Digital Judaica Done Right

Table of Contents

Introduction

What is this About?

We want to have a computer environment that supports and facilitates study and research of Jewish Orthodox texts. (Most of the technological pieces that we need to develop are not at all specific to Jewish Orthodox Texts. On the other hand, desire to facilitate study of such texts is our primary motivation.)

This "dream system" will record various information about the texts and support doing interesting things with them.

On the text level, we want to allow marking up names of people and places, index entries, logical structure of statements, user-defined tags and the like. On the structural level: allow capturing of multiple structures of the same text, for example - chapter/verse and pages of a printed edition. On the intra-text level: record links from one text to another with their types and other metadata, and support link reversal.

It should be possible to print any text with glosses formatted nicely.

The system should be accessible through a clean, even minimalist, but powerful web-based interface. Mass participation - proofreading of the texts, marking them up, clarifying the cross-references and such - should be very much supported.

History

Initial conversations in this area took place in 1992-1994, between Leonid Dubinsky and Baruch Gorkin. Most of the requirements were understood then, but not the need for universal web-availability and crowd-sourcing: access to the Internet was not what it is now.

We realized that a standard approach to text markup has to be used, and settled on SGML (XML did not yet exist).

In the summer of 2006, discussions restarted between Dubinsky and Gorkin. Crowd-sourcing and potential commercialization were discussed. In the Fall of 2006, Michoel Koritz joined in.

Format

Some of the research notes, discussion items and conclusions related to this project were captured and published on the project's web site.

Standard XML formats for text publishing were tried: DocBook and TEI. Since the format that is likely to be used for the storage of the texts themselves is TEI, we wanted to use TEI as the format for publishing the project papers too. This turned out to be impractical, since there is no convenient way to edit TEI document over the web using wiki-like interface. It seems that some people tried creating packages with such functionality, but nothing is being developed now. There is the DocBook Wiki, but it's approach to section editing leaves much to be desired, and anyway, TEI and not DocBook is our preferred format. Also, the machinery for turning TEI into PDF had some issues [http://listserv.brown.edu/archives/cgi-bin/wa?A2=ind0712&L=TEI-L&T=0&F=&S=&P=63].

We then published in a form of wiki pages - some of them in Russian. A blog was deemed an appropriate format, since it captures a sense of development in time, allows comments and, hopefully, encourages smaller, but more frequent publication :)

As part of this project, we need to develop technology for web-based editing of big texts with rich markup and complex structure. When we have it, we'll use it for publishing our papers, so this is not the final format ;)

Onward!

What do we want? (The Dream)

#########

###############, #######, # ########## #### ########## # #######, # ###### ###-## ######.

########

(###### 2), ######### ### ##########, ###### ### ######### ########## ######## (###### 3) # ########. ######### (###### 5) ###### ############# ##### ################# #############, # ####### - ####### ###### #### ######### ###### ########## (###### 6), # #### ########### ######. # #########, ########## ##### # ###### ####### ########## ### ##### ###### ("build it and they will come").

#######

#####

######, ######### ## ####### ####### ###### ######. # ########## ####### ## ###### ##### #########, # ## # ##########.

"######" ##### (###########, ###########, #############), # #### ## ########.

####. ########, # #####- # #####- #########.

#########

textual features, ## #### ######## # ######## ##########, ############# # ######:

- ##### #####

- ######### ####

- ########## ##### #############

- ######## #############

- ##### ####### # ############# ########

- #####, ### ######### ####### ## #### ######### ###### # ######

###########
######:

- ## ##### ###### ####### ######### ########

- "####"

#########

#######, ############,
#########, ### ######. #######: ##### ######### ##### #####, #####,
(### ####### -
- #### #### #######) #### ## ##### #####, #####,
####, ############ ##### #######, #####, ##### ###### #######, ####/
####### ###### #######, #####, ############# ###### [#####,] ######
#, ##### [, ############] ###### #### ######, #####, ####[, #### #####]

#######, ########, #######,
- ### ########## ####### ###### # #### ## ######.
############ ####### ####### ##### #### ##### ############.

########
(####/#### # ##### # ######; ######## # ############# # #######).
"##############": #####
######### ####### #####, # ######### ###### - ####### #####.
######### # ###### ######### ########## # ###### ######, ##### -
####### ###########.

######

#####. ###### ######### ####### ### ##### ###### ###### # ######### ### ###### #######.

#######, ### ###########. #########, ###### ############# #### # ####### ### ########## ###### ##### ###### ####. ######...

#####e #### ######:

- ###### ############ #####

- ###### ######### ##############/############ ######

- ###### ############## #####

"####" ######.

##############

######, ################ ######## ###### ####### - #########.

(#### - ######) ### #### #### (###### - ###### ## ####### ########### # ###). ##### ##### ############## # ######, # ######### (######### ## ####### # #######).

#########

("#######") ###### ######. ########, #########/#######/######## ########## ####### ###### ##### ############# - ### "#######" (######).

#####

(XXX ######?)

source control system - ####### ######. ### ## ###-###### ####### # #############...

It seems that branching and merging are necessary.

########## #########

###, ### ##### ####### ##### ######### ############, ###### ####
############ ####### # ##### ########## ######### ####.

########.

Add/change. Add/change metadata.

Look into AtomPub, WebDav and distributed source control - say, Mercurial.

######### ########

#########.

######### ######

######, #.#. ###########, #####
######### # ######### #########.

########

TODO

######### #####

- # ###### #########
- # ###### ###### ######. ########:

- ## ######## ######

- ### ########## ####### ######

- ### ############ ############## ######

- ## ######

- ## ###### ######

- ## ######### ###########

- ## ####### #############

- ####### ## "######### ######"

- ## "######### ########"

######: [Information Retrieval from Annotated Textxs]

#########

#######. ########:

- ####### # #########

- ###### ####### ###### ######

- ###### #### ##### #### # ####### ###### ####

#########. ########, # #### ############# ########.

##########. ##### ###### ########### ## ### ### #### ########## ##### (######; ####; ####### #########) # ########### ## ############ ############.

#########

XXX "# ## ##### #". ## ## - # ################ ########. ########### ########### - #### ######. XXX ### # ###########.

(##### ######## #####) ## ###### ####; ((##### ####### ####) ####### ####) ## ##### ####.

####### ##### ######### ###### - #### #####; ##### # ########### #############.

######

#######

#################

- ###### ######### #####

- ####### ##### - # ####### "######"

- ######### ########### ######## ####### ### ## ###### ####### ##
############# ### # ## ############ ######

- ########### ########### #######, ######### ## # ######,
###########

#########

############ ## ####### - ############## # ##### #### (#########);
#####; ##### "######": ###/#####; #######: ######/########### ###;
######### #####, ######### # ######### ######, ####### ######
#####, ######### ########### # #.#. ## ###### ####### ## ########
########## ####### ######, ########### # #### (# ###########
#######), ######### # ############## ## ######, #########
#########. ###### ############# ####### #######. "####### ####":
######### ###### # ####### #### ### ####### ##### - #### #####
(######## ########### ######). Google Notebook. JotSpace
[http://www.jotspace.com/].

########. ##
, ### ### ######### ## "#####",
"#####",
########### ############# ## ####, ######### ########## ######
######### ### #####. ##### # ########## #### # ##### ######

########### ##### # #######. #### ######## ## ###### (### #####, ###
#####, ####, #### ...) #######, ##### ## ######## ###### - ###### ###
########## (######### #####, #####, ######, ######### #########) #
#######.

K#### ###### ######## # ####### "######" ##### ############
######## ###### ##### ######.

########
####-############### # ###### ########. (### ######### # #######
########## #### ######### ######## ###############, ##### ##### ##
########)

###########.

##########

#######:

############ #######

####### #######: Wikipedia, Wikisource, Distributed Proofreaders

######## #######

######## #######

#############

(XProc/XQuery/XSLT)

######

######

############### (web)

########## # #######

Wikipedia

Live Journal

Del.icio.us

Digital libraries

######

##########: Wikipedia

########: Wikipedia

Citizendium [64]

http://many.corante.com/archives/2006/09/18/
larry_sanger_citizendium_and_the_problem_of_expertise.php   [http://
Fmany.corante.com/archives/2006/09/18/
larry_sanger_citizendium_and_the_problem_of_expertise.php]

http://many.corante.com/archives/2006/09/20/
larry_sanger_on_me_on_citizendium.php

#########/############ (#########,
##################, ########)

########, ####### #########

##################/

####

########. #####,
######## ######## ### ###### ######## ### ###### ###### ######
######. # #############, ## ########## ### 10 ############ #######
####. ### ######## ##### #### - # ###
#####. ######## # ######### #### ######### ###### #####
###. ######## ###### ##### ##### ###. ####### ######### ##### # ###
- ### ###### # arXiv#.

############### #### (### ###### #####) ## ######### #######.

######

############### ############### ##### ## ######. ########, #######
######. ### - ###### # "###############" #######.

Google

###############: GData.

#####.

- ## #######.

- ## ## ######.

######

####### ####: ###### ### ##### ### ######### #####?

######

####### # ######## #####

##########

TODO

########

######

#######? ##### ######### ###: OpenTorah # "##### ####".

Sources

[Fraenkel97] *The Responsa storage and retrieval system-whither?..* Aviezri
Fraenkel. 1997. http://www.wisdom.weizmann.ac.il/~fraenkel/Papers/
trs.ps. http://www.wisdom.weizmann.ac.il/~fraenkel/Papers/pha.ps.

[CAB] *The Cathedral and the Bazaar*. Eric S Raymond. http://www.catb.org/
~esr/writings/homesteading.

[Ontology] *Ontology is overrated*. Clay Shirky. 2005. http://www.shirky.com/writings/ontology_overrated.html.

Distributed Proofreaders. http://www.pgdp.net/c/default.php".

TEI Lite Tutorial. http://www.tei-c.org/Lite/teiu5_split_en.html.

TEI P5. http://www.tei-c.org/release/doc/tei-p5-doc/html/.

TEI CE. http://www.tei-c.org/Activities/CE/.

TEI Overlap. http://www.tei-c.org/Activities/SIG/Overlap/.

TEI WD. http://www.tei-c.org/release/doc/tei-p5-doc/html/WD.html.

TEI NH. http://www.tei-c.org/release/doc/tei-p5-doc/html/NH.html.

TEI I18N. http://www.tei-c.org/I18N/.

TEI Roma. http://tei.oucs.ox.ac.uk/Roma/.

?. http://www.w3.org/People/cmsmcq/2000/poddp2000.html.

?. http://www.mulberrytech.com/Extreme/Proceedings/html/2004/DeRose01/EML2004DeRose01.html.

?. http://www.idealliance.org/papers/extreme/Proceedings/html/2005/Bauman01/EML2005Bauman01.html.

?. http://www.tei-c.org.uk/wiki/index.php/SIG:Overlap.

?. http://www.tei-c.org/wiki/index.php/Talk:SIG:Overlap.

?. http://www.tei-c.org/Talks/OUCS/2005-02/talk-access.pdf.

Subversion. http://subversion.tigris.org/.

eXist XML database. http://exist.sourceforge.net/.

JXHTMLEDIT. http://www.tecnick.com/public/code/cp_dpage.php?aiocp_dp=jxhtmledit.

Syntext Serna. http://www.syntext.com/products/serna/index.htm.

Stylus Studio. http://www.stylusstudio.com/.

ALTOVA xmlspy. http://www.altova.com/.

oXygen. http://www.oxygenxml.com/.

Exchanger XML. http://www.exchangerxml.com/editor/.

editix. http://www.editix.com/.

topologi. http://www.topologi.com/.

Ajax. http://en.wikipedia.org/wiki/AJAX.

Applets. http://jroller.com/page/tackline?entry=using_applets_in_place_of.

Applets. http://weblogs.java.net/blog/chet/applet-jax/Yapplet.html.

Thank God - Java EE Is Not Like Ajax. http://www.coachwei.com/blog/ _archives/2006/9/27/2367882.html.

Unicode. http://www.unicode.org/.

Unicode. http://www.w3.org/International/articles/inline-bidi-markup/.

XML. http://www.xml.com/axml/testaxml.htm.

XLink. http://www.xml.com/pub/a/2002/03/13/xlink.html.

Wikipedia: Tim Bray. http://en.wikipedia.org/wiki/Tim_Bray.

http://www.w3.org/People/cmsmcq/.

Theological Markup Language. http://www.ccel.org/ThML/ThML1.04.htm.

Tanakh ML. http://tanakhml2.alacartejava.net/cocoon/tanakhml/index.htm.

Open Scripture Information Standard. http://en.wikipedia.org/wiki/ Open_Scripture_Information_Standard.

Project Gutenberg. http://www.gutenberg.org.

Citizendium. http://www.citizendium.org/cfa.html.

No new XML languages. http://www.tbray.org/ongoing/When/200x/2006/01/08/No-New-XML-Languages.

http://www.unicode.org/charts/PDF/U0100.pdf.

http://www.unicode.org/charts/PDF/U0590.pdf.

Topic Maps. http://en.wikipedia.org/wiki/Topic_map.

?. http://www.ontopia.net/topicmaps/materials/tmrdf.html.

Information retrieval from annotated texts.. A.S. Fraenkel and S.T. Klein. J.. Amer. Soc. for Information Sciences. 50. 1999. 845–854.. http://www.wisdom.weizmann.ac.il/~fraenkel/Papers/annot.ps.

http://scripts.sil.org/cms/scripts/page.php?site_id=nrsi&item_id=XSEM.

http://books.chabadlibrary.org/default.aspx.

XML# ##### ######## HTML.

Implementation

Approach

##########

########, ## ####, ### ##### ############ ###, ### ####, ############. ###### ## ##### #######: ######## ############## ##### (### #######), # ############# - #####; #########, ########## ########, ############# ###### # ######## #####, ### ##### ######## ############ (####### ### # ###### ########); #### ### ######## ######## # #####, ### # ### ######### # #.#. ## ####### - ### ########, ###### ########## ########### ########## ##### #############, ### ####### "#######" #####, ### ##### ####### ##########.

######, ### ######## "## ########". ##### ###### ## ##### ## ### ############# #########. ## #### # #####-## ######## #### "###########" ########, #######, ### ############ ### ##### #####. ######## ## ##, ### ##-## "############" ########### ###### ########## # ### ######### ##########, # ##-## ############ ######## ############## "######### #####" # ### ##### #### # ## ########.

######, ###### ## ####, # ###########...

####, ### ###### ###### ######### # Unicode [53]. ############ #### ######### #########.

####, ### ###### ###### ######### # XML# [55] [1] , ######## ## ##, ### ## ## ######### ## ############# ########### ######## ###### ###### (##. ####). ### ## #####, ########### ####, "##########" XML #########.

TEI

XML#, ### #### [58], ##### ## ########### ##### ######## XML#, # ############### ##### ## #### "#######" [65].

XML# "############" (######### ## #########) ######## #### ######### (## ########## ###### # ##### "########"): ############# TEI (########## ########## #######) [6]. ###### #### ### ########## ########## ###### ## ####### XML# - ##### ########-####### [59]. ####, ### #### ## ###############.

(# ###### #######, ###### ## ######, ###### ###### ## ## ########## ### ########## #### ########: Theological Markup Language [60], TanakhML [61], Open Scripture [62], Project Gutenberg [63].)

#####

(###### 2.1). # TEI ######### ########## ####### #### ########## ############ ####### ###### [7]. ## ######### ##### ############ [9].

#########

######### - #####, ### ... - # TEI ####.

#########

(###### 2.3). ###### ### ######## ## ###### ###### ### ####### # ####### ######### # ######### #######. #### ## ################ ########, ## ######## ###### ##### ######## #### ###############, ### "##### ###### ######### ## ######". ##### ## ##### ###### ####### ### ### ############# ### ##### # ####### ######### "########" ###### ## ######, ## #### ###########. ## "#####" ###### - ####### ###### ##### "########" ######### ###### - ### ########## ###### ######, # ### ######### ##### ##############.

#####-###### ####### ##### ######## ##### ######### ## ##### # #####. ########, #### #########. ## ######### ## ## #### #######, ########### ## #########, ######### # #.#. ###### ####. ####, ############ TEI [6], ## ### #########, ######## ##### ####### ##### ############ [10], ############# ####### ###### [8], # ########## #######: [13] [14] [15] [16] [17] .

XXX ######### #########?

######, ########## # ##### - ###### ## #### (###### 2.1). #### TEI (# #### ######### ### XML-########) ####### ##-######### #, ###########, ##### ## #####. ###### ######## ### ########## ############### ######## # XHTML# ###, ##### ### ### # ###### #######, # ##### ### #### ## ############## ######### ####### Unicod#, ######## ########### ###### [54]. ###, ######, #### ######### ############## ##### ####### # ######### ######### (########, ########## XML). #### #### ####### ## # ### ##########, ### #####, ##### ############### ###########, ## ### ######, #########. # ### ############# ######### ######## ######### ########### - ##########.

##########: ######### ### ####### ########, ####### #### ###### <verse> # </verse>, # ##### ########## ##### #### ##### ## ######. ## ########### ## ## # ###############? ########, ## ######## ## ########## ###### ##### ## ####### ###### ##### # ####### ####### ########## ###? ## ######## ## ### #### ##### # ######## #######? # ###### ##### ###### ##### ########## ### - # ####### ### # #########?

######, ### #### #### #### ##### ## ######, ## ### ######## ##### ###### ######## - ## # ######, ### #### ## #######. #### ### #### ####### ######, ##### #### #### ## ######: ###### ##### ########### # ######## ########### ## #####, # #### # ######## ###### ##### TEI ######## ### ## ####### - # # ## ########. #### ## ########### ###### ########### ######### ######### ###### ###, ### ## ### ###### #### ########. # ######### ######### ######## ###### # ######### #########: ## ###### ######, ### ##### ### #######, ### ########### web-#########, ## # ######, ### web-#########, ############### ### #### TEI ### ## ########. # # ######### ###### ##### ###### ## ### #####.

######## ##, #### ##### # ##### ####### ##### ## ######, ## ### ### ## TEI? ## ### ## ####! TEI### ###### ######## ### ###############: ##### ####### #### ######### ######### ############## [11]. ######, # ### # ######### ###### - P5 - ############ #####

############ #####, ####### ########## ######### ROMA [12], ## #### ########.

######, ## ##### ####### ###### # TEI, ## ######### ########## ## # ###### ########, ###### # ########## #######. ###### ### # ######. ### ## #####, #########, ###### ########### ####, ##### ######### # ########## #######, ### ######### ##### ########### #### TEI.

######

########## ############ # #### ######## ######. ### # TEI ##### ######### ############ ####### # #######. ###, ######, ###### XLink (XPointer?) [57] ## ########: #### ########## ## Topic Maps [70] # RDF [71].

XML#

web-######### ###### ########### ######## ######### ############## ########## ## XML. ######### ###### #### ##########. ########: JXHTMLEDIT [41] (### HTML#), Syntext Serna [42] (#############). ######, ### ## #####, # #### #### ############# #### ###### (XML, TEI) # ########## ########## ###### 3.3 ####. # ###### ##### ######## ### Stylus Studio [43] # ALTOVA xmlspy [44] - ### ###### ### Windows. #### ######### ######### ## Jav# - # ###### ########### #####: oXygen [45], Exchanger XML [46], editix [47], Topologi [48]. ### ## ### ######### ###### ########## Oxygen, ### # ### #####.

#########

{Browser} {Tabs}

AJAX [49]. # ########## ## ### ######## ### ###########, # ### ########## ############### ######-###### ########## Java###### ######## [50] [51], ### ##... ### ############## ####### [52] # ######### # AJAX##. ######### ### ##########, ####### ######### ####### ########## ### ### 10.

Technology

XML Databases

It is possible to store the texts as XML files in the file system and use XSLT (as implemented by Saxon) to select requested pieces and transform them into presentation form. Indeed, I'll have a copy of all the texts in simple XML files anyway, since I need to check the texts into a revision-control system.

It seems likely, though, that I'll need to store the texts (also) in an XML database. Here are some requirements that make me think so:

- Access parts of documents in response to a query

- Fetch fragments of the documents referenced from a given one

- Find documents referencing a given one (link reversal)

- Full text search

Only first of these requirements can realistically be satisfied without some indexes. On the other hand, only first two are trivially satisfied by an XML database (like Exist). Integration between Lucene text indexing package and Exist needs to be looked into. As for link reversal, we'll probably have to write the indexer and accessor ourselves...

It is clear that a query language to be used is XQuery [http://isbn.nu/0321180607]. It is a nice, functional, non-statically-typed language, that have recently acquired update and text search capabilities. (XXX)

TEI### #### ########, ### #### ############ XML##### ###### ###### # XQuery [18].

########## # ######### XML#### ##### ###### ######## ? Bourret [http://www.rpbourret.com/]. ######## ########## #### ###### ### XML#:

- eXist [http://exist-db.org/]

- Berkeley DB XML [http://www.sleepycat.com/products/bdbxml.html]

- Sedna [http://modis.ispras.ru/sedna/index.htm]

- Timber [http://www.eecs.umich.edu/db/timber]

- MarkLogic [http://xqzone.marklogic.com/]

- Lucene [http://lucene.apache.org/]

XQuery

Some use XQuery as the (almost) only implementation language for the application (e.g., AtomicWiki [http://judaica.podval.org/moin/AtomicWiki]). XQuery *is* a functional language. But XQuery does not have static typesystem or exception processing. I will use Java (or Scala) as my main implementation language, and XQJ to access XQuery/XSLT processors.

XML and Java

There are APIs for

- parsing: javax.xml.parsers

- XSLT: javax.xml.transform

- XPath: javax.xml.xpath

- XQuery (XQJ): java.xml.query

javax.xml.xpath only supports XPath version 1

It seems that I can do pipelines using XQJ.

XML Pipelines

- Cocoon [http://cocoon.apache.org/]

- Pipelines [http://moinmo.in/FeatureRequests/PipelineArchitecture]

- XProc

- Calabash [http://fgeorges.blogspot.com/2008/10/poor-mans-calabash-integeration-into.html]

TEI: Versioning Machine Versioning Machine [http://mith2.umd.edu/products/ver-mach/] # <teiPublisher> [http://teipublisher.sourceforge.net/docs/index.php]. ### ###### #### # ## ## #### - Susan Schreibman [http://www.greenstone.org/cgi-bin/library] # Amit Kumar, # ### ########. ###### #### ############# eXist.

########, ## # ####### ### #########: ###, ##### # ###. ### #### ########### ######### # ###### #########, ########## #######, ############## ####### ####### ######### ###### # #.#. ## ##### #### ########## # #### ###### ####### version control system (XXX ## ########), # ###### - SubVersion [http://judaica.podval.org/moin/SubVersion]? Hg?.

#########, # ## ###### ##### web-#########. ########## ###### # ############### ###### #### ######## ######:

- WebDAV [http://www.webdav.org/]

- REST

- SOAP [http://en.wikipedia.org/wiki/SOAP]

- XML-RPC [http://en.wikipedia.org/wiki/XML-RPC]

- XML:DB [http://xmldb-org.sourceforge.net/]

- RSS [http://en.wikipedia.org/wiki/RSS_%28file_format%29]

- Atom [http://en.wikipedia.org/wiki/Atom_%28standard%29]

- SubVersion/Hg

  - ########## ###### # #####

  - ###### (#####) # ###### ############# (#####) ####### ########

  - Annotea [http://www.w3.org/2001/Annotea/]

- Collate/Anastasia

WebDAV

XML and Wiki

- AtomicWiki [http://code.google.com/p/atomicwiki]

- WikiModel [http://code.google.com/p/wikimodel]

- WikiModel [http://wikimodel.sourceforge.net/]

- XmlWiki [http://moinmo.in/XmlWiki]

- WikiXmlDtd [http://www.usemod.com/cgi-bin/mb.pl?WikiXmlDtd]

- DocBookWiki [http://doc-book.sourceforge.net/homepage]

- Single Source Publishing [http://www.cecc.com.au/cb_pages/publishing.php]

Java Content Reporitory

####### # ################ ######### (# ###-### ### :)), ########
Content Repository System.

Jav#: JSR 170.
- # ##### # ## ####. ######,
###
#######, ##### ######### ########### [http://www.ics.uci.edu/~fielding/pubs/dissertation/rest_arch_style.htm#sec_5_2_3] ### REST [http://rest.blueoxen.net/cgi-bin/wiki.pl?FrontPage] # #### ## ############
######### - # # ######### ######, ########## # ######## Day [http://www.day.com/site/en/index.html], ######### ##### ######### :). ####
########## ##########: Apache Jackrabbit.

How do I return XML documents as XQJ-valued properties? My stuff is going to be stored in XML database (or a bunch of files processed by Saxon). I

can not present them as files (because they may not be) that need to be parsed (especially when they are already parsed and indexed). So my only (standard) choice to glue the low-level store to Java is XQJ.

How do I project one hierarchy on top of another in JCR? Let's say I store files with texts in the repository. I need to expose the result (both through WebDAV and locally) as a hierarchy where I can drill into the documents' hierarchical structure (chapter, paragraph etc). I will do it through XQuery (likely XPath, but running within XQuery engine). Also, there may be multiple such structures, and before drilling in I will need to turn main structure into an alternative one using XSLT. I'd like the *external* view to have the same JCR API.

- At Wikipedia [http://en.wikipedia.org/wiki/Content_repository_API_for_Java]

- A Practitioner's Perspective [http://www.theserverside.com/tt/articles/article.tss?l=JCRPract]

- JSR 170 [http://jcp.org/en/jsr/detail?id=170]

- What is Java Content Repository? [http://www.onjava.com/pub/a/onjava/2006/10/04/what-is-java-content-repository.html]

- JSR 283 [http://jcp.org/en/jsr/detail?id=283]

- Version 2 [http://www.infoq.com/news/2007/07/java-content-repository-2]

#######:

- Apache Jackrabbit [http://jackrabbit.apache.org/index.html]

- Apache Lenya [http://lenya.apache.org/]

- Apache Graffito [http://incubator.apache.org/graffito/]

- OpenCMS [http://www.opencms.org/opencms/en/]

- DSpace [http://www.dspace.org/]

- Fedora [http://www.fedora.info/]

- Greenstone [http://www.greenstone.org/cgi-bin/library]

Content management Systems

- A list [http://java-source.net/open-source/content-managment-systems]

- http://www.day.com/site/en/index.html

- Daisy [http://www.outerthought.com/en/products/daisy]

- Daisy [http://cocoondev.org/daisy]

- Apache Lenya [http://lenya.apache.org/]

- Apache Forrest [http://forrest.apache.org/]

- Apache Graffito [http://incubator.apache.org/graffito]

- OpenCMS [http://www.opencms.org/opencms/en]

- DSpace [http://www.dspace.org/]

- Alfresco [http://www.alfresco.com/]

- Magnolia [http://www.magnolia-cms.com/]

XForms

- Chiba [http://chiba.sourceforge.net/]

- Orbeon Forms [http://www.orbeon.com/]

- Firefox Extension

- FormFaces [http://judaica.podval.org/moin/FormFaces]

Tagsoup

URLs

XPointer in the URI, not in the fragment! No delimiters, just URI parts - which can be implicit (not "chapter=3", but "chapters/3", or just "3")! Editions in the

URI ("Chumash/boston+toronto/Genesis")! Metadata ("about"), raw XML etc. - in the URI, not as query parameter ("Genesis/about", "chaters/1/raw")! More URI promotion: natural references ("Genesis/2:1", "Genesis 2:1")! Intervals ("Genesis/2:1-3")! Concatenation ("Genesis/2:1-3;5") probably shouldn't be done through URIs!

Books URIs:

/books/Tanach/editions/.../[parts/.../]books/.../[weeks/...]/chapters/.../verses/...

editions: a | a+b (side-by-side) | a-b (differences)

parts: Torah | Neviim | Ksuvim

books: Genesis | Ionah | ... (appropriate for part if present)

weeks: Genesis | Noah | ...

chapters: n | m-n

verses: n | m-n (can be present only if one chapter is selected)

Alternative names may be used.

URL may be truncated.

Parts of the URL may be implied - and need to be derived.

Metadata

Metadata is used to:

• guide navigation

• provide listings and names

• create classifications (links)

• stitch together data directories

- store application-specific metadata

Some of the data in it has to be duplicated in the text document (for self-containment, *and* for non-position-based navigation).

We need to be able to handle things like "Chumash/books/Genesis/weeks" and "Chumash/weeks" with one metadata document...

Locators for the navigational steps can be: - subdirectory/file - element XPATH - milestone XPATH

1) I need to be able to provide a list of selectors (book name/ chapter #/ verse# etc.) on any level.

2) A selector can have multiple names, which I do not want to duplicate (and maintain) in each edition of the text. So, selector names have to be part of the metadata.

3) A text can have multiple structures.They are important for the metadata also. Restructuring of the text is done by XSLT. It seems logical to use the same for the restructuring of the metadata.

It follows that the metadata needs to be processable as XML (and have format similar to the texts). Do I also need it to be processable (in part) as Java objects (using JAXB) - is not clear.

We are going to use milestones [?] to represent multiple structures.

```
<book n="Genesis">

  <chapter n="1">

    <week n="Genesis" milestone="begin"/>

    <paragraph type="open" milestone="begin"/>

    <verse n="1">

    ....

    </verse>
```

```
    </chapter>

    ...

    <chapter n="6">

       <verse n="1">

       ...

       <week n="Noach" milestone="begin"/>

       <verse n="..">

       ...

    </chapter>

</book>
```

Tanach Markup

What are the TEI-appropriate tags for Tanach?How do we represent the paragraph in the middle of the verse?

Super-Wiki

Wiki with multiple formats => function reversal (TEI->HTML; edit; back)...

Wiki page rename and links correction - if the wiki itself is in an XML database (like AtomicWiki [http://judaica.podval.org/moin/AtomicWiki]) *with* our link-reversal index, wouldn't it be easier? History will be kept by the revision-control system...

Navigation:

- expand/contract viewport

- move viewport

- switch to a different structure preserving focus (from "lesson" to "chapter" in Tanya, for instance)

- switch to a different edition / look around at editions

Internal Wiki Markup

<section level="" milestone=""/> 1 2 3 4 start

<list type=""> bullets numbers

- <item>

<table>

- <row>

    - <item>

<paragraph>

<blockquote>

<line>

<style type=""> emphasize bold italics underline strikethrough

<anchor id="">

<wikiLink relative="" reference="">

<interWikiLink name="" refererence="">

<urlLink reference="">

<include> image?

Notes

crowd-sourcing TEI files [http://comments.gmane.org/ gmane.text.tei.general/7031]

Web-based IDE with WebDAV's versioning

BUGS

Upstream:

- http://sourceforge.net/tracker/index.php?
  func=detail&aid=2056090&group_id=17691&atid=117691 exist resolve-
  url

- http://xmlroff.org/ticket/131 xmlroff tables (fixed)

Sebastian:

- File a bug against FO stylesheets (title, table of contents).

- File a bug about reference shape consistency.

- File a bug about use of @name for reference.

Saxon, Tomcat and relative URIs for the stylesheets. XQuery Server Pages
(and eXist).

space before a word that has read/write annotations (Psalm 60)

Styles of biblio references.

Google SSO. GData. RSS/Atom - second edition? Hacking...?

Start working on XSLT: Genesis -> FO

leningrad-import:

- remove stylesheet link

- add TEI P5 All declaration; namespace(s)

- makaf

XProc

Discussions as text.

Convince CiteULike to make their XHTML really XHTML, or at least - well-formed XML. Better - parse RIS.

Laying out classic Jewish texts

It is natural for a user, after researching with our system, to desire to print selected texts and fragments for personal - or group - study away from a computer. Such printouts are one-use artifacts. It is clear that ability to produce such printouts must be present in the system from the beginning. The question is: how good the typographically does it need to be?

We need to format a tree of texts: main one, commentaries of it, commentaries on commentaries etc. It is known about each piece of commentaries what is it commenting on. All the font metrics are also known: glyph sizes, what is haging how low and what is sticking up and how high. Result needs to be readable and (is it a separate requirement?) beatyfull.

To format "like in a book", we need to optimize the following contradicting constraints (the list is probably incomplete):

• the page must be fully covered with print

• comment must start on the same page where what it comments on is

• comment must end on the page it started

Koritz says that we do not need to print books, but "leaflets" instead: text with comments that fit on one page. In the "forum format", whatever that means.

Gorkin says that printing "like in the book" of the multi-layered text is extremely challenging typographically, and thus very interesting, but design of the overall interface of the system is even more interesting - and difficult. And more importants. Also, what exactly are the requirements for the printing facility, and what is their order of importance, will become clear only in the process of using the system. So, initially printing needs to be acceptable, but primitive - we do not have resources to do fancy stuff from the beginning.

Dubinsky says that the format that will "grow" from the use of the system, will turn out to be a familiar to us all format "like in the book", or so close to it, that a solution for one will fit the other; that good leaflet is not easier to print than a book; and that ability to print familiar "book-like" format is neccessary

for the psychological comfort of the users. But he also agrees that features and interface of the system are more important.

Thus, everybody agrees that initial printing facility will be "primitive". Gorkin does not want to expend any effort to even find out how primitive. Dubinsky would like to see something acceptable. Nothing of the sort has been found so far. XSL-FO [7] is insufficiently expressive for our problem - even version 1.1, it seems.

Beyond   Pretty-Printing: Galley   [http://lambda-the-ultimate.org/node/2419] Concepts in Document Formatting Combinators

Nonpareil [http://www.it.usyd.edu.au/~jeff/nonpareil/]

iText [http://www.lowagie.com/iText/]

XSL-FO   2.0   Requirements   [http://www.w3.org/TR/2008/WD-xslfo20-req-20080326/]

Sources of Free Texts

Jumping points

- Wikipedia Torah database [http://en.wikipedia.org/wiki/Torah_database] - done

- Wikisource Judaica Bookshelf [http://he.wikisource.org/wiki/%D7%90%D7%A8%D7%95%D7%9F_%D7%94%D7%A1%D7%A4%D7%A8%D7%99%D7%9D_%D7%94%D7%99%D7%94%D7%95%D7%93%D7%99]

- psychomystic [http://psychomystic.blogspot.com/search/label/Torah%20Online%20Links%20Database] links - done - closed access

- Chabad Library [http://chabadlibrary.org/books/]

- Sichos Kodesh [http://www.sichoskodesh.com/] - empty

- Otzar 770 [http://www.otzar770.com/]

- hebrewbooks.org [http://www.hebrewbooks.org/]

- chabadlibrarybooks.com [http://www.chabadlibrarybooks.com/]

- Seforim Online [http://www.seforimonline.org/]

- Grimoar [http://www.hebrew.grimoar.cz/] - Kabbalah

- jewishcontent.org [http://www.jewishcontent.org/] - for PDAs

- Torah Texts [http://www.torahtexts.org/]

- chassidus.ru [http://chassidus.ru/rambam/index.php] - broken

- Halacha Brura [http://www.halachabrura.org/alephlink.htm]

- Digitized Book Repository (JNUL) [http://www.jnul.huji.ac.il/dl/books/html/bk_sub.htm] - broken

- Otzar HaHochma [http://www.otzar.org/otzaren/indexeng.asp]

Texts

- Tanach (Leningrad Codex) [http://www.tanach.us/Tanach.xml]

- Mishna [http://chaver.com/Mishnah/TheMishnah.htm]

- Targumim [http://cal1.cn.huc.edu/index.htm]

- Midrash Raba [http://www.tsel.org/torah/midrashraba/index.html]

- Midrash Tanhuma [http://www.tsel.org/torah/tanhuma/index.html]

- Yalkut Shimoni [http://www.tsel.org/torah/yalkutsh/index.html]

- Ovos DeRabi Noson [http://www.tsel.org/torah/avotrnatan/index.html]

- Sefer HaHareidim [http://www.daat.ac.il/daat/mahshevt/kitsur/tohen.htm]

Wikisource

- ... and Mechon Mamre [http://he.wikisource.org/wiki/
  %D7%A9%D7%99%D7%97%D7%AA_
  %D7%95%D7%99%D7%A7%D7%99%D7%98%D7%A7%D7%A1%D7%98:
  %D7%95%D7%99%D7%A7%D7%99%D7%98%D7%A7%D7%A1%D7%98_
  %D7%95%D7%9E%D7%9B%D7%95%D7%9F_%D7%9E%D7%9E
  %D7%A8%D7%90]

- Tanach [http://he.wikisource.org/wiki/%D7%9E
  %D7%A7%D7%A8%D7%90]

- Mikraot Gdolot [http://he.wikisource.org/wiki/%D7%9E
  %D7%A7%D7%A8%D7%90%D7%95%D7%AA_
  %D7%92%D7%93%D7%95%D7%9C%D7%95%D7%AA]

- Targumim [http://he.wikisource.org/wiki/%D7%AA
  %D7%A8%D7%92%D7%95%D7%9D]

- Mishna [http://he.wikisource.org/wiki/%D7%9E
  %D7%A9%D7%A0%D7%94]

- Tosefta [http://he.wikisource.org/wiki/%D7%AA%D7%95%D7%A1%D7%A4%D7%AA%D7%90]

- Masechtos Ktanos [http://he.wikisource.org/wiki/%D7%9E%D7%A1%D7%9B%D7%AA%D7%95%D7%AA_%D7%A7%D7%98%D7%A0%D7%95%D7%AA]

- Mechilta [http://he.wikisource.org/wiki/%D7%9E%D7%9B%D7%99%D7%9C%D7%AA%D7%90]

- Sifro [http://he.wikisource.org/wiki/%D7%A1%D7%A4%D7%A8%D7%90]

- Sifri [http://he.wikisource.org/wiki/%D7%A1%D7%A4%D7%A8%D7%99]

- Midrash Rabba [http://he.wikisource.org/wiki/%D7%9E%D7%93%D7%A8%D7%A9_%D7%A8%D7%91%D7%94]

- Talmud Bavli [http://he.wikisource.org/wiki/%D7%AA%D7%9C%D7%9E%D7%95%D7%93_%D7%91%D7%91%D7%9C%D7%99]

- Talmud Yerushalmi [http://he.wikisource.org/wiki/%D7%AA%D7%9C%D7%9E%D7%95%D7%93_%D7%99%D7%A8%D7%95%D7%A9%D7%9C%D7%9E%D7%99]

- Rif [http://he.wikisource.org/wiki/%D7%A8%D7%99%22%D7%A3]

- Rambam [http://he.wikisource.org/wiki/%D7%9E%D7%A9%D7%A0%D7%94_%D7%AA%D7%95%D7%A8%D7%94]

- Tur [http://he.wikisource.org/wiki/%D7%90%D7%A8%D7%91%D7%A2%D7%94_%D7%98%D7%95%D7%A8%D7%99%D7%9D]

- Shulchan Oruch [http://he.wikisource.org/wiki/%D7%A9%D7%95%D7%9C%D7%97%D7%9F_%D7%A2%D7%A8%D7%95%D7%9A]

- Kitzur [http://he.wikisource.org/wiki/%D7%A7%D7%99%D7%A6%D7%95%D7%A8_%D7%A9%D7%95%D7%9C%D7%97%D7%9F_%D7%A2%D7%A8%D7%95%D7%9A]

- Oruch HaShulchan [http://he.wikisource.org/wiki/ %D7%A2%D7%A8%D7%95%D7%9A_ %D7%94%D7%A9%D7%95%D7%9C%D7%97%D7%9F] and? [http:// he.wikisource.org/wiki/AHS:OCH]

- Shulchan Oruch HaRav [http://he.wikisource.org/wiki/ %D7%A9%D7%95%D7%9C%D7%97%D7%9F_ %D7%A2%D7%A8%D7%95%D7%9A_%D7%94%D7%A8%D7%91]

- Siddur Tora Or [http://he.wikisource.org/wiki/ %D7%A1%D7%99%D7%93%D7%95%D7%A8_%D7%AA %D7%95%D7%A8%D7%94_%D7%90%D7%95%D7%A8]

Texts in English

- Babylonian Talmud: Soncino [http://www.come-and-hear.com/talmud/ index.html] Rodkinson [http://www.sacred-texts.com/jud/talmud.htm]

- The Guide for the Perplexed [http://www.sacred-texts.com/jud/gfp/ index.htm]

- Shulchan Aruch [http://www.torah.org/advanced/shulchan-aruch/]