Demand (DoD) [13]. The idea of DoD is to generate a custom database according to a set of user-selectable criteria using commonly used database templates as source. The database PRIDE [14] is one DoD to provide users to build a database in five steps. However, the DoD in this case only works for data for one - the proteomics - area. In the field of e.g. the Humanities, this concept of a DoD could also be useful. With DOD's approach, it is then possible not only to simply implement an information system such as presented in [10], but also to link these into a network in such a way that even simple, federated, or more complex searches are possible [11]. When presenting the data from the humanities scholars, it is still important to note that text formatting is also a constraint and very important for the humanities scholars. Text formatting is explicitly supported with the TEI format. Since TEI is a machine format that has yet to be converted into a human-readable format. Efforts have been made to implement TEI viewers such as the Edition Visualization Technology (ETV)<sup>1</sup>, but these are only usable for very specific TEI schemes. For example, an viewer was developed for epigraphers, which requires a slimmed-down TEI variant, the EpiDoc [5] scheme, as input. While the texts are correctly displayed according to appropriate conditions, e.g. the Leiden conventions [6], the tool called EpiDoc Front-End Services (EFES) [8] must be installed locally before a data set can be viewed. With EFES the texts are displayed, but only one text at a time. An evaluation of all texts or other data is not possible. Therefore, an approach that is based on the combination of a database and the corresponding formatting of the texts is highly recommended when it comes to efficiently carrying out an analysis with the existing data. In addition, the storage of data also has the advantage here of being FAIR (**F**indable, **A**ccessible, **I**nteroperable, **R**eusable) from an ethical point of view [7] and that of good research practice [3].

## 4 Critical Edition - Critical Texts of Cankam Literature

The point of departure for the current experiment is a traditional critical edition of average complexity, from one of the smaller disciplines within the humanities, namely Indology, or, to be more precise, Tamilistics. Tamil is the language of classical literature in South India; the text selected is an anthology of love poetry that dates back roughly to the beginning of the Christian era. The anthology consists of 400 poems between 13 and 33 lines of length, accompanied by miniature commentaries for each piece and a more detailed old anonymous commentary for the first 90 poems. The part already completed (and available in print) comprises book 1, that is, the first 120 poems plus the old commentary [15]. The critical edition is based on eighteen manuscripts (palm-leaf and paper),

---

<sup>1</sup><http://evt.labcd.unipi.it/>

three editions from the early and middle 20th century and numerous quotations from the commentaries of an extensive tradition of poetics. The edition consists of a metrical Tamil text, a word-split transcription, a critical apparatus, an interlinear English version and an English translation, with heavy annotation for every layer. On the basis of the text and the variants reported in the critical apparatus a word index cum concordance for every word and all its derivations is produced which also contains an analysis and a translation as well as references to the existing lexical works. At the basis of such a work is a complex web of cross-references on various levels. The Tamil text itself needs to be seen in an intertextual context of parallels within the corpus of literary texts; many of its elements reoccur in the commentary. The critical apparatus relates directly to the manuscript and print sources that contain the corresponding variants. The word-split transcript separates metrical items into lexical items, the interlinear version adds their morphological and semantic analysis. The translation adds an interpretation of the whole both in terms of syntax and in terms of a piece of poetry that follows a set of conventions, has parallels and has been a long-term subject of scholarly discourse. The annotation adds direct references to parallels and ongoing scholarly discussions. The word-index contains a detailed morphological and semantic analysis of each item in relation to various external tools such as dictionaries, grammars, commentaries and scholarly articles. In book format this work consists of three volumes that, including the bibliography, make up a total of about 1100 pages. The tool that was used to produce this edition, in DOCX format, was simply Microsoft Word.

## 5 Transforming Documents into TEI

Editions contain natural language text paragraphs and sections written in a specific syntax. The latter are, for example, translations of Tamil poems with footnotes or word-indexes with associated information including occurrences of the words in the poems and external sources such as dictionaries. A word-index entry of the word *acaitta* is depicted in Figure 4 on the left hand side. The associated information of the word *acaitta* is separated by a tab, including the information *pey. p.a.* and the occurrences *155.14* and *301.22* meaning that the word *acaitta* occurs at poem 155 in line 14 and poem 301 in line 22. A reader, who has access to a printed edition or digital edition violating the FAIR principles, has to search for the poems 155 and 301 manually, if he or she wants to gather further information about the context, in which the word *acaitta* is being used. A website could help a reader to reduce the time required to find specific information, by providing links for each occurrence of the words in the indexes, such that the reader can click on them to be redirected to the corresponding poems. As depicted in Figure 3, the word *acai* in the word-index on the right hand side of the picture appears in line 5 of poem 40 on the left hand side. In our work, we focus on processing Microsoft Word DOCX documents, as we collaboratively work together with Eva Wilden, who is using primarily Microsoft Word. Direct accessing translations or word-indexes in the

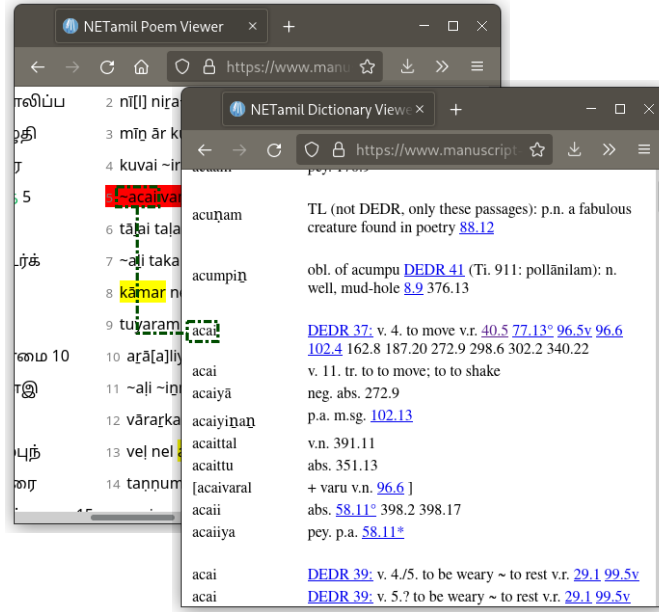


Figure 3: Poem linked with word-index

editions written in DOCX format, for implementing a website is impossible. Therefore, we provide a web application, implemented in Java, for uploading DOCX documents by humanities scholars as a service, as depicted in Figure 2. The application analyzes the uploaded documents and searches for parts that are written in a specific syntax, such as word indexes. If a word index is found, it is processed and a web service is created, access is via the internet to implement a website.

More specifically, a DOCX document is a zip archive standardized as Office Open XML (OOXML), containing among other files a document.xml file. We created a parser, using Antlr4 [12], for parsing and processing the contents of the document.xml file. The result is a document in TXT format containing the contents of the DOCX document without any XML tags, with annotated parts such as sections or tables. The TXT document is again parsed with a parser, we have generated, using Antlr4. A parser is generated, as a Java code snippet, by Antlr4 from a grammar, as depicted in Figure 4 on the right hand side. The grammar consists of four rules and in this case, the rule *entry* matches with the vocabulary index on the left hand side that is a sequence of tokens *acaitta* (text), a tab (tab), *pey. p.a.* (text), *155.14* (occurrence), and *301.22* (occurrence). Each token matches with the corresponding rules *text*, *tab*, and *occurrence*. If the syntax of the documents changes, the grammar can be easily adapted, without the necessity to change anything else part of our web application. We instantiate Java objects from the parsed TXT file and deserialize them as TEI documents.



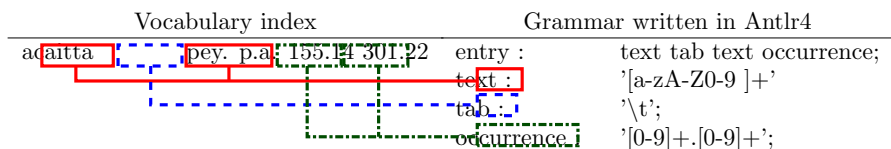


Figure 4: Parsing a Text using Antlr4

A webservice could be created on demand, using the web services description language (WSDL), that accesses the created TEI documents.

## 6 Interactive Critical Editions

As depicted in Figure 2 and described in Section 5, our web application supports the work of a humanities scholar by converting his Word documents into TEI when parts of the document are written in a certain syntax. A TEI document, that stems from the transformation, is not very pleasant to read for humans, but in a well known scheme readable by our web applications or in general by a machine. Therefore, we create viewer for all kind of TEI documents, our web application creates, using HTML5 together with JavaScript and CSS. The viewer has several advantages over a document written in Word or a printed book. For instance, the viewer allows for (i) a search engine, (ii) linking of various elements such as words in word indexes with their poems at where they appear, (iii) restricting access to only authorized humans, (iv) export the visualized TEI document into various formats, and (v) merging of TEI documents with other TEI documents or other resources such as images mentioned in the documents. The latter can save humanities scholars weeks or even month of work. For instance, as depicted in Figure 5, two vocabulary indexes need to be merged. Vocabulary index I and II are tiny excerpts of two very large word indexes, containing thousands of words that need to be merged. Doing that by hand would take several weeks of work and is a very laborious and cost-intensive task. Instead, our web application allows for merging the indexes automatically on demand. The merged vocabulary index is again a TEI document, that could be exported into various formats or visualized by our viewer. A humanities researcher does not need to search for the same word in two different documents anymore.

## 7 Databasing on Demand

The advent of more flexible and precise tools in recent years has enabled the way to evaluate critical editions as well as new methods for manuscript evaluation, thus expanding the field of the humanities. In addition, a large amount of data accumulates over time, which can only be evaluated by machines in a short time and are available to be included in the analysis processes.

Vocabulary index I		Vocabulary index II		I & II merged	
acai	DEDR 37: v. 4. to move - v.r. 40.5 77.13° 96.5v	acai	DEDR 37: v. 4. to move - v.r. 77.13° 187.20	acai	DEDR 37: v. 4. to move - v.r. 40.5 77.13° 96.5v 187.20
acaiyā	neg. abs. 272.9	acai	v. 11. tr. to move; to shake	acaiyā	neg. abs. 272.9
				acai	v. 11. tr. to move; to shake

Vocabulary index I		Vocabulary index II		I & II merged	
acai	DEDR 43: v. 11. to tie – v.r. 54.7	acai	DEDR 43: v. 11. to tie	acai	DEDR 43: v. 11. to tie – v.r. 54.7
acaii	abs. 188.12*	acaitta	pey. p.a. 155.14 301.22	acaii	abs. 188.12*
				acaitta	pey. p.a. 155.14 301.22

Figure 5: Merging of word indexes (Index I: blue, Index II: yellow, and I & II merged: blue and yellow as green)

Databases are ideal for storing a large amount of data and, depending on the database query language, the available data format and the tool, allow simple to more complex search queries to be performed. In addition, database management systems such as Heurist also support the creation of websites, including a search mask, display of search results, and presentation of data, in short time.

To make the creation of project-specific, autonomous databases as easy as possible, we have developed a databasing on demand approach so that users may create customized databases automatically. Therefore we use Heurist to create database instances.<sup>2</sup>

The precondition for the databasing on demand approach is that the poems, dictionaries, and comments are stored in TEI or CSV format. Otherwise, the transformation from Word to TEI, as described in section 5, must be performed first. Then, a TEI-specific database scheme must be created (either manually or by creating a template, which is then reusable). The databasing on demand process encompasses the three steps 1. transforming from TEI to CSV based on the extended EpiDoc XSLT Stylesheets to parse the TEI files and create the particular data view, 2. importing data into the database, and then 3. publishing data, respectively.

1. **Transform from TEI to CSV:** Input data are TEI (poems, commentaries) files. Since there are very different possibilities to structure such

<sup>2</sup>[https://heurist.fdm.uni-hamburg.de/html/heurist/?db=CSMC\\_UWA\\_NETamil2](https://heurist.fdm.uni-hamburg.de/html/heurist/?db=CSMC_UWA_NETamil2) (internal access)

a scheme. It is recommended to follow the TEI project example. We have written java program to process the transformation from all TEI files (poems, commentaries) to one CSV file.

2. **Import data into the database:** The created CSV file from the previous step is the input file. The CSV file was imported into the database. When importing, the user must make the mapping from the columns in the table and the fields in the database manually.
3. **Publishing data:** During or after the importing process, a web page, see Figure 6, can be created easily and in short time.

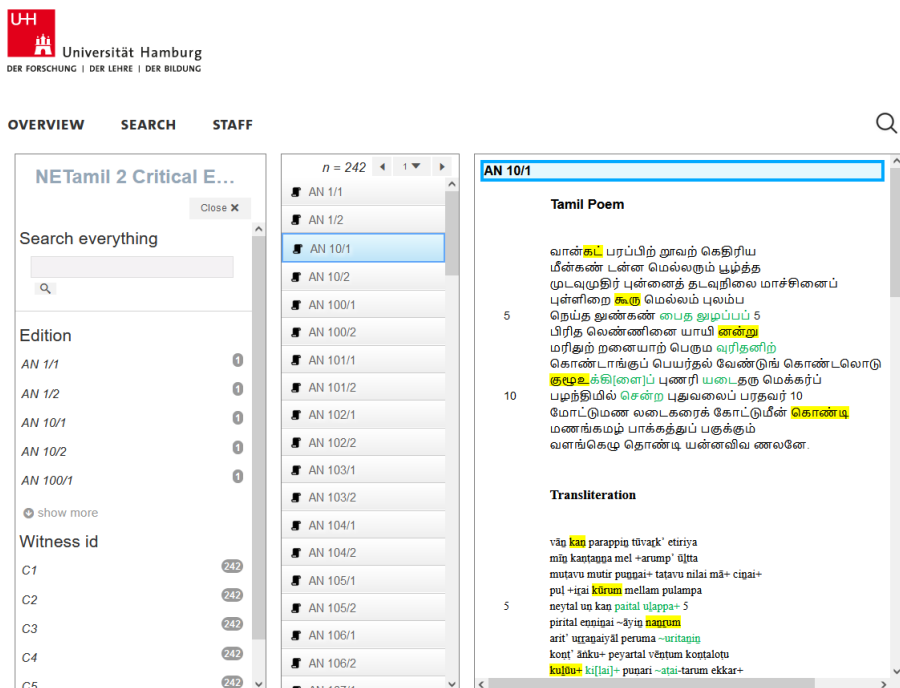


Figure 6: Website created with Heurist; left: search area, middle: result set, right: data representation

Even though Heurist offers a kind of drag and drop functionality for creating a webpage, the user needs knowledge of the automatically generated source code (PHP and HTML) so that project-specific adjustments can be made. The usage of some database functionalities requires a certain know-how, so that a close cooperation between humanities scholars and computer scientists is very advantageous and recommended.

We used the above three steps for poems and comments. For creating links to other dictionaries, such as the second Dravidian Etymological Dictionary,

they were automatically created in the PHP script for the users' view. However, creating links automatically to the poems in the database require another approach. For this, a database query was made to obtain the database IDs of the poems. A Java program was then written which automatically creates a link to the database entry for the reference to a poem (see Figure 7). As a result, users can use the created information system<sup>3</sup> as an interactive version of poems, dictionaries, and commentaries.

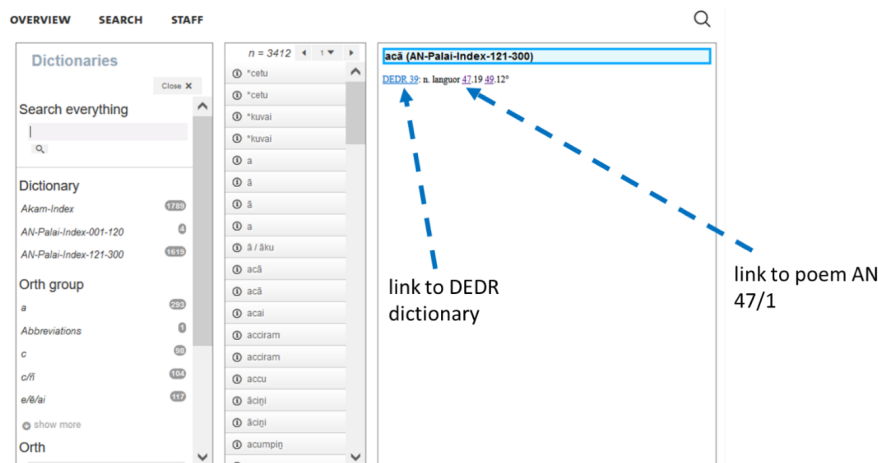


Figure 7: Dictionary with links to the DEDR dictionary and poems; left: search area, middle: result set, right: data representation with links to other dictionaries and poems

## 8 An Annotation System for the Humanities

As depicted in Figure 2 and presented in Section 6 we created a viewer for TEI documents, that contain for instance poems written in Tamil, that stem from the transformation of DOCX documents into TEI. The poems were transcribed from palm-leaf manuscripts and editions that were photographed and stored as images in a database. We add the images to the viewer for those who are authorized to access them. Images of manuscripts and editions are associated with the images, as the DOCX documents, at where the poems stem from, contain the names of the images. A humanities scholar may want to note in his or her texts that a specific region in an image is damaged by a tape worm. One way of doing that is to add a description to the text, that contains the name of the image and a description at where the damaged region is located at. A description could contain for instance the coordinates of the region in

<sup>3</sup>[https://heuristic.fdm.uni-hamburg.de/html/heuristic/?db=CSMC\\_UWA\\_NETamil2&website&id=981&pageid=976](https://heuristic.fdm.uni-hamburg.de/html/heuristic/?db=CSMC_UWA_NETamil2&website&id=981&pageid=976) (internal access)

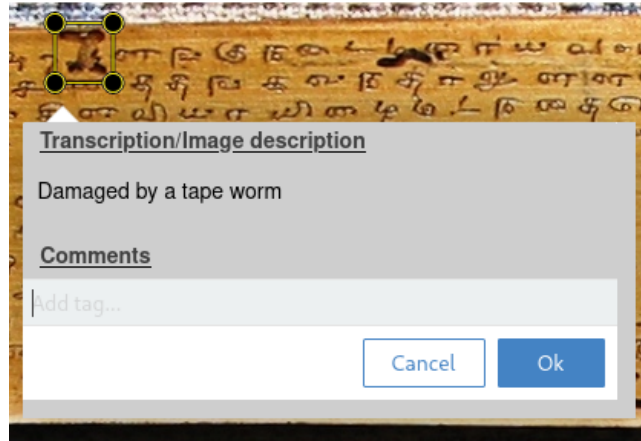


Figure 8: Annotation of a palm-leave manuscript

pixel units. If another scholar reads the text, then he or she needs to search for the image and if the image is found, for the correct position in the image. We follow another approach, at where the humanities scholars can interactively add annotations to the images, as depicted in Figure 8. For the annotation of the images, we use Annotorious<sup>4</sup>, which is open-source and implemented in JavaScript. As Annotorious is written in JavaScript, it runs only on client side. However, annotations are serializable into JSON, we send to the server at where the viewer and Heurist are hosted at. Annotations are stored in the Heurist database and were exported together with the images and contents of the TEI documents into any format.

## 9 Application and Results

From the point of view of a scholar in the humanities, the approach described here promises a number of advantages, both with respect to ongoing as with respect to already accomplished work. Since nowadays manuscripts are no longer documents that are stored in a library, that can be consulted, often with great effort, and exploited for text edition, but are available or can be made available online, it is very tempting to complement a printed critical edition by an electronic one. This way it is possible to accompany the critical apparatus by direct access to all the (digitized) sources, a direct view of the manuscript images and the variants they attest. The advantages for ongoing text-critical work and continued improvements on the hands of the research community, no longer a single scholar, are obvious. Similarly, the direct interlinking of edition, word indexes and existing dictionaries simplifies the necessary lexical work and allows

<sup>4</sup><https://github.com/recogito/annotorious>

an efficient use of work already done in the same area. An additional benefit is the possible use of such a construct in teaching: direct access to all the relevant material for students, and even the possibility to make their own contributions. Attractive is also the user-friendliness of the conversion procedure which does not require advanced technical skills in IT. Many scholars prefer to continue working with a simple word-processing program, all the more when they have huge amounts of data which have already been produced in such formats. The same holds good for cases of collaboration with colleagues in technically less advanced areas.

## 10 Conclusion and Future Work

The presentation of data and the results of analysis plays an important role in the humanities. Tools such as Word that support the WYSIWYG (What You See Is What You Get) principle are often used. For this purpose, humanities scholars mainly use Word to present their research data such as critical editions. In addition, printed versions of critical editions are mainly required as a project result. Individually structured documentation and print versions of research data complicates automatic data exchange and linkage to other (similar) research data. Increasingly, however, humanities scholars' research data are being archived in widely used standards such as TEI. Since TEI is a machine-readable format and we want to support humanities scholars to continue using their established processes and tools, we have shown in this paper how to create a transformation from Word to TEI. Furthermore, we have shown how to use this standard to perform a databasing on demand to support the analysis process by performing a more advanced search over a large amount of data in a short time. We also presented an annotation system for adding annotations to a palm leaf, for example. The annotations were stored in a database and used for further analysis. All our approaches satisfy the FAIR principles that are elementary to good scientific practice in particular.

As future work, researchers from other fields are also supported by these approaches to represent critical editions and work with the new services. We are also working on merging the databases from different projects into one overarching information system (see [11]) so that, for example, federated searching is supported. The database of Tamil poems, dictionaries, and commentaries generated by databasing on demand could be integrated as part of the federated database system in a next step. From a data linking point of view, it is possible to include dynamic links in addition to the static links in order to enrich the existing information with further information.

## References

- [1] Mohammad Khaled Al-Shboul and Abdullah Abrizah. Information Needs: Developing Personas of Humanities Scholars. *The Journal of Academic*

*Librarianship*, 40(5):500–509, 2014.

- [2] Magnus Bender, Tanya Braun, Marcel Gehrke, Felix Kuhr, Ralf Möller, and Simon Schiff. Identifying and Translating Subjective Content Descriptions Among Texts. *International Journal of Semantic Computing*, 15(4):461–485, 2021.
- [3] DFG. Guidelines for Safeguarding Good Research Practice. [https://www.dfg.de/download/pdf/foerderung/rechtliche\\_rahmenbedingungen/gute\\_wissenschaftliche\\_praxis/kodex\\_gwp\\_en.pdf](https://www.dfg.de/download/pdf/foerderung/rechtliche_rahmenbedingungen/gute_wissenschaftliche_praxis/kodex_gwp_en.pdf), 2022. Accessed 17 January 2022.
- [4] Tom Elliott, Zaneta Au, Gabriel Bodard, Hugh Cayless, Carmen Lanz, Faith Lawrence, Scott Vanderbilt, Raffaele Vigiante, and et al. EpiDoc Reference Stylesheets (version 9). Available: <https://sourceforge.net/p/epidoc/wiki/Stylesheets/>, 2008-2017. Accessed January 22, 2022.
- [5] Tom Elliott, Gabriel Bodard, Elli Mylonas, Simona Stoyanova, Charlotte Tupman, Scott Vanderbilt, and et al. EpiDoc Guidelines: Ancient documents in TEI XML (Version 9). Available: <https://epidoc.stoa.org/gl/latest/>, 2007-2022. Accessed January 22, 2022.
- [6] Everipedia. Leiden Conventions. [https://everipedia.org/Leiden\\_Conventions](https://everipedia.org/Leiden_Conventions). Accessed 17 January 2022.
- [7] M. Friedrich, Jost Gippert, Davidson McLaren, Cécile Michel, Ralf Möller, Doreen Schröter, Stefan Thiemann, and Eva Wilden. Ethical and Responsible Research at the CSMC. <https://www.csmc.uni-hamburg.de/files/uwa-csmc-ethics-code.pdf>, 2022. Accessed 17 January 2022.
- [8] GitHub repository. EFES: EpiDoc Front-End Services, 2022. accessed January 22, 2021, <https://github.com/EpiDoc/EFES>.
- [9] HEURIST. A unique solution to the data management needs of Humanities researchers. <https://heuristnetwork.org/>, 2022. Accessed 17 January 2022.
- [10] Sylvia Melzer, Hagen Peukert, Hongxu Wang, and Stefan Thiemann. Model-based Development of a Federated Database Infrastructure to support the Usability of Cross-Domain Information Systems. In *IEEE International Systems Conference (SysCon 2022), Montreal, Canada, April 25 - April 28, 2022*. IEEE, 2022. accepted.
- [11] Sylvia Melzer, Stefan Thiemann, and Ralf Möller. Modeling and Simulating Federated Databases for early Validation of Federated Searches using the Broker-based SysML Toolbox. In *IEEE International Systems Conference (SysCon 2021), Vancouver, BC, Canada, April 15 - May 15, 2021*, pages 1–6. IEEE, 2021.

- [12] Terence Parr. *The Definitive ANTLR 4 Reference*. Pragmatic Bookshelf, 2013.
- [13] Florian Reisinger and Lennart Martens. Database on Demand - An online tool for the custom generation of FASTA-formatted sequence databases. *PROTEOMICS*, 9(18):4421–4424, 2009.
- [14] Juan Antonio Vizcaíno, Florian Reisinger, Richard Côté, and Lennart Martens. *PRIDE and “Database on Demand” as Valuable Tools for Computational Proteomics*, pages 93–105. Humana Press, Totowa, NJ, 2011.
- [15] Eva Wilden. *A Critical Edition and an Annotated Translation of the Akanāṅṅūru: Part 1, Kalirriyāṅainirai. Old commentary on Kalirriyāṅainirai KV - 90, word index of Akanāṅṅūru KV - 120*. École Française d’Extrême-Orient, 2018.