Application Note

A MediaWiki extension to open remote access to literature data from Pubmed

Andreas Bohne-Lang[1*], Elke Lang[2]
1. Medical Faculty Mannheim of the University Heidelberg, Ludolf-Krehl-Straße 13-17, D-68167 Mannheim, Germany, bohne-lang@medma.uni-heidelberg.de
2. Media Faculty, University of Applied Sciences Darmstadt, Haardtring 100, D-64285 Darmstadt - Campus Dieburg, elke.lang@fbmedia.h-da.de

Abstract
MediaWiki is one of the most popular wiki applications in the world, mostly known from the free encyclopedia WikiPedia. One of the success factors of this software is its extensibility. The following application note describes the 'Pubmed Extension' (http://www.mediawiki.org/wiki/Extension:Pubmed) that adds the new tag <Pubmed> to the wiki markup language and allows remote access to literature data from the Pubmed database. Literature Data can be included automatically into the wiki page by using layout template files.

Introduction
Wiki is a Hawaiian word for 'fast', and the underlying idea of a wiki program is fast and easy alteration of a web page, ensuring that the information on the page is always up to date. The most famous example of a wiki is the free encyclopedia WikiPedia [1] based on the MediaWiki software [2]. Today, companies and educational organizations use wikis as collaborative software in their intranets, and furthermore they use wikis beyond the traditional wiki functionality as content management systems to present static information to Internet users. ArrayWiki [3,4] is a nice example that allows sharing public microarray data repositories and meta analyses. Using wiki systems invokes several incontrovertible advantages: The wiki programs normally come along with a dedicated simple layout language and a "WYSIWYG" ("What You See Is What You Get") editor, enabling unexperienced users to edit pages without any need of detailed knowledge in HTML . Web pages can easily be edited using any web browser. The second advantage is the powerful extensibility [5] facilitating to add new features or enhance functionality of the main MediaWiki core. Sometimes it is necessary to include literature data within a wiki page. Pubmed [6], the largest article collection in the world with about 18 million article references collected from about 5000 journals [7], can be used in the biomedical area. Pubmed is one of several services provided by the U.S. National Library of Medicine, storing articles back to the 1950ies. Up to now there was no way to include Pubmed data automatically into a MediaWiki page.

Meterial and Methods
The MediaWiki software provides several extension types that can be categorized according to the programming techniques: Subclassing, hooks, tag-function associations, and Ajax [8]. MediaWiki software is entirely programmed in the popular and easy-to-learn programming language PHP [9]. The dynamic design of the MediaWiki core code allows registering new functions opening by this way new functionality to the whole system.

The described Pubmed extension [10] takes advantage of hooks and tag-function associations on the wiki software part. The tag extensions can possess attributes controlling the generation

of HTML code. The example <pubmed limit=10> … </pubmed> illustrates the use of this parameters.

The Pubmed wiki extension bases on a SOAP class (a protocol for exchanging XML-based messages over computer networks, normally using HTTP/HTTPS, successor of XML-RPC) programmed in PHP. It contains methods to connect to the Pubmed API functions that are listed in [11]. The extension program code only uses the Esearch and Efetch function from the API interface. Briefly: Esearch primarily searches and retrieves IDs (for use in EFetch, ELink, and ESummary) and Efetch retrieves records in the requested format from a list of one or more primary IDs or from the user's environment.
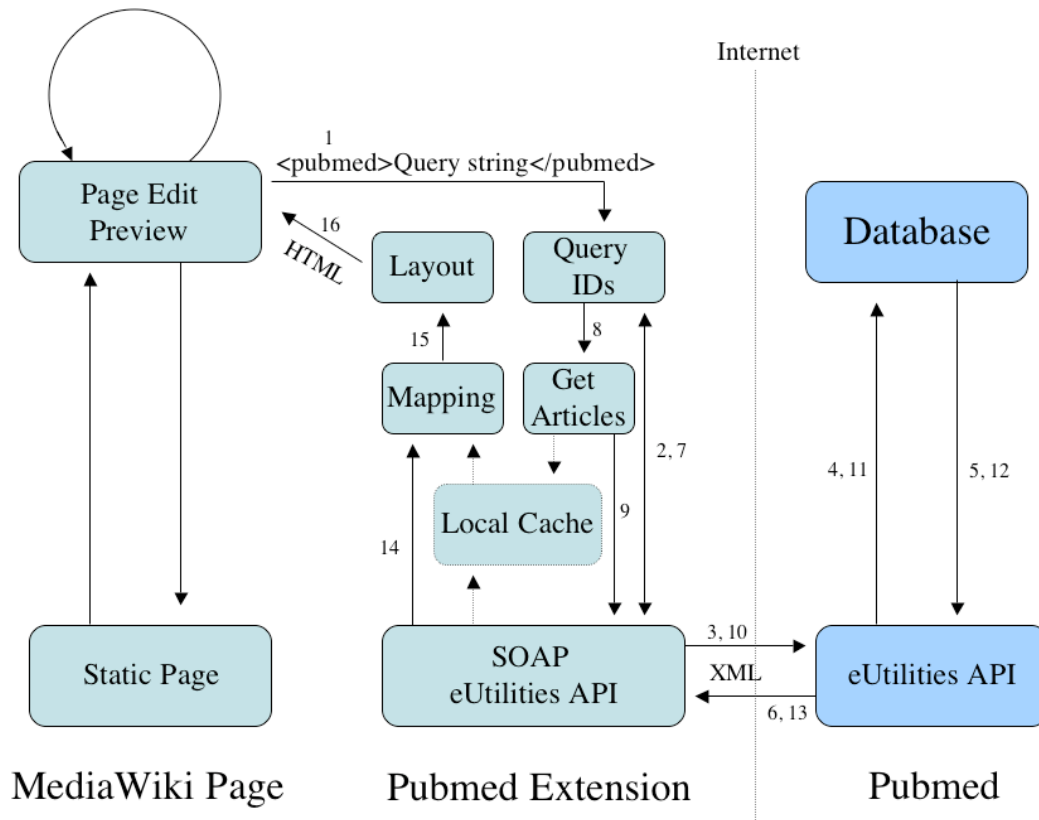
Figure 1: Workflow of the extension (without enabled local cache)

The extension is invoked by the <pubmed>-tag and transmits a query to Pubmed using the described method. The result set consists of the Pubmed article IDs. If the local cache is enabled the program tries to fetch the article data from former hits stored on the local disk space. Otherwise they are directly loaded from the Pubmed server. A Pubmed request causes the server to send the answer that is a large set of information in a XML record. In a next step the algorithm extracts the necessary parts by mapping. This step is necessary because some data within the XML result are unfortunately not unique like the 'Forename' or 'Firstname' attribute in the author definition. The following step is the layout process building the HTML code from the mapped data using a HTML template file.

In addition to the primary literature data some links are generated and included. At the moment there are three default links: A link back to the Pubmed page by the PMID, a link to the journal on Worldcat.org by ISSN and a link to the full text at the publishers page if DOI information is available.

Example layout file:

```
<p style=" border:1px outset gray; margin:.5em; padding:.5em; background-color:#efd;">
<i>##Authors##</i>  <br>
<b>##Title##</b> <br>
##Journal##: ##Year##, ##Volume##(##Issue##);##Pages## <br>
##PMID## ##WORLDCAT## ##DOI## ##Special##
</p>
```

1) In the wiki editor: <pubmed>17052341</pubmed>
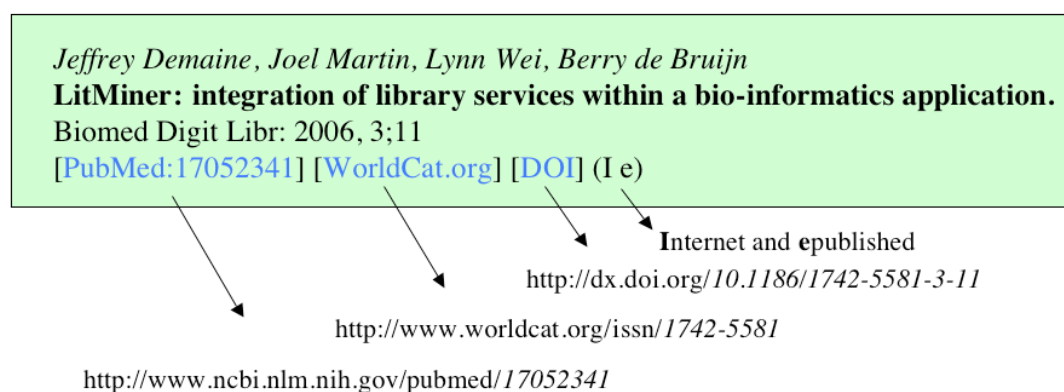
2) In the wiki page:

Jeffrey Demaine, Joel Martin, Lynn Wei, Berry de Bruijn
LitMiner: integration of library services within a bio-informatics application.
Biomed Digit Libr: 2006, 3;11
[PubMed:17052341] [WorldCat.org] [DOI] (I e)

Internet and epublished
http://dx.doi.org/*10.1186/1742-5581-3-11*

http://www.worldcat.org/issn/*1742-5581*

http://www.ncbi.nlm.nih.gov/pubmed/*17052341*

Figure 2: Example of the layout and HTML generation

Result
The MediaWiki provides a new tag within the markup language. The <Pubmed> tag provides a powerful connection to Pubmed by submitting the value between the opening and closing tag as a query string to Pubmed. A complete search field description can be found at [12]. Thus it is possible to select single articles by the PMID like this:
<Pubmed>6418270</Pubmed> or a complete list <Pubmed>6418270, 7819986, 15912238</Pubmed>. In addition, the query can be a complex search string like this:
<Pubmed> ("bohne-lang"[au]) and (free full text[sb]) </Pubmed> that returns all free full text publications of the author Bohne-Lang.
The <Pubmed> tag can possess three attributes that influence the layout of the result set. The attribute ‚limit’ controls the number of result records that the layout generator returns. The natural limit of the NCBI utilities is by 500 result records per query. The parameter should be used with complex queries – maybe to present the first 20 hits. The two other attributes ‚layoutfile’ and ‚layoutlinks’ can set the layout to different layout template than the default one. Two other pieces of information about publication status are added in brackets – Information about the publication medium and status (P for paper or I for Internet) and information about publishing status (p for published, a for ahead of print and e for epublished).

Discussion
Due to the fact that wikis are more and more used to present department or institute information to users, the extension can help presenting biomedical literature much more comfortable. Wikis and other Web 2.0 technologies are in the main stream of presenting

information to users [13]. Due to the success of the free MediaWiki software and its extensions, wikis are more and more used as simple content management systems, allowing only special users (community?) to alter the wiki page. Thus scientists can easily create publication lists – maybe of their own publications, for courses or project pages. The administrator can setup own HTML template files for the layout procedure and can provide them to the users. They can use the generic code or can switch to one of the others. The included default layout file links the WorldCat.org by the journal ISSN. WorldCat.org is a global library catalog project, and the extension links the catalog to the journal section by the journal ISSN. WorldCat.org can be used to find a library that has subscribed the journal. Furthermore the layout templates can be adapted to national services as EZB in Germany. The extension does only connect to the Pubmed service to retrieve data during the editing phase, and if the page is saved, the final created HTML page is stored within the database. Thus viewing the page is quite fast because there is no need to fetch any data from Pubmed in this moment. But one should be aware that in a case of complex search the final list is stored and frozen until the page is edited again.

Altogether the extension provides a pleasant way to automatically include literature data from Pubmed into MediaWikis.

References

1. WikiPedia [http://www.wikipedia.org/]
2. MediWiki [http://www.mediawiki.org/wiki/MediaWiki]
3. Stokes TH, Torrance JT, Li H, Wang MD: ArrayWiki: an enabling technology for sharing public microarray data repositories and meta-analyses. BMC Bioinformatics 2008, 9 Suppl 6:S18
4. Bio-Medical-Informatics Laboratory @ Georgia Tech [http://www.bio-miblab.org/arraywiki]
5. MediWiki Extensions [http://www.mediawiki.org/wiki/Category:Extensions]
6. Pubmed [http://www.Pubmed.org]
7. List of Journals Indexed for MEDLINE [http://www.nlm.nih.gov/tsd/serials/lji.html]
8. Manual Extensions [http://www.mediawiki.org/wiki/Manual:Extensions]
9. PHP [http://www.php.net]
10. Pubmed Extension [http://www.mediawiki.org/wiki/Extension:Pubmed]
11. NCBI Entrez Utilities Web Service of the NLM [http://www.ncbi.nlm.nih.gov/entrez/query/static/esoap_help.html]
12. Pubmed Search Field Descriptions and Tags [http://www.ncbi.nlm.nih.gov/books/bv.fcgi?rid=helpPubmed.section.Pubmedhelp.Search_Field_Descrip]
13. Boulos MN, Maramba I, Wheeler S: Wikis, blogs and podcasts: a new generation of Web-based tools for virtual collaborative clinical practice and education. BMC Med Educ 2006, 6:41

Glossary

API	Application Programming Interface
Ajax	Asynchronous JavaScript and XML
DOI	Digital Object Identifier
EZB	Elektronische Zeitschriftenbibliothek (Electronic Journal Library)
HTML	Hypertext Markup Language
HTTP	Hypertext Transfer Protocol

HTTPS	Secure Hypertext Transfer Protocol
ISSN	International Standard Serial Number
NCBI	National Center for Biotechnology Information
NLM	National Library of Medicine
PHP	Programming language
PMID, UID	Pubmed Identifier, Unified Identifer
SOAP	Simple Object Access Protocol
XML	Extensible Markup Language
XML-RPC	Extensible Markup Language Remote Procedure Call