

Forum Kultur & Informatik.

An Online Pathfinder to Digital Cultural Heritage

Leonhard HUBER; Johann STOCKINGER

Zentralanstalt für Meteorologie und Geodynamik, Wien; Oesterreichische Computer Gesellschaft, Wien

Online resources on computing technology and digital media for capturing, archiving and delivering of cultural heritage artifacts are countless. For an interested beginner, the field of humanities computing can hardly be overviewed. Information provision about cultural heritage informatics forms a multi-flavoured patchwork of online resources; most extensive contents are offered on portals of research institutes (DigiCULT), associations/expert networks (Virtual Library Museen; Archimuse), and university libraries (e.g. Cornell University Library Digital Initiatives).

Within its initiatives in the area society and culture, the Austrian Computer Society runs a forum entitled "Forum Kultur & Informatik". In this context, lectures are regularly given on topics such as digital preservation, collection management etc. Furthermore, with Cultural Heritage Digitisation License (CH@DL), a training-on-the-job course for librarians, archivists and museum curators is being developed. The forum is currently working on the conception of an extensive online pathfinder for the field of cultural heritage and computing – this activity is presented in the paper in detail.

Building on a student project, an online portal is evolving that tries fulfill a pathfinder function for beginners, intermediates, and professionals in the area of cultural heritage informatics. The project's effort is two-fold: to develop an ontology for the field of humanities computing. And, based thereon, to provide the interested audience on the WWW with qualitative content in its topical context.

Literature Review

The basis literature review was conducted in respect to the following subject areas:

humanities computing / digital library / museums informatics – technologies, methodologies, activities, and project management

discovery of best practice projects

from the following authors / institutions (exemplified):

Int'l Council of Museums, International Council of Archives, International Federation of Library Associations Arts and Humanities Data Service, Archives & Museums Informatics, Salzburg Research, EVA Conferences, VL Museums Pages, H-Soz-u-Kult

Jonathan P. Bowen, Lorna M. Hughes, Suzanne Keene, Elizabeth Orna & Charles Pettit; and others

Ontology development for humanities computing

The literature review resulted in the creation of an inclusive ontology. The main classes in the ontology and the property categories for the class "Projekte" (projects) are displayed in the following two tables (Tab. 1 and Tab. 2).

<i>class</i>	<i>coverage</i>
scientific disciplines, content	thematic area of the digitized/digital information resources
IM activities and processes	collection management, user research, project management
information management / information technology tools	information organization, input and output devices, data exchange formats and interfaces
institutions	education, science and research, business, public administration
continents, countries	continents, countries, towns
culture and society	culture, society
presentation and information provision	evaluation methods, encoding system, types of presentation, sensory perception, building blocks for information provision, information organization paradigms, accessibility
persons	experts in the area of digital cultural heritage
projects	best-practice examples (especially user-oriented and technologically innovative designed information systems, offering digital cultural content)
publications	books, diploma theses/dissertations, case study collections, conference proceedings, journals, electronic sources
collections	digital repositories
events	interviews, conferences, lectures, presentations, radio broadcasts, seminars, TV documentaries, workshops

Tab. 1 - Main classes in the ontology for "Forum Kultur & Informatik"

<i>property category</i>	<i>coverage</i>
basic information on the project	project title, target, start and end date, project partners, financing
requirements analysis	user groups, type of delivery interface, topic scope
system architecture	input devices, output devices, information architecture, storyboard, interface design
prototype development	prototype content, usability testing
content research & production	content research (existing contents, authoring of new content), multimedia production, copyright clearance; content / digital asset management life-cycle
live system	system maintenance, usage statistics, usability evaluation, user feedback (QA)

Tab. 2 – property sets of class "Projekte"

Content authoring

The knowledge acquisition process - ontology creation as well as content authoring – was performed with Protégé Ontology Editor¹. This toolkit enabled collaborative work. At first, the classes' and properties' structures were defined (Fig. 1); afterwards, instances were inserted (Fig. 2), and the semantic network started to evolve.

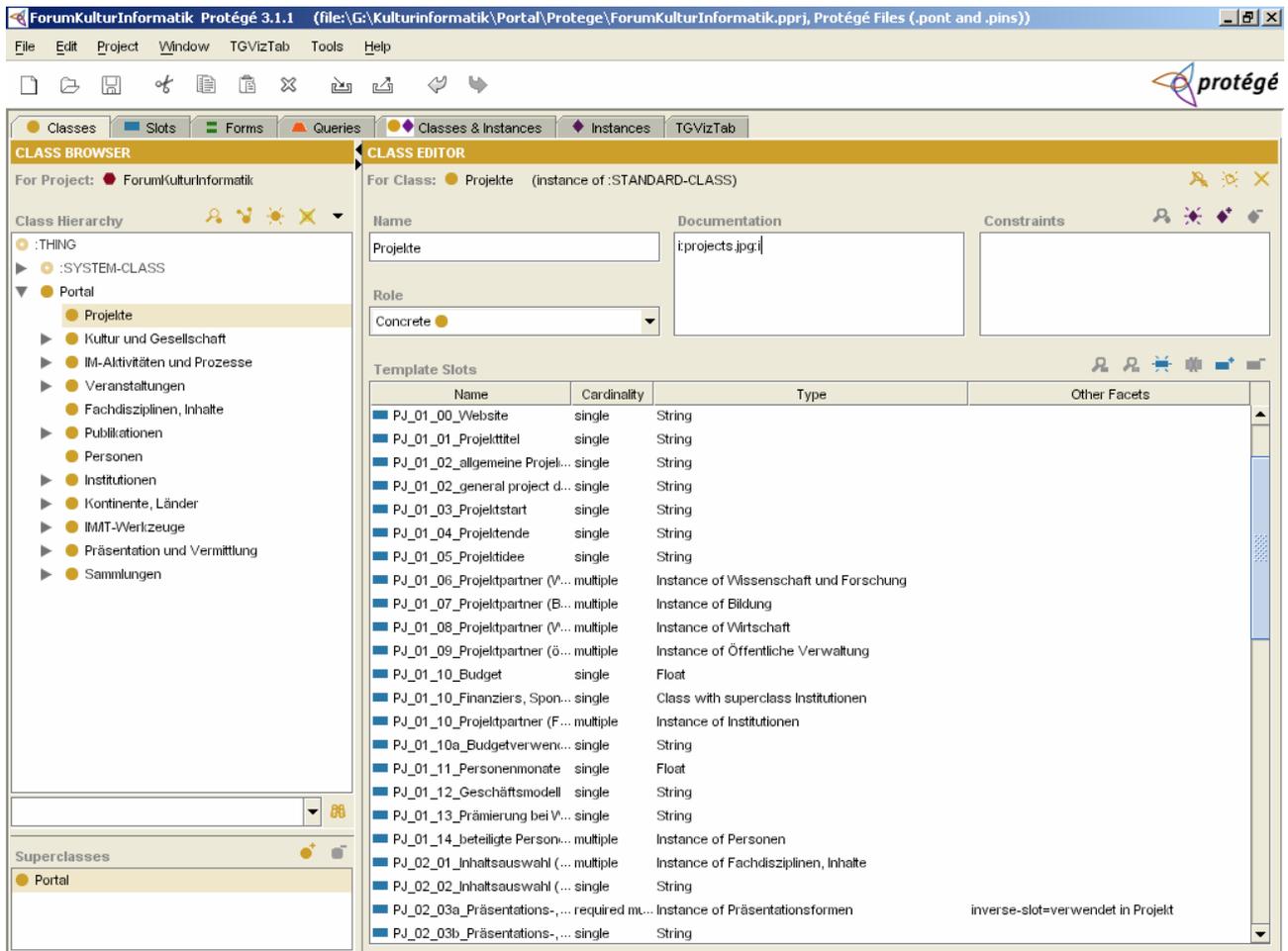


Fig. 1 - Class hierarchy and property structure in Protégé Ontology Editor

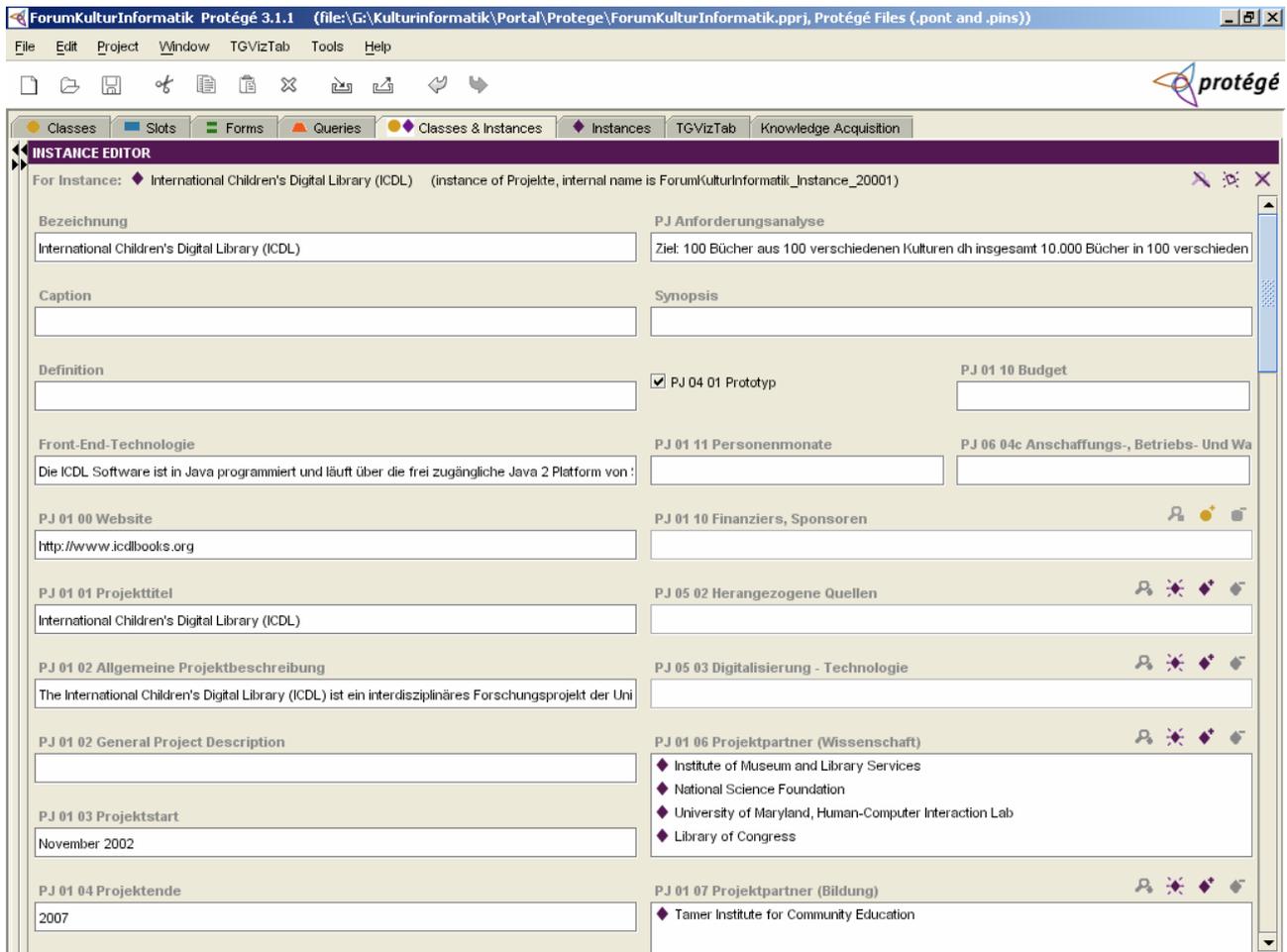


Fig. 2 - Creation of a project instance

Software development of a Web interface to display RDF data

The target of the software development was to create a Web interface for the presentation of Semantic Web content, as exported in the form of RDF(S) by Protégé. RDFS offers a basic Markup facility for the syntactic encoding of ontologies. RDF-tagged instance data then references RDFS ontologies for the expression of properties and relations. In sum, RDF(S) enable the codification of a semantic network.

Apache HTTP server ² ver. 2.x	meistverwendete Web-Server-Software (71% von 74 Millionen Websites weltweit ³)
MySQL-Datenbank ⁴ ver. 4.x	relational data base management system, exceeding 5 million set-ups world-wide
PHP ⁵ ver. 5.x	Hypertext Pre-Processor, programming language for the implementation of dynamic Web services
phpMyAdmin ⁶ ver. 2.x	administration interface for DB/MS
RDF API for PHP ⁷ ver. 0.9	software toolkit for parsing RDF data
Treeblic 1.8.0 ⁸	visualization software (Java-Plugin)

Tab. 3 - The basic software configuration

The following functionalities are offered by the software:

- systematic display of RDF contents
- query functionality
- visualization

To enable Information Retrieval based on the RDF data, SPARQL query language for RDF⁹ is used. It enables – similar to SQL with ordinary relational databases – efficient and precise query formulation.

The portal

The portal's layout and information architecture was developed with regard to accessible design. (Compare the book by Peter MORVILLE and Louis ROSENFELD (2002). The ISO (1996/1998) 9241 guidelines on "Dialogue principles and "Guidance on usability" were referred to, as well as the conformance criteria set by the Web Accessibility Initiative. (For cultural heritage online offerings for people with special needs compare Daniela MITTERHÖFER 2004.)

During the design process of the Web portal, special attention was given to the Visual Information-Seeking Mantra postulated by Ben SHNEIDERMAN (1996). The portal seeks to conform to the logic of "Overview, Zoom, Filter, then Details-on-Demand". The portal's entry page is shown in Fig. 3 below.

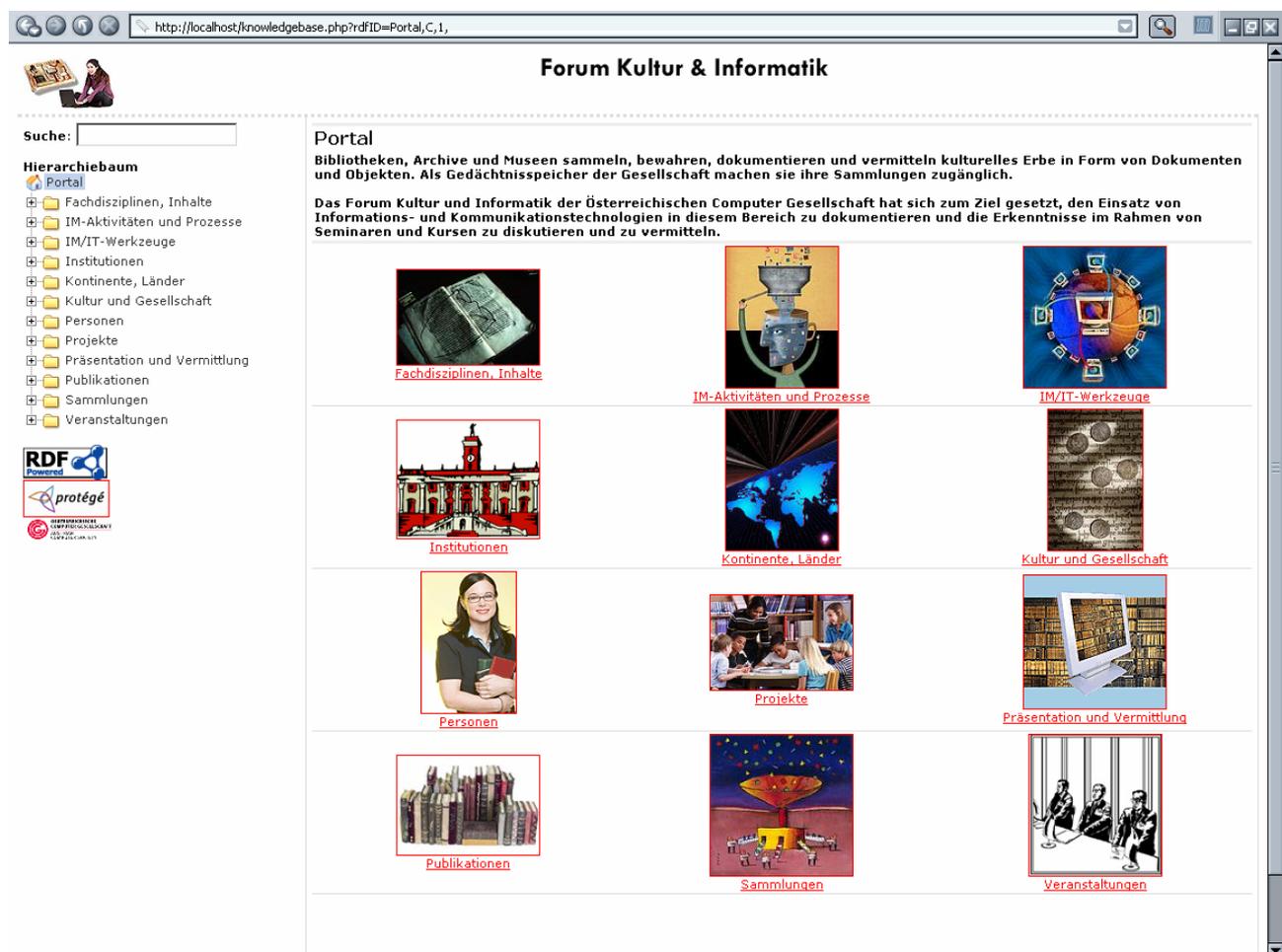


Fig. 3 - The home page of the portal showing the main classes as entry points.

Interface components

Search field

The search box enables information retrieval for the semantic network. The search can be based on a boolean combination of terms via "and", "or" and "not" operators. Furthermore, a phrase search is possible; search terms also can be truncated. The search results consist of the matching classes and instances.

Navigation tree

The hierarchic navigation tree displays the conceptual structure of the semantic network, containing both classes and instances, as shown in Fig. 4.

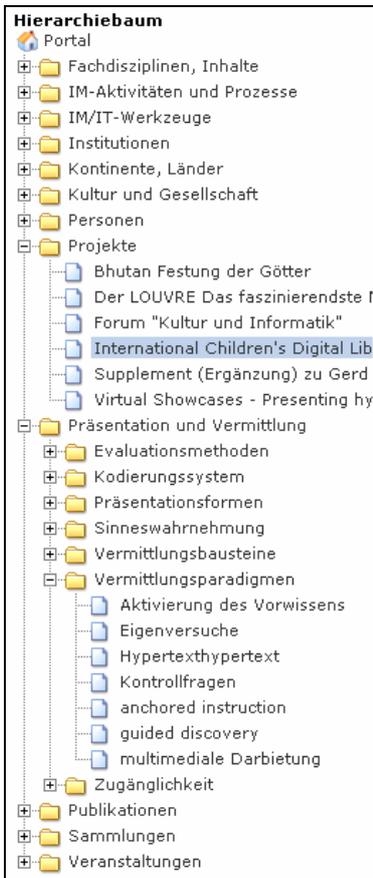


Fig. 4 - Navigation tree

Content window/main

The text displayed in the content window is dependent on the type of user interaction.

With a class being selected, its properties are displayed accompanied by a short textual description.

Alternatively, instance data is shown (the contents of the knowledge base) (Fig. 5); a search result listing is displayed respectively.

<p>Software-Schnittstellen Screen-Design / Layout (Graphik)</p>	
<p>Front-End-Technologie</p>	<p>Die ICDL-Software ist in Java programmiert und läuft über die frei zugängliche Java 2 Platform von Sun. Die Java Web Start-Technologie ermöglicht es dem User, mit einem einfachen Klick auf einen Web-page-link, die Software herunterzuladen, zu installieren und hochzustarten.</p> <p>Basic-Zugang: Text-Suche, visuelle Suche Enhanced-Zugang: Vergrößerung, visuelle Suche, erweiterte Textsuche</p> <p>So kann die Software entweder über die Web-Site oder direkt vom Desktop abgerufen werden.</p> <p>Die meisten Bücher liegen unverschlüsselt in jpeg-format auf einem Web-Server(Basis-Zugang) und sind auch direkt über die Java-Client-Software zugänglich (erweiterter Zugang).</p> <p>Verschlüsselte Bücher liegen auf einem Adobe-Content-Server, sind über das ICDL-Application's-Visual-Search-System abrufbar und werden über die frei zugängliche kommerzielle Adobe eBook Reader Applikation gelesen.</p>
<p>Prototyp</p>	<p>true</p>
<p>Prototyp-Inhalte</p>	<p>Bookreader-Prototypen: Die Inhalte werden für 3 unterschiedliche Lesertypen gestaltet.</p> <ul style="list-style-type: none"> Im "Comic Strip Reader" werden die Inhalte wie Comic-Streifen dargestellt. Dadurch kann der Leser schnell erkennen ob viele Bilder oder mehr Text den Inhalt bestimmt bzw schnell innerhalb der Seiten navigieren. Das digitalisierte Buch kann so entweder Seite für Seite gelesen oder innerhalb der Seiten hin und her gesprungen werden. Andererseits entspricht der Aufbau dem Leseverhalten einer "Lese-Spirale". Die aktuelle Seite wird zwischen 2 Spiralen präsentiert. Während die linke Spirale die vorangehenden Seiten zeigt sieht man auf der rechten Seite die folgenden Inhalte. Auch hier wird die schnelle Navigation zwischen den einzelnen Seiten ermöglicht. Der "Traditional-Reader" ähnelt den handelsüblichen Readern. Man sieht jeweils eine Seite und hat über Zurück- bzw. Weiterbuttons die Möglichkeit jeweils um eine Seite nach vorn oder zurück zu navigieren. Der User kann darüber hinaus auswählen, ob er zwei oder nur eine Buchseite gleichzeitig sehen möchte.
<p>Prototyp-Evaluationsmethoden</p>	<p>Vorab-Evaluation</p>
<p>Rechteklärung</p>	<p>http://www.icdlbooks.org/adults/copyright.html</p>
<p>Neukonzeption von Inhalten</p>	<ul style="list-style-type: none"> "Dynamic content generation" - ermöglicht in Zukunft einen Support bei der Internationalisierung der Inhalte "Book overviews" zeigen an wo auf welcher Seite man sich zum aktuellen Zeitpunkt befindet
<p>Zugangsbasis</p>	<p>Online-Zugriff freie Verfügbarkeit</p>

Fig. 5 - Web display of project instance shown in Fig 2.

Visualization

A visualization of the semantic network in the form of an interactive circular hierarchic graphic is offered. The semantic structure is depicted by a hyperbolic tree (see Fig. 6). The classes and instances are represented by nodes. Different color ranges enable a distinction between classes and instances. The instance relations are drawn as edges.

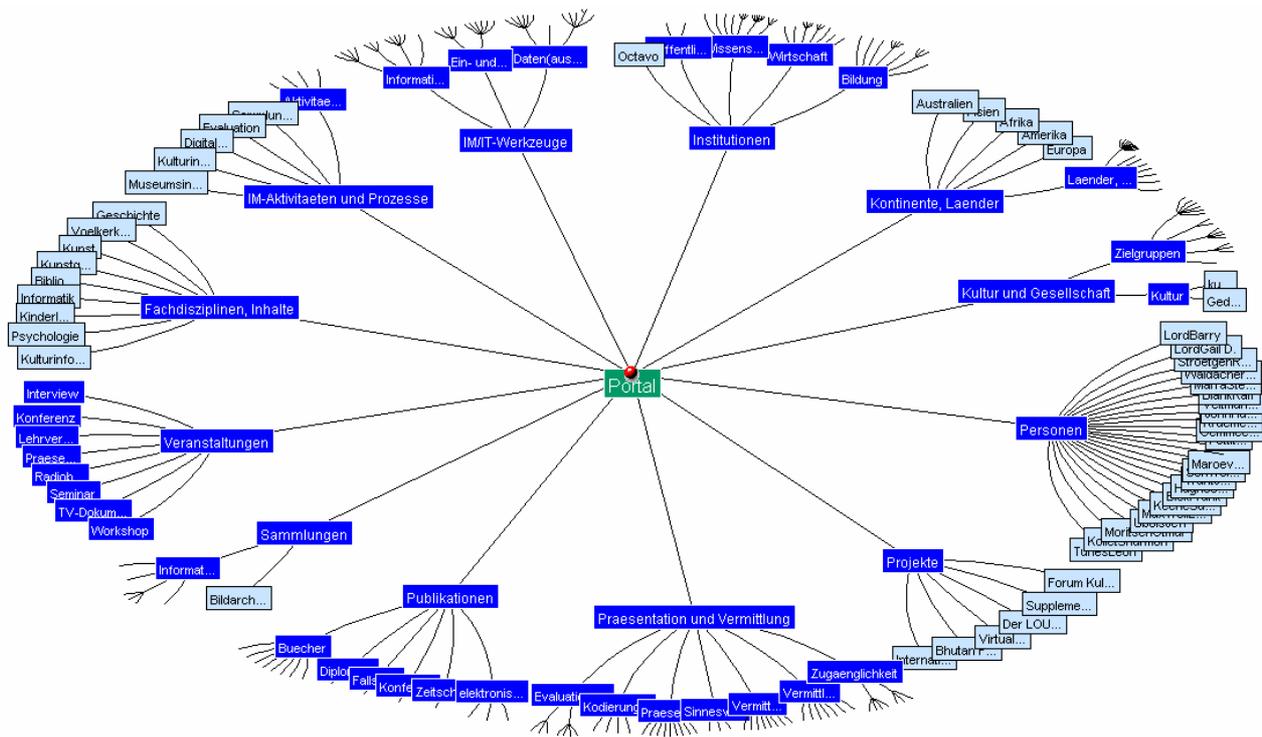


Fig. 6 - hyperbolic visualization of the ontology structure

Information behaviours supported by the Interface

Three generic types of information retrieval are supported: browsing, search, and visual browsing. These can be carried out separately as well as combined.

Browsing

The selection of an element from the hierarchical navigation tree initiates a browsing activity through the content repository. Afterwards, the user may follow the relations between single instances, so they are able to display every entries tied to a specific aspect.

Search

In contrast to an explorative entry into the knowledge base, a more or less specific search is also possible. Every search request is followed necessarily by a browsing activity through the result listing.

Visual browsing

The visualization of a semantic network helps the user to arrive at a "bird's eye view of the knowledge network. Parts can be magnified, selected, and displayed separately. The visualization offers a higher efficiency to the user than pure textual display, on account of the parallel – non-sequential – human perceptive faculty of information.

Further work

The project aims at perception larger than specialist scientific community. Therefore, the integration of discussion functions will be helpful, as well as promotion activities, both online and offline. The portal site will be available at <http://kultur.ocg.at>

References

- Mitterhöfer, D. 2004. Zugang zu digitalisiertem kulturellen Erbe. Angebote für spezielle Benutzergruppen. 54-68. (unpublished diploma thesis)
- Morville, P. & Rosenfeld, L. 2002. Information Architecture for the World Wide Web: Designing Large-Scale Web Sites. O'Reilly Media
- Shneiderman, B. The Eyes Have It: a task by data type taxonomy for information visualization. Proceedings of the IEEE Visual Languages 1996 (ed. IEEE Computer Society) 336-343.
- World Wide Web Consortium. 2005. Web Accessibility Initiative (WAI) - home page. <http://www.w3.org/WAI/>
- ISO (ed.) 1996/1998. Ergonomic requirements for office work with visual display terminals (VDTs), Part 10: Dialogue principles. Part 11: Usability principles in ISO 9241.

¹ <http://protege.stanford.edu>

² <http://httpd.apache.org>

³ http://news.netcraft.com/archives/web_server_survey.html

⁴ <http://www.mysql.com>

⁵ <http://www.php.net>

⁶ <http://phpmyadmin.sourceforge.net>

⁷ <http://www.wiwiss.fu-berlin.de/suhl/bizer/rdfapi/>

⁸ <http://treebolic.sourceforge.net/>

⁹ <http://www.w3.org/TR/rdf-sparql-query/>