A landmark in biomedical information: many ways are leading to PubMed

Introduction

Enter PubMed, provided by NCBI (National Center for Biotechnology Information), offers access to the most comprehensive literature database in the field (PubMed), and it also provides users with access to a broad range of databases, fact databases, ontologies and vocabularies, and educational material. It provides information suitable for consumers and professionals, as well as complex multimodal information for scientists. This portal stands for valuable open-access information, offering most of its information sources for free, including bibliographic information about articles and access to freely available full-texts. Due to its comprehensiveness, reliability, and availability, it is a primary information resource for many specialized portals including and distributing tailored, theme-focused information. NCBI offers several well-documented programming tools for inclusion of automated information access into external applications. We present an example showing how these automated ways of access can help to build customized special-purpose access portals that can be used by users who are not familiar with programming.

Material & Methods

Wiki programs provide fast and easy alteration of web (HTML) pages, aiming to ensure that the information on a page is always up to date due to continuous collaborative curation of contents. The most famous example of a wiki is the free encyclopedia Wikipedia [http://www.wikipedia.org] based on the MediaWiki software [http://www.mediawiki.org/wiki/MediaWiki]. 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. Wikis commonly are not estimated to be suitable information system platforms in the life science realm, but there are remarkable approaches to prove success. ArrayWiki [Stokes 2008, Bio-Medical-Informatics Laboratory of Georgia Tech, http://www.bibmi Sinai/arraywiki] is an example for a biomedical wiki that allows to share and publish public microarray data repositories and meta analyses. Wiki programs normally provide 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 extensibility [http://www. mediawiki.org/Category:Extensions], facilitating to add new features or enhance functionality of the main MediaWiki core.

PubMed Extension Implementation

Sometimes it is necessary to include literature data into a wiki page. Up to now there was no way to include PubMed data automatically into a MediaWiki page. The described PubMed extension [http://www.mediawiki.org/wiki/Extension:PubmedExt] uses hooks and tag-function associations in the wiki software part. It is based on a fgi/XML class in PHP and contains methods to connect to the PubMed API functions that are listed in [http://www.ncbi.nlm.nih.gov/sites/ESgeneral?help=html]. Programming details can be found in [Bohne 2008]. The extension is invoked by the "pubmed"-tag and transmits a query to PubMed using the described method. The result set consists of an PubMed article IDs related to the query. A PubMed request causes the server to send the answer as a large set of information in a XML record. In a subsequent step, the algorithm extracts the necessary parts by mapping the related fields. This step is necessary, as some attribute names of the XML result unfortunately are not unique, like the ‘Forename’ or the ‘Surname’ 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 publisher’s page, if DOI information is available.

Result

The MediaWiki provides a new tag within the markup language. The "pubmed"-tag establishes a powerful connection to PubMed by submitting the value between the opening and closing tag as a query string to PubMed. The "pubmed"-tag can possess three attributes that influence the layout of the result set. The attribute 'limit' controls the number of result records returned by the layout generator. The natural limit set by the NCBI utilities is 500 result records per query. The parameter should be used with complex queries – e.g. to present the first 20 hits. The two other attributes 'layoutfile' and 'layoutlink' can set the current layout to layout templates different from the default mode. Two other pieces of information about publication status are added in brackets - Information about the publication medium and status (P for published, I for Internet) and information about publishing institution (p for published, a for ahead of print and e for estimated to be published).

1. Medical Faculty Mannheim of the University Heidelberg, Lindolf-Krebl-Straße 12-17, D-68167 Mannheim, Germany, bohne-lang@medmum.uni-heidelberg.de
2. Media Faculty, University of Applied Sciences Darmstadt, Haarstr. 100, D-64285 Darmstadt - Campus Dieburg, elke-lang@fmd медиа-ха.de

Figure 1: Compositional scheme - data flow

Figure 2: Example: http://www.ncbi.nlm.nih.gov/geo/query/acc_index-php?GEOid_Page

Figure 3: Compositional scheme - data flow

Figure 4: Example layout file

Figure 5: Compositional scheme - data flow

Figure 6: Compositional scheme - data flow