This paper describes the project “From ‘Oral’ Tradition to Hypertext: Changing Media in Rabbinic Literature” of the collaborative research center “Media and Cultural Communication” housed at Cologne University. The project aims to present an example of rabbinic literature (the tractate Megilla from the Babylonian Talmud) in a way that will enable readers unfamiliar with rabbinic writings to understand rabbinic textual and interpretation techniques. An electronic text is employed for this purpose. Therefore, a collaboration with the research center’s project “Impacts of Networked Multimedia Information Systems on Cooperation and the Organization of Knowledge in Cultural Science Projects” run by computer scientists has been set up. This paper describes the outcomes of these joint efforts.

The need for a comprehensive study environment is twofold, both for students of Jewish Studies and for the interested students from other fields. Teaching Talmud at a European university – as well as studying it there – is a challenging task, since students need to understand both the language and the particular style of talmudic discourse in addition to the contents. Classes can either focus on the original Hebrew-Aramaic text, with many explanations of words, phrases and argument structure, or it can use a translation into the students’ native language that mirrors the standards of the original text as closely as possible. Using a translation into a third language – not the student’s native tongue – would inevitably create the need for additional interpretation and engender additional misunderstanding. Furthermore it is often necessary to check parallels of any given talmudic passage against other rabbinic literatures. In order to study Talmud one thus needs a library containing dictionaries, editions (and translations) of the main rabbinic texts, encyclopedias and other resource books. The average European student often has no access to this kind of library while studying talmudic texts. The comprehensive educational environment provided by this project’s electronic medium should (at least partly) substitute for the real library by providing information relevant for proper textual study.

We thank N. Koltun-Fromm for her considerate corrections and language-editing of this paper.


3 It cannot be the objective of any comprehensive electronic study environment to substitute completely for the real library and its manifold functions beyond supplying information, but one might wish for a tool that includes a maximum of information. Cp. software like “BibleWorks 4.0” (BibleWorks™ 1999) including a
In addition to students and scholars of Jewish Studies the Talmud has in recent years found a new audience in scholars of literary studies and philosophy, who – following Emmanuel Levinas and Jacques Derrida – have tried to explore the talmudic discourse. Most of these scholars depend for their studies on translated excerpts that represent but a few talmudic passages out of their literary context. To further interdisciplinary contacts between these scholars and the realm of Jewish Studies it is necessary to provide sizeable parts of the Talmud and its commentaries in translation. These otherwise inaccessible texts transfer knowledge about how the talmudic discourses are structured and what roles are assigned to a given passage within the greater framework of talmudic and rabbinic literature. The needs of these scholars coincide to a great extent with the needs of the students of Jewish Studies.

Our plan to create a hypertext for a talmudic tractate results from the wish to create one study tool for both groups of readers and to let more people understand the difficulties and the beauty of the Talmud.

1. Talmud and Hypertext

Technical Definition of Hypertext

While the idea of a technical implementation of hypertext is quite old\textsuperscript{4}, the idea was renewed by Ted Nelson in the 1960’s and implemented with the hardware available at that time. He coined the term “hypertext” and wrote:

Hypertext is the combination of natural language text with the computer’s capacity for interactive branching, or dynamic display, when explicitly used as a medium. Or, to define it more broadly, ‘hypertext’ is the generic term for any text which cannot be printed (or printed conveniently) on a conventional page, or used conveniently when bound between conventional covers. ‘Non-linear text’ might be a fair approximation.\textsuperscript{5}

The most important concepts of hypertext are therefore non-linearity of text, computer supported links and the interactive way of using hypertext. Thus, 20 years later, Jeff Conklin reduced the technical definition of hypertext to:

The concept of hypertext is quite simple: Windows on the screen are associated with objects in a database, and links are provided between these objects.\textsuperscript{6}


Objects can contain any digital data, text or image. These data-filled objects will be referred to as “hypermedia objects” below.

Talmud and Hypertext

The Babylonian Talmud is the largest collection of texts from rabbinic literature, covering some 5000 folio pages in the standard edition printed by Romm in Vilna (1880-1886). It is also one of the favorite examples of existing hypertexts before hypertext theory. Its traditional layout with the Mishna and the Gemara as the main block on the page and several commentaries grouped around them, first developed by the Christian printer Daniel Bomberg in Venice (1520-23), has invited many interpretations both from scholars of Jewish Studies and from hypertext-specialists and scholars of literary hypertexts. D. Porush notes:

The Talmud is just as likely to “send” the reader to a page elsewhere in the Talmud as to the next page. You can open the Talmud almost anywhere to begin, although standard Talmudic learning progresses, at least at first, in a highly-arbitrary sequence of books. We also see that the Talmud promotes marginalia, scribbling or commentary, and a non-linear non-directed form of knowledge. ... Notes and marks refer the reader to arguments elsewhere in the Talmud. In short, modern computer users will quickly recognize that the Talmud is a hypertext system: a means of gathering clusters of information that is non-linear, promotes interpretation, is multi-vocal (or collaborative in the extreme), tends towards anonymity, is non-directed, packages information in multi-referential but discrete pages, and de-constructs the authority of a single author.

It might be worthwhile to analyze this equation before discussing the possible creating of a “hypertext-Talmud”. The hypertext metaphor can be doubly applied to the Talmud: Both the external structure of the Talmud and its internal structure can be described as hypertext.


The external structure is the combination of texts on the traditional page of Talmud. The main texts on the page are the Mishna, edited in the 2nd century C.E., and the Gemara, that was probably closed in the 8th to 9th century C.E. They are carefully edited compilations that show the traces of several editing processes, each of them with its own agenda for the logical, rhetorical and thematic outline of the texts. The medieval and early modern commentaries included in the printed editions were composed by single authors (with the exception of the Tosafot, whose publishing-history differs from that of all other commentaries) either as commentaries on the completed and closed texts of the Babylonian Talmud or on single tractates, or as supercommentaries on preexisting commentaries. All commentaries take up *lemmata* from the talmudic text and create links into the text, attaching their own explanations. While we read and study the talmudic page by looking for links from the inside to the outside, the commentators actually created links from the “outside” to the “inside”. Most of the older commentaries were not originally written on the margins of copies of the Talmud. They required the use of two manuscripts simultaneously. The page of Talmud therefore fits the description of hypertext: it brings together and links documents in a way that enables the reader to use several connected documents at the same time.

As to the internal structure of the talmudic text, it can be also described as a hypertext because, like all rabbinic texts, it is a unique combination of new and old material. In each text we find a great number of passages that occur in parallel transmissions of textual units copied from their sources and inserted at suitable positions in the target texts. This created an internal hypertext-structure that connected Midrashim and Talmudim via nodes of shared text. Associations and structures were created that connected the different parts of the Talmud and enabled different ways of reading. The Gemara itself is – beyond the linear text – a web of units connected by associations that offer a multitude of possible readings. The Talmud itself is a hypertext-structure. In spite of this it should be noted that the talmudic text is not fragmented in the way modern hypertexts are usually structured.

---

11 Cp. A. Ahrend, *Rashi’s Commentary on tractate Megillah* (hebr), dissertation, Bar Ilan University, 1995, who studied the use of keywords and quoted *lemmata* in order to prove that the Rashi commentary on bMegilla was originally written as separate text outside the Talmud manuscript.
12 With the exception of some ashkenasic manuscripts even late manuscripts of talmudic tractates do not include Rashi on the page.
14 Cp. the card-stack metaphor employed by most early hypertext-systems like Hypercard or KMS.
The idea of the Talmud as a hypertext has invited attempts to render the Talmud as electronic hypertext. The available products focus on the traditional mode of Jewish Talmud study and connect the Talmud passages with their respective commentaries and with passages from halakhic works that deal with the same topic.

We decided to take a different approach, developing an electronic study environment that focuses on the literary characteristics of the Talmud. The hypertext-metaphor of the talmudic text thus required that we recreate the structures and associations as different paths of reading through the electronic text. The hypertext-metaphor of the talmudic page enables us to attach additional material to the talmudic text as a supplement and substitute for the existing commentaries.

2. Developments in Digital Libraries

Jewish Studies was one of the fields where electronic texts became available for scholars rather early in the development of digital libraries. Given the technical developments we can now look back on a history of very different questions and solutions, most of which have undergone several changes within the history of one “product”.

Electronic texts of rabbinic literature with more or less sophisticated search-engines were sold from the 1980’s onwards. In UNIX and DOS they needed to overcome the barrier created by ASCII and the national code-pages, usually by forcing the user to use the Israeli national code-page and thus not being able to access any regional national characters. The fact that Hebrew had no universal ASCII standard but existed both in lower ASCII and in higher ASCII added to the representation problem. The main function of these early electronic texts was the concordance: search-engines would find all occurrences of a given string in the selected database. Programming energy accordingly went into the search-engines that are by now able to combine different strings, to include all or select prefixes and suffixes with the string, or even to perform grammatical searches that will find all occurrences of a lemma or root. Hyperlinks that will call a connected passage to the screen were a comparatively late feature. Generally it can be said that these databases are retrieval systems for authoritative texts. Some

---

15 In the earlier times of Hercules Graphic Adapters manual switches were invented as substitute for the constant switching of jumpers on the Graphics Adapter. Later small batch-files could be employed to change to the Israeli national code-page, usually unidirectional. Applications under the operating system Windows were able to solve the problem of representation.

16 Cp. e.g. the Bar-Ilan-Responsa-Project or similar products by commercial providers. Cp. also the elaborate search-engine used by Maagerim (Israel Academy of Language) that will find both different form of the same (even irregular) root and grammatically identical forms of different roots.

17 The Bar Ilan Responsa Project introduced these hyperlinks in version 5, autumn 1997.
of them have export-filters but due to the problems of Hebrew representation cut-and-paste is usually not supported. While marketing strategies often connect these tools to the study of rabbinic texts, their actual ability to teach the knowledge and techniques necessary for a thorough understanding of rabbinic texts is minimal to nonexistent.

With SGML and the Text Encoding Initiative’s attempt to provide a standard document type description (DTD), the description of electronic texts in a meta-markup-language became possible and spread widely in European literatures. It did not however catch on in Hebrew literature, possibly because of the previously existing concordance-based works in this field, possibly because SGML does not support any of the Hebrew standards and no SGML-editor or -browser supports right-to-left (RTL) text. L. Barth attempted to adopt the TEI DTD for his electronic edition of *Pirqe deRabbi Eliezer*, a visionary project that includes transcription and tagging of all available sources of this Midrash. He defines many new standards and overcomes the TEI DTD standard of describing mainly the physical structure of a published book as opposed to its contents-structure. His ingenious reworkings in order to use Hebrew in SGML evince the need of a complex mark-up-language that masters multilingual and bi-directional texts.

Text tagged with the TEI DTD can be used for complicated retrieval processes and comparisons. Several other projects have shown how these SGML texts can be further enriched with additional information and made the core of complicated information systems that include text, commentary, illustrations, etc. A special case of SGML tagging and application is “The Wife of Bath’s Prologue on CD-ROM”, published in 1996 by P. Robinson. This application enables extensive research into the presented sources. An optimal depth of tagging and commenting has been reached combined with collation, hyperlinks and digital images. This depth is possible only if the amount of text studied and prepared is limited: in this case the corpus contains 861 lines – albeit in 53 manuscripts. Since all possible research activities had to be preprogrammed into the electronic text, the work is a fine example of state-of-art-scholarship as well, with hardly any questions on the text, its transmission and its linguistic features left open.

---

18 Cp. names like “Bar Ilan Study Buddy” or descriptions like “Incomparable tool for Torah research”.
20 L. M. Barth describes his project at http://www.usc.edu/dept/huc-la/pre-project. Cp. also his publication on the electronic publication of Hebrew texts, *AJS Perspectives* 2,1 (2000). His is the only project TEI lists as using the TEI DTD for Hebrew texts, no projects in other Semitic languages are listed.
21 Cp. e.g. the Perseus-Project, at http://www.perseus.tufts.edu.
Any attempt to produce a full electronic Talmud-text that is as exhaustive in its textual quality and commentary will have to fail, given both the size of the talmudic corpus and its long history of commentaries that would have to be included. The open character of the Talmud and its ongoing commentating processes interdicts any attempt of completeness. Instead every electronic presentation of the Talmud will have to address a specific audience with its specific needs.

3. Concepts for an Electronic Study Environment

All existing computerized study tools for the Talmud focus on aspects of traditional Jewish learning. They can roughly be grouped into two classes: 1) drill programs for talmudic vocabulary and 2) modern commentaries. While the first group does not refer to the complete text of a tractate or any unit of argument (sugiya), the second group contains programs that combine the image of the page with audio-files containing a traditional lesson (shi’ur). Some resources add dictionary functions and other reference resources to the traditional commentaries. They aim to bring the Yeshiva educational experience to a Jewish audience that wants to study at home. The user is expected from the outset to be familiar with certain talmudic concepts and Hebrew technical terms.

Outside of traditional Jewish learning, the Talmud is studied in academic Jewish Studies departments and programs around the world. Students and scholars of Religious Studies are as interested as Christian Theologians and – since Levinas and Derrida – scholars of literary theory. The difficulties of language, style and talmudic argumentation patterns present themselves as almost insurmountable hurdles to anyone wishing to read talmudic passages for the first time. Scholars who do not wish to devote most of their time to talmudic studies therefore often rely on secondary sources or translated extracts, without ever being able to read and understand a longer passage of Talmud in its context. Students of Jewish Studies on the other hand often find themselves painstakingly scrutinizing single talmudic sentences in order to master – in addition to the language – both the literary structure and the halakhic contents. The literary quality of the Talmud unites both groups of potential users of an electronic study environment. It may also be the quality of the Talmud that presents the greatest challenge for an electronic study environment.

23 A. Steinsaltz e.g. spent more than 20 years on publishing his vocalized Talmud edition with his commentary in modern Hebrew. Of the critical edition of the Babylonian Talmud prepared in Jerusalem, 9 volumes, containing each only part of a tractate, were published in more than twenty years. The only complete new commentary on the Babylonian Talmud that was published in recent years is J. Neusner, The Talmud of Babylonia. An Academic Commentary, 46 vols., Atlanta, 1994-1996, a work for which Neusner could rely on his complete American translation of the Babylonian Talmud and on his many previous studies on the Talmudim and rabbinic literature in general.
environment since it is not easily translated into databases or other forms of fractionalized and reorganized data.

In rabbinic literature the passing on of knowledge is always connected to predefined forms and structures, like *Leitworte*, hermeneutic rules and stock phrases. Readers who are not familiar with the rules that govern the use of these entities need either teachers explaining the texts to them or other means that will provide the necessary tools in order to detect the literary structures of talmudic passages. By stressing the visualization of the parts of the text that can be studied independent of the specific topic discussed and by enriching it with on-demand accessible information on the rules that govern their respective uses we hope to enable readers who are not familiar with rabbinic literature to develop the necessary skills to read, study and understand these texts.

The main functions of the envisaged environment thus will be the possibility to read the talmudic text in the original and in translation in a way that visualizes structures and associations and opens up the different ways to read the text and the possibility to study literary characteristics of the Talmud and the linguistic forms they take. In addition, the study environment will make long talmudic and midrashic texts (with the accompanying commentary) available to a broader audience not necessarily fluent in Hebrew and Aramaic.

The Individual Parts of the Study Environment

The first element of our study environment is a German translation of a rabbinic text. We chose the Babylonian Talmud’s tractate Megilla, since it contains both halakhic and aggadic material. The subject matter of the tractate, the reading of the Esther-Scroll on Purim, is expanded into general regulations on Bible readings in other liturgical context. The advantages of a subject that can be related to everyday Jewish life for a study unit is evident. The long exegesis of the book of Esther at the end of the first chapter of Megilla is one of only three long midrashim in the Babylonian Talmud. In addition to the Babylonian Talmud’s tractate Megilla we translated Mishna and Tosefta Megilla, the biblical book of Esther and all the rabbinic commentaries on Esther. Since the projected audience for the study

---


environment will mainly be readers who are not yet familiar with rabbinic literature the translations are a core element of the study environment. The translation is marked and enriched with additional information. The same mark-up and information will be applied to the Hebrew texts as well.

For the cooperation of scholars of Jewish Studies with German Computer Scientist it seemed logical to first develop a general design using material that all involved scholars could read and then analyze where Hebrew standard applications can be used and where additional methods have to be developed. In the final product readers will have the choice of reading either the Hebrew/Aramaic original text or the German translation. If we can find a way to divide the screen without losing legibility we should like to offer both texts on the same screen, possibly with parallel scrolling. This feature will be necessary if the study environment is used for teaching Talmud in university courses. It will appear dispensable to users without prior knowledge of Hebrew and Aramaic and should thus be optional.

The first elements that have to be distinguished in their visual appearance are quotations from other texts. In talmudic translation it is customary to print quotes from the Mishna in small capitals and quotes from the Bible in italics. For the Esther-commentary in bMeg 10b-17a we decided to further highlight the Biblical verses discussed by setting them in bold and italics. In addition to that we introduced italic small capitals for quotes from the Tosefta. In our printed edition, parallels to rabbinic texts are marked by curly brackets at the margin of the text. For the electronic text the curly brackets had to be substituted by a more distinct element.

Different levels of text that belong to different levels of discussions should be separated from each other as well. For the printed text we used different paragraph indentation thus highlighting those parts of the text that could be identified as “excurses” to the surrounding argumentation. Indentation continues in the electronic version, but additional manipulation of the levels of argument is necessary and possible.

In addition to these visual markers that are usually employed in printed texts, the electronic version will contain the possibility to highlight classes of text-elements that structure the text to ease reading. Since the original text is printed without punctuation, marking the names of rabbis that usually introduce a new argument enables the reader to separate the text into units. Standard phrases create the logical framework of the text and organize the contents. In addition to that, we marked Leitworte that structure longer units of texts. For these items it is possible to access brief citation, for some of the items longer commentaries will also be available. From the outset we

26 For the technical realization of our project this method of ignoring the possible problems of Hebrew turned out to be a sensible solution: while we developed our Study Environment, developments in the representation of Hebrew caught up with our demands.

27 Cp. the layout of the printed texts.
decided that information should be given on-demand, i.e. the user may choose whether or not a group of textual elements is highlighted. The question whether additional information should be available only when a given item is highlighted or not was discussed controversially in our team. On the one hand we appreciated the need to provide as complete a set of information at any given time, on the other hand, users might be overwhelmed by too much information. For the final product we may decide to offer the choice between these two options. The user would decide at the onset of a session whether s/he wants the complete set or only the selected information.

The additional information differs from class to class. The names of Rabbis are enriched by short “biographical” information as stated in the traditional commentaries, naming the generation a scholar is affiliated with, his academic home, and, if known, his teachers and/or pupils. Longer information on rabbis gives slightly longer “biographies” with bibliographical data and some assumptions about a rabbi’s role in the rabbinic universe. Genealogies of the rabbinic class are part of Talmud commentaries since the Middle Ages, the information distilled from them is essential for the traditional understanding of the Talmud. The many problems inherent in this pseudepigraphic and literary construction will be dealt with in an essay in the Study Environment, though they can not be fully discussed in this framework.

The most common marked elements are the standard phrases that structure the argument. Here the brief citation in the German translation will provide the original text while the brief citation in the Hebrew text will provide our standard German translation. Additional information will also be given on the use of a certain key phrases. The application of hermeneutical rules in the talmudic text will be enriched with short explanatory texts.

J. Neusner and many others have proven in many publications that rabbinic texts are not unstructured collections but intentionally structured and carefully built texts.\textsuperscript{28} \textit{Leitworte} are especially important to understand the order of the argument structure. By offering the possibility to jump from one highlighted \textit{Leitwort} to the next we want to construct a new, non-linear reading of the talmudic text that basically follows the rabbinic technique of connecting material by the use of \textit{Leitworte}.

In addition to these structurally important elements, we decided to supply information on two more general classes of elements. On the basic level we will enrich the text with a kind of dictionary that will explain talmudic terminology, foreign words and – on a more general level – difficult terms. This feature will be more important for the reader using the original text than for the translation, though it will be present in the German version as well. These dictionary entries for difficult Hebrew and Aramaic terms and

\textsuperscript{28} Cp. e.g. J. Neusner, \textit{The Rules of Composition of the Talmud of Babylonia}, Atlanta 1991.
loanwords in the German version will engender curiosity and invite the reader to explore every possible aspect of the talmudic text.

In addition to the talmudic and midrashic texts we decided to include a selection of commentaries and interpretations. A generally different form of highlightening is necessary in order to separate this type of subjective information from the more objective information in the previous cases. Commentaries and interpretations can be attached to passages of text or to single elements, i.e. words or phrases. Again brief citation and longer descriptions can be offered. Among the longer texts is an extensive commentary that analyzes the tractate Megilla and describes how passages relate to each other. This commentary is divided into units that attach to selected paragraphs.

Although we used the standard printed text of the Babylonian Talmud for our translation, the importance of manuscript transmission has to be realized. A critical edition of the tractate Megilla does not exist and our three-year project could hardly aspire to provide one. We therefore selected four main manuscripts that will be compared with the Vilna-text. Significant variants will be added as brief citations both in the translation and in the original text.

With the possible exception of the *Leitworte* the described features stay at the surface of the text. We need further possibilities to create new ways of reading. One option is to hide or display textual levels. Since levels are relative to the Mishna discussed in a talmudic chapter, with the Mishna on level 1, commentary on the Mishna in level 2 and associated excursions or super-commentaries in levels 3-6, the option to hide all text in the subordinated levels is necessary. It is thus possible to read only those passages that deal with the Mishna itself or to chose how far a reader wants to follow the textual fabric into its super-commentaries and discussions of minor details.

Another option would be to translate the textual structure into tree-structures, marking the different branches of argumentation and indicating where sub-arguments begin and end, where structures like lists of three, seven or twelve parallel items occur, and what associations triggered the deviations from the main argumentation.

---


30 Ms Columbia X893 T141; Ms Göttingen cod. Heb. 13, Ms Oxford Bodleiana 366 (Opp. fol. 23) and Ms Vatican ebr. 134.
This core of the study environment – the marked and enriched text and translation – will be surrounded by larger units of information, essays etc. that can be accessed via menus. It is necessary that any longer texts will be printable, since reading from a computer screen is not considered a highly successful method of studying.\(^{31}\)

To realize this environment we needed computer programs that would be able to deal with long texts, support Hebrew, support hyperlinking, allow us to attribute “brief citation”-strings to items and would enable us to manipulate the visualization features. None of the standard hypertext-packages do this.

4. Technical Realization

Software Engineering\(^{32}\) is the discipline responsible for the technical realization of information systems. Software Engineering depends on so called phase models to divide the whole process of engineering an information systems in manageable parts. Every phase has a clear and distinct goal as well as a defined output. Every phase is connected to the next phase in the process. Jumping back in earlier stages of the engineering process is usual in real world projects but costly. We concentrate here on three essential phases: 1) The requirements engineering\(^{33}\) phase, which transforms a “vision in context” into a system specification (the system specification defines the “what” of the system but not the “how”); 2) the design phase, which defines sequentially the “how” of the system; and 3) the implementation phase, which is stamped by coding and testing activities following the specification and the design documents.

Requirements engineering

Besides technical features for the study environment the analysis of requirements also incorporates usage and development aspects relevant for different classes of readers - students, teachers, interested laities – and developers. To capture these different usage aspects we adopt a use case-driven approach to scenario-based requirements engineering.\(^{34}\) The main requirements engineering document was written cooperatively so that the

\(^{31}\) Psychologist suggest to minimize the information content on screen to the amount the reader can process in her/his short term memory, cp. e.g. R.E. Horn, Mapping Hypertext. The analysis, organization and display of knowledge for the next generation of on-line texts and graphics, Waltham 1989, p. 86.

\(^{32}\) Cp. e.g. I. Summerville: Software Engineering, Reading, MA, 2000.


different stakeholders could add their viewpoints. The specification contains not only functionalities of the study environments but also visionary details of a future multimedia computer-based training environment. However, in this project we want to set up a simple and user-friendly study environment following the translated texts and additional material only available in the electronic edition. So, we concentrated on students and designed an appropriate environment based on the existing material.

Design of the environment

Rabbinic literature is, as mentioned earlier, a kind of hypertext before hypertext theory existed. None the less, it is usually printed in traditional book-form. Accordingly, our German translations were also published as books and therefore were originally prepared for camera-ready copy. The electronic edition however will use the facilities of hypermedial computer-based representation and presentation. As we rely on the printed edition, we have a document space containing Microsoft (MS) Word documents. All these documents are formatted according to guidelines that transform the structure of the text into a visualization scheme on paper. As we said earlier, none of the available hypertext packages support all the initial requirements captured in the first phase, such as Hebrew language and hypermedial representation and the hypertextual presentation of the documents. Hence, we need a representation and presentation language which is capable of fulfilling the requirements mentioned. Fortunately, the set of presentation and representation languages grouped around the XML technology solve our problem. To transform the document space given by MS Word documents we chose XML for the following reasons. First, XML allows – like HTML and SGML – hypertext linking features. However, HTML allows only presentation oriented tagging of documents, i.e. no semantics are supported directly. Furthermore, HTML lacks the possibility to define new tags. In short, HTML is a language for representation. On the other side, SGML is a very complex meta language for the electronic representation of text and the available SGML-editors are very costly and do not support Hebrew. Beyond this, browsers for SGML depend on a local installation of the software and the documents and are intended for use in CD-ROM based productions only. With XSL there is a convenient way to

---

35 Extensible Markup Language (XML, cp. http://www.w3.org/TR/REC-xml). XML has been designed to enable the use of SGML (Standard Generalized Markup Language, ISO 8879) on the World Wide Web. It is not fixed like HTML (Hypertext Markup Language) and “it removes many of the underlying complexities of SGML in favor of a more flexible model, so writing programs to handle XML is much easier than doing the same for full SGML.”, cp. http://www.ucc.ie/xml/#FAQ-ACRO (the frequently asked questions on XML).

36 Well known vendors are e.g. ArborText, Frame Technology Corp. and Interleaf, cp. http://www.exeter.ac.uk/SGML/report/annexd.html.

37 Extensible Stylesheet Language. For a short history of XSL, frequently asked questions, available software and a tutorial look at http://www.w3.org/Style/XSL/.
transform the XML documents into other XML documents, e.g. HTML pages for representational purposes. Second, XML provides extensive hypermedia linking facilities with extensions like XLink\textsuperscript{38} and XPointer\textsuperscript{39}. Third, XML is specified for Unicode use, thus supporting also Hebrew.\textsuperscript{40} Fourth, there are available XML databases like Tamino from Software AG to construct a database for future usage and research perspectives. To summarize, XML technology is an ideal candidate for designing and implementing the study environment. It allows the division of representation and presentation, the support of hyperlinking, Hebrew language and advanced database functions.

Implementation of the environment

The implementation of our study environment follows the requirements extracted from the analysis of former projects and the literature and the design decisions described above. To realize a XML-based study environment a DTD is not really mandatory. This is an advancement regarding SGML, where every used tag has to be provided by the DTD. Nonetheless, we need a DTD for two main reasons. First, we want to realize a database containing hypermedia objects semantically tagged by XML. Hence, for describing schema information a DTD is a good choice. Second, we want to semi-automatically transform existing materials, such as German translations of Mishna, Tosefta and Babylonian Talmud Megilla into XML documents to fill our database.

We developed the conversion tool DocToXML to transform the existing MS Word documents into an XML representation. DocToXML is written in Visual Basic for Applications. In the current version DocToXML is able to transform the whole given document space into XML\textsuperscript{41}. Yet, not all information in the DTD can be extracted from the document space and conversion errors are possible.

To add tags to the transformed XML document and to correct errors we use the XML editor XMLSpy 3.5. XMLSpy is a general purpose XML editor for little projects and one of the few handling Unicode. For every given XML document we can check well-formedness according to the definition of XML, e.g. that every position of opening and closing tags is correct, and validate against a given DTD. After performing these two procedures we

\textsuperscript{38} XML Linking Language, cp. http://www.w3.org/TR/xlink/.

\textsuperscript{39} XML Extended Pointer specification, cp. http://www.w3.org/TR/WD-xptr.

\textsuperscript{40} As to Unicode glyphs for Near Eastern languages in general, cp. C.A. van de Repe, ‘Computing for philologists (4); Appendix: Unicode and Near-Eastern scripts; Postscripts on the Unicode proposals’, DS-NELL 3 (1996), pp. 205ff.

\textsuperscript{41} We plan to extend DocToXML to a general purpose tool. With so called wizards, little parameter windows, every given formatting information in a MS Word document can be assigned to arbitrary XML-tags given by DTD.
can ensure that the XML document can be represented on the screen with XSL.

XSL is a stylesheet language to specify the presentation of a class of XML documents by describing how an instance of the class is transformed into an XML document, e.g., an HTML page. We developed a stylesheet for representing the XML documents as HTML pages using JavaScript as script language. Since we rely on HTML any popular browser, e.g., Microsoft Explorer or Netscape Navigator, is capable of representing our study environment.

Figure 1 shows a screenshot of our study environment in an internet browser with a German document loaded. Documents available in the study environment are listed in a scrollable list box. Users can choose one of the documents by clicking on its call-name. The related document is then loaded from the document base. To minimize loading and processing time, the talmudic text had to be cut into small units. Instead of using the highly arbitrary division by talmudic pages, we cut along the lines of Mishnayot commented on and followed the textual structure wherever possible.

The loaded document carefully conserves the layout of the printed text. Indentation is kept. The curly brackets in the printed text that marked

Figure 1: Screenshot of the study environment in an internet browser (all readable information in German)
passages that have parallels in other rabbinic texts were substituted with bright red arrows in the electronic text. Footnotes from the printed translation indicating the exact parallels had to be translated into a different visualization scheme. In this case we chose commentary-boxes that appear when the mouse is moved onto the arrow. The commentaries, many of which are not part of the printed translation, are indicated by blue bullets. Additionally, as in the printed edition, the mishnaic and biblical texts under discussion in the talmudic text are marked on the left margin of the text. The same column contains indicators for the textual structure.

Every document represented in the browser has different parameters that can be manipulated by the reader in the study environment to give further information on-demand. The parameters are visualized by checkboxes in a non-scrolling frame. Thus the chosen parameters are visible all the time. Additionally, the given parameter setting is saved when the user changes the document, enabling her/him to follow a chosen path of study through longer units of text. The parameters are:

1. Ebenen (Levels): We have differentiated five levels of textual structure (cp. p. 11) which are enlisted in a scrollable listbox. The reader can choose how many of the levels are to be displayed and the whole page layout adopts itself to the chosen level. Hidden paragraphs leave a small amount of white space as their trace, thereby indicating roughly the amount of hidden text.

2. Bibel (Bible): If marked, all hypertext nodes containing quotes from the Bible are highlighted in orange and italics, preserving the tradition from printed books in addition to the color-code.

3. Fachbegriiffe (Technical Terminology): If marked, technical terminology from the Talmud and from Jewish Studies are highlighted in grey. If the pointer is moved over the highlighted text a small text-box with light yellow background and a red margin containing a short definition and explanation of the concept in question, pops up next to the highlighted word. The window closes again when the pointer is moved away.

4. Handschriftenvarianten (Manuscript Variants): If marked, a small green glyph is displayed before the text to which significant variants exist. If the mouse moves over the glyph, a text-box with the text of the variant and its source is displayed.

5. Interpretationshilfe (Interpretation hints): If marked, little blue squares are inserted before words or phrases in the screen representation. If the mouse is moved over such a square, a textbox pops up. The textbox is only visible while the mouse is over the square and an interpretative help is given.

6. Leitworte: If marked, Leitworte are highlighted in a bright blue. We plan to create links so that the reader can jump serially to the next
Leitwort by clicking on the highlighted word once XLink and XPointer have been fully implemented.

7. Rabbinen (Rabbis): If marked, names of rabbis are highlighted in red, if the mouse is moved over them a little textbox with a short “biographical” note pops up. A separate window with longer “biographical” and bibliographical information can be opened with a mouse-click.

8. Standardformulierungen (stock phrases): If marked, phrase nodes are highlighted in dark blue and if the mouse is moved over them, a little textbox with the original text (in the German translation) or the German phrase (in the original text) and a short explanation of the function of the phrase in the talmudic argument or the hermeneutical rule employed pops up.

9. Worterklärungen (language explanations): If marked, words are highlighted in purple and if the mouse moves over them, again a little textbox with a language explanation pops up.

Every document available in the listbox is linked forward and backward, so the user can click on these links at the bottom of the screen-representation of the documents to navigate in the document space.

Creating Hebrew documents and their representation on screen proved to be easier than expected. Since we had decided to use Unicode, the previously
enriched German files served as models for identical Hebrew files that can be represented in standard browsers, where any text with a tagged item will be represented in the direction required by its language. Hebrew and German files thus use the same DTD. In a separate stylesheet for the Hebrew files, the direction for these documents was set to “rtl” (right-to-left). This automatically changes the direction of the columns defined for symbols and structure indication and guarantees that the tagged items are represented in the correct order. With the exception of the stock phrases, where the original text shows the German translation on demand while the translation shows the original phrase on demand, the Hebrew text is enriched with the identical information as the German translation. Both groups of texts together create a unified document space.

5 Conclusions and Outlook on Further Work

The need to study texts beyond merely reading them, possibly in translation, creates the need for study tools that offer texts and additional information in a unified environment. Most electronic study environments are not fit to feature large amounts of texts and do not care to visualize the content structure of the texts. Our Comprehensive Electronic Study Environment strongly focuses on the visualization of content structures. It thereby aims to enable the reader to understand how the texts are composed. This cannot nor will not substitute for the teaching situation in class, but it ought to add a new quality to regular teaching, since the student will have prepared the texts far better than s/he can with the help of traditional resources alone.

In the final Study Environment this document space will be combined with other document spaces that include secondary material on the transcribed texts, images, audio-files and multimedia objects. Further elements are planned that will be implemented with XML and its extensions. A searchable bibliography will be created using a standard XML database model. Search routines should enable the reader to navigate in the document space both according to content and according to XML description. Combining text related technologies with database technologies we hope to achieve unlimited navigation in the document space. Additionally, special reading paths can be created that lead the reader through the text once XPointer and XLink have been fully implemented.

Further work might follow several different paths. On the one hand, the Study Environment should be developed further with pedagogical units and tests so that the user can measure her/his success. The Study Environment

---

42 This command also switches the position of the scrollbar, which now is on the left side of the window, a slightly annoying feature in IE 5. For a detailed comparison of Hebrew representation in different browsers cp. http://www.nirdagan.com/hebrew/review.

43 The Perseus project (cp. note 21) may be an exception, but it might be argued that it offers no direct study help for the content structure of the texts. It multimedia surrounding and the image database are its main points of visual learning.
could then be used for home assignments in university courses or similar teaching environments. On the other hand, the concept of the Comprehensive Study Environment for texts ought to be developed as an independent shell that can be filled with different data, i.e. as a tool to create individual study environments for any given set of texts. Many fields outside the comparatively small scope of Jewish Studies face similar problems of teaching reading skills beyond pure lexical understanding. Any application created for Hebrew texts would as well work with Arabic, Syriac, Persian, and other right-to-left languages, with small changes in the type of short and long information supplied. Ancient Near Eastern texts for which a unified XML DTD has been planned\textsuperscript{44} could be enriched for study purposes using the same method. On an even broader scope the shell could be developed to cater to all literatures as a tool to create Study Environments on highschool- or university-level, i.e. educational software for the regular market.

\textsuperscript{44} Cp. the description of the conference “Electronic Publication of Ancient Near Eastern Texts” at the Oriental Institute of the University of Chicago on October 8 & 9, 1999, published at http://www.ariadne.ac.uk/issue22/epanet.