

Summary of the second German VuFind Community Meeting

Day 1: 2013-Sep-16

After a short introduction from Inken Feldsien-Sudhaus (director of the University Library of the University of Technology Hamburg, TUBHH) and Oliver Goldschmidt (TUBHH) the program starts.

Heiko Weier (TUBHH) features the first session dealing with VuFind and indexes.

First speaker is Gerald Steilen (Service Center of the German Common Library Network VZG GBV), who shows some facts and figures about how the index from the Common Library Network is being used. Several members of the Common Library Network are using this index and the Service Center also offers a hosting service for VuFind, which uses the index also. The index contains round about 100 Mio records. Gerald shows, that some facet data can be very confusing, because the data in the index contains records from different sources and the handling of some fields is pretty different in each of these sources. For example it is very different, how many of the records have a language attribute or a certain classification. This makes it hard to build facets in VuFind: most of them will be incomplete and using them would make the research recall very inefficient (records without this attribute will never be displayed).

Second speaker is Bettina Sunckel (Library Network Hesse HeBIS). She speaks about their VuFind environment. HeBIS has developed a special ILS-driver based on the PICA driver, but covering some special requirements of PICA in Hessen. Also, they are using the DAIA driver, the additional Tab for EDS and Shibboleth authentication. HeBIS offers a VuFind service for their members and are currently building some new VuFind instances (for the university libraries in Mainz and Marburg). Two Instances in Frankfurt and Kassel are productive. The development model for the vufind-based HDS (HeBIS Discovery System) is cooperative: Besides the developers at HeBIS, developers work in different member libraries (university libraries Frankfurt and Kassel) and in a DFG funded research project.

After a short break the program continues with four parallel workshops:

1. VuFind 2.0 and the search system (David Maus, HAB Wolfenbüttel)
2. Optimizing Solr index structures and analysis (Uwe Reh, Library Network Hesse HeBIS)
3. Relevance ranking and query type detection (Hajo Seng, University and State Library Hamburg)
4. Developing drivers for VuFind (Oliver Goldschmidt, TUBHH)

Summaries of the workshops:

Workshop 1:

David presents and explains the differences of the design between VuFind 1.x and 2.x. He explains, how a query is built and performed in VuFind.

Workshop 2:

Uwe explains, how the Solr index is structured, how query analysis works and how one could use the features of Solr. He shows for example phonetic search, special characters like Umlauts, stemming, elevation, ... He also explains, how queries like C++ can work and how EDismax works.

Workshop 3:

Hajo explains technical details about how relevance is being calculated in VuFind and how it can be configured. He shows the relation between relevance ranking factors and the

metadata in the index.

Workshop 4:

Oliver shows, what different types of drivers are existing in VuFind (ILS-drivers and RecordDrivers) and what functions they are covering. RecordDrivers are allowing to use record types different to MARC records in VuFind. In TUBfind this is used to integrate Websites and blog posts from the library into the result list and show them differently than MARC records. The group discusses, what other media types or record types could be integrated in VuFind. Some examples taken from that discussion are:

- Moodle content (Courses, Discussion groups, ...) (Moodle is an E-Learning environment)
- Video platforms (or even youtube content or other multimedia content)
- more data about persons or authors (taken from a separate index)
- maps (2D or 3D)

For each of these record types one could construct a new driver and create a template to display this content. Before that, it would be necessary to get the content into an index first. This can be done using the VuFind importer (using XSLT to transform the input file) or by creating an XML file, which is compatible to SOLR directly.

Day 2: 2013-Sep-17

The next day startsoff with a VuFind wishlist session moderated by Gerald Steilen (VZG GBV). Most suggestions from last year's wishlist session in Leipzig were transformed into JIRA tickets by Oliver Goldschmidt (although some of them are still missing since list had not been completed).

The wishlist from last year contained the following things:

Search functions

- Browse Bookshelf
- multilingual search
- personalized preferences

result list

- Show on shelf (map of the library containing book position)
-> some libraries have integrated BibMap showing a floor plan of the library
- Improve visualization of results
- OR-facets (values out of one facet should be selected and searched as an ORed facet)
- handling of series or multi-volumed records
- Recommendation Services
-> some libraries are using Bibtip (e.g. FH SWF)
- FRBRisation (similar records associated with the current one should be visible in the result list)
- export only selected titles from the result list
- factor popularity of a title into relevance ranking

Administration

- allow to configure a fallback index for VuFind, in case the main index fails
- verify email-addresses if the email address of a user changes
- a usable admin interface to configure VuFind
- restrict access to several functions (like sending email; this should be only allowed after login, or licenses from records to suppress certain records in result list, if it is not available to the one who searched)

- an interface to allow plugins for VuFind

Some of these points have been resolved in the meantime, some of them are addressed in JIRA tickets.

Wishlist 2013

1. VuFind FAQ / Manual / Overview on the structure

It could be useful to have an overview on the structure of VuFind, like an administration manual or an explanation of the tree structure and the files/classes/namespaces

David Maus explains, that a documentation on file level won't be very useful, but a documentation of namespaces could be. He suggests using the X-Debug-Extension for PHP to get an overview about what classes or methods are used at all.

But it's not practical for the developers to describe everything they do in detail; it could be better to answer concrete questions (e.g. questions about how to migrate from VuFind 1 to VuFind 2).

The community is planning to collect questions in the VuFind Wiki, that can be basis for an FAQ about VuFind.

David Maus mentions that VuFind 2.2 is scheduled for summer 2014 and currently a feature collection is going on.

In Germany we are missing an organization or foundation coordinating the developments here. Mr. Stäcker (director of HAB Wolfenbüttel) made a move to introduce something like a foundation, that could do that. He is still looking for (library) partners to initiate and finance the project.

2. Improving the search system

David Maus would like to hear about ideas for improving the search system, for example in the design of an EDismax Handler. He offers to operate as a coordinator and collector for that.

3. Using Responsive Design in VuFind.

Oliver Schihin is willing to act as a coordinator or contact for this case for the next weeks and says, the Basel University will continue working in that area.

4. Improving Tagging functions

There was some discussion on this topic. To improve tagging, it's suggested to use recommendation services to make it easier to tag records. Nevertheless most discussion participants agreed, that tagging is not used very often and we should not spend too much work in functions no one needs.

Presentations: "Best practice with VuFind"

In this block we saw some presentations of VuFind instances in Germany and Austria.

Stefan Winkler reports about the BOSS system (BSZ One Stop Search), which is offered by the Library Service Center (BSZ) in Konstanz. He explains why they decided to include the EBSCO index instead of other commercial index solutions (it was the result of a survey). EBSCO plans to develop a module to include their index for VuFind, that should be included in the VuFind trunk. BOSS features dealing with authority data by using a synonym file in Solr. Doing that, synonyms of person names can be respected while searching. The

University Library Leipzig is using authority data for some time yet, as **Björn Muschall** explains, but this is based on a different system: they are indexing all the authority subject terms with the records. Its planned to switch to Solr 4 for BOSS (currently its still Solr 3) and next year to migrate to VuFind 2. BSZ is currently working on a Pazpar2 metasearch integration into BOSS. Stefan also shows plans for a cooperative index, that should contain all of the different german library networks. <https://wiki.bsz-bw.de/doku.php?id=projekte:boss:start>

Sebastian Böttger (University of Kassel) presents the social bookmarking and publication-sharing system PUMA (clone of BibSonomy). It aims to integrate the features of bookmarking systems as well as team-oriented publication management. A module for VuFind integrates PUMA as the favorite list of VuFind with advanced tagging functions. Extended export functions allow to export stored records in many formats (like BibTeX, Endnote, etc.). The PUMA module for VuFind is integrated in HDS (the VuFind environment of HeBIS). <http://puma.uni-kassel.de/>

Christian Dabrowski (University of Applied Science South-Westphalia) presents the marketing strategy, the library launched within the university one year prior to the introduction of their VuFind instance KAI. The name KAI is easy to remember and was found in a name-contest by the university staff. They did a lot of advertisement for the new catalog by postcards and other means. The traditional Aleph Opac was removed completely out of focus in favor of KAI. <http://kai.fh-swf.de/>

Jessica Drechsler (Georg-Eckert-Institute for InternationalTextbook Research) demonstrates their VuFind catalog. A particular problem with school books is, that they often have equal titles and are hard to distinguish in the result list. Therefore special codes are used for faceting.

Philip Kahle (University Library Innsbruck) demonstrates a VuFind implementation project. The VuFind catalog of Steiermärkische Landesbibliothek contains heterogenous data mainly based on digitized catalog cards, which are read by OCR. The result was verified by a automated query in other catalogs. http://webapp.uibk.ac.at/alo_cat/startpage.jsp

Hannah Ullrich (University Library Freiburg) also demonstrates their VuFind system. Hannah mentions that they are using the DAIA driver, but had to extend/modify it, because it did not cover their requirements completely (e.g. they needed more than one DAIA source and the driver made some unnecessary requests. <https://rds-ui.ub.uni-freiburg.de/opac/>

Lightning Talks

1. Uwe Reh: "DAIAinfo" as an alternative to Screenscraping and direct database access

- trying to get information about availability out of PICA
- problem: i.e. closed branch-offices(?) are calculated directly via OPAC, so the database is not always "consistent"
- invented own microformat "DAIAinfo"
- Bibmap @ UB Frankfurts OPAC bound to callnumber, thus "DAIAinfo" is integrated to that field, but hidden for users
- "DAIA-Info" is readable via screenscraping from OPAC and is base for the DAIA-server in vufind

- at the end this microformat inside the PICA-catalog is invented to get a more solid screenscraping and to tell vufind the actual media location (bibmap-information)
- see <https://lbsopac.rz.uni-frankfurt.de/PRS=daia/DB=30/SET=3/TTL=1/SHW?FRST=19> how DAIAinfo is integrated

2. Oliver Schihin: "swissbib goes VuFind: A quick glance at the development with VuFind 2.x, our own themes and modules as well as external tools"

- Development View (not always working): <https://testvf.swissbib.ch>
- Prototype of Swiss national catalog swissbib.ch based on VuFind 2
- First step was to port the swissbib-Layout to VuFind (up to now on OCLC TouchPoint)
- Solr Index now used for VuFind was originally designed for TouchPoint
- Project management via github <https://github.com/swissbib>
- swissbib started with VuFind 2
- Index includes parent-IDs for presenting multi-volume works; multi-level hierarchy is presentable as well

3. Daniel Zimmer: Visualizing search results: just nonsense or are there any usefull use cases for that?

- D3JS (<http://d3js.org/>) JavaScript Library to visualize search results; compatible with JQuery (meaning it can be combined without problems)
- Examples:
 - pie chart on the apportionment of subject headings (data is taken from the facets)
 - Top20 facets are displayed in a pie chart
 - Problem: facets are incomplete (see Gerald's presentation)
 - Though visualization can be annoying or confusing
- Histogram with the results with the corresponding creation years
- Comment Ursula Schulz: visualization is a desire of librarians, but not necessarily by the users also. Studies prove, that visualization is probably overrated
- Comment Steilen: A study from Peter Kostätt proves, that mobile interfaces are just optional; they are only needed for location based services (show location of a book), but not necessarily for the query

4. Ursula Schulz: Usability vs. Quality or consequences after Gerald's presentation (see above)

- Knowing statistical data about the coverage of facet building fields is very important because we need to know the quality of our catalogue service
- Incomplete facets blast recall without the user being able to be aware of this loss of information.
- Feasible consequences:
 1. Waiving incomplete facets so that those data fields with good coverage result in filters with reliable quality. A short term solution.
 2. Enriching poor field coverage with automatic procedures like matching with WorldCat records and automatic language recognition. This needs to be done by the service centres of the German library networks; said service centres must be equipped with appropriate man power. A medium-term solution.

- 3. Waiving library discovery systems altogether as tools like Google Scholar and amazon are already being used for discovery by most of the academic community. Instead libraries can focus on their functions as delivery services and learning spaces.
- Offering facets only because the data happen to be there is not a meaningful service. Instead cataloguing rules must provide data that users need and ask for. One example is the need for a facet that allows filtering resources according to target groups and level.
- Comment Timm Martin Siewert, ExLibris: Swinburne University of Technology (<http://bit.ly/16hjQGC>), Australia, did research into the use of facets. Patrons used only 3 facets out of 10. As a result only these 3 facets are shown.
- Comment Gerald Steilen: DS can only be interim solutions.
- Comment Uwe Reh: Black-Box-Systems like Google have the same problem but they remain invisible to the user.

5. Marius Zierold: VuFind from the perspective of smaller projects. Opportunity or burden?

- <http://gendervu.linuxd.org>
- "Black Hole before using VuFind"
- What do you do if you don't have MARC records?
- Frequent problem in Humanities: Hierarchies!
- Summary-tab contains a list of essays in this work; from the essay-record you return to the parent title.
- Experience: If very small libraries offer good quality, people come to the library after all, instead of Googling.
- Small libraries often have no OPAC, VuFind is actually pretty useful, but for mini-libraries the operating costs are too high.
- Opportunity: Yes, but often also a burden due to necessary financial and human resources.

6. Uwe Reh: "explain.solr.pl" Kurze Vorführung, des Tools zum Verständnis des Relevanceranking

- Solr parameter *explain* explains the relevance ranking of a Solr result list (how the relevance was calculated exactly)
- `explain.solr.pl` takes a Solr result list XML (completely with `explain=true`, just copy and paste) and makes the result more readable and even visualizes the explanation by drawing some charts about the result
- It explains, why a document is in the result list and which factors had an impact on the relevance score

7. David Aumüller: Factual search on RVK

- RVK notations for single records in the index / the MARC-data available.
- Complete path of the RVK-classification covered in Solr-index.
- It can be browsed by search onto the RVK-field (on any level of hierarchy).
- <https://katalog.ub.uni-leipzig.de/vufind/Search/Rvk>

8. Gerald Steilen: Solr 4 Index

- Use of GBV-Index has increased sixfold in short time

- Traffic resulted in performance decrease
- Solution: Change of infrastructure: redundant and distributed data holding
- Garbage-Collection may result in a totally unusable index system
- Solr-Cloud version 4.4 distributed on 5 servers