

Access to Knowledge - Better Use of Internet

Volkmar H. Haase,
Technische Universität
Graz, Austria

vhaase@ist.tugraz.at

Christian Steinmann,
HM&S GmbH,
Graz, Austria

cs@hms.org

Stephan Vejda,
HM&S GmbH,
Graz, Austria

stv@hms.org

Abstract

WIP-Austria is an internet portal developed by Austrian Research Centers and by HM&S GmbH Graz. It gives access to databases, electronic and printed documents as well as to consultancy and services. The WIP database is based on "knowledge objects": the contents of a document or service is defined by a two dimensional matrix (knowledge map) based on a decimal classification scheme and a set of pragmatic attributes. WIP is accessed by posing questions; questions are interpreted as linguistic variables which can be used to build an "interrogation map". Pattern matching together with a Fuzzy Logic reasoning process leads to best fits.

The WIP prototype is in pilot use by the Austrian Research Centers.

Preamble

The schedule of the (WIP) project is as follows:

- 6/2000 - 9/2000 planning, specifications, development of data structures and algorithms (finished)
- 10/2000 - 6/2001 detailed design, implementation, test, contents provision (finished)
- 7/2001 - 6/2002 pilot operation within ARC research center
- 7/2002 (?) nationwide operation

Information in the Web

Information is no more a scarce resource but an abundant one. In the open market of the www "every/wo/man" has the problem to reach the information she/he really needs, as well as to judge whether it is correct or not. Studies say that half of the www-pages are either outdated, biased or just wrong. (Sorka, 2001; Wittmann, 2001)

e-commerce seems to boom. National and supranational bodies (e.g. the European Union) are heading for to reach a stronghold in this market. An Austrian government initiative (eAustria) in connection with the EU's eEurope proposes a "virtual Marketplace" combining e-commerce, e-government-services, as well as e-based knowledge dissemination. Initiatives creating prototypes are encouraged. (EU 2002)

Challenges in these initiatives as well as in the www-world as a whole are threefold: Customers (information seekers) expect that they will be one answer per question (not many); information providers expect that their problem solution will be found by the persons who have the need (needle in a haystack); information brokers want to use a tool

Material published as part of these proceedings, either on-line or in print, is copyrighted by Informing Science. Permission to make digital or paper copy of part or all of these works for personal or classroom use is granted without fee provided that the copies are not made or distributed for profit or commercial advantage AND that copies 1) bear this notice in full and 2) give the full citation on the first page. It is permissible to abstract these works so long as credit is given. To copy in all other cases or to republish or to post on a server or to redistribute to lists requires specific permission from the publisher at Publisher@InformingScience.org

which delivers best matches of questions and answers, and which is easy to handle.

Virtual Marketplace: The Prototype WIP Austria

In the Internet many products called “portals” or “knowledge portals” are available (Heindl, 2001). We developed our own solution in order to investigate strengths and weaknesses of our concept of a universal description of a “knowledge object”. WIP Austria (WissensPortal = knowledge portal) is the prototype of a knowledge marketplace in the Internet (Austrian Research Centers, 2002). It is an access point for persons or companies who want to access specific knowledge (both scientific and commercial: from the theory of relativity to Yellow Pages). The system is able to identify appropriate knowledge resources from a pool of selected knowledge providers (today these are the research facilities of the Austrian research centers), and from a nationwide, a Europe wide stock (tomorrow).

WIP finds answers in a two step process: a) matching of questions and knowledge resources in order to identify the most promising candidates; b) interrogation of a selected resource.

In our approach resources can be both documents (in paper or electronic form, including www-pages) and people (experts), as well as services implemented through persons and/or software and/or document-provision. WIP offers for each question the options of “reading” and “asking”. In the latter case the customer is directed to a person (via e-mail or telephone).

The Problem of Finding a Knowledge Resource

Search engines, portals and “classical” information retrieval techniques offer some kind of keyword search in the web. The results may be thousands of hits many of them worthless or – the major problem - not decidable whether correct/applicable or not. There are several reasons for this situation:

- a) the interrogator provides some syntax of a question, not a meaning (which the system has to guess);
- b) she/he stays anonymous (the system does not know her/his mental structure, it is unable to infer unsaid prerequisites);
- c) homepages frequently are more like “advertisements” (i.e. biased) than “abstracts”;
- d) secondary information (A quotes B quotes C quotes ...) is abundant.

The major goal of WIP is the matching of “knowledge-sought” and “knowledge-provided”. The objectives are:

- a) creation of a user profile so that the system knows her/his mental structure;
- b) to provide a framework to catch the meaning of questions and problems via (free) text analysis or via a “wizard”-interface that interrogates the user;
- c) to describe knowledge resources in an objective and standardized manner (using identical structures as in the description of questions);
- d) handling of documents and experts in the same way, namely as knowledge resources.

A Knowledge Pattern Matching Algorithm

A remark at the beginning: we looked into many classification schemes. In our mind was the scope of possible contents accessible via WIP (i.e. Austrian and European research results and services): the classification should provide a “flat” distribution. As we could not find the ideal one, we constructed our own scheme, a modification of Dewey’s (OCLC 2000).

The kernel of the matching process is a knowledge map, one for each question, one for each knowledge resource. The map is a matrix: There are 100 lines which represent topics from a specifically developed classification scheme (e.g. 40: Information Systems and Technologies, 87: Media and Communication Studies), and there are 10 columns representing “attributes”, i.e. scope (basics, surveys, details, applications) and cross relevance (environment, health, automation, society, safety, mobility). Each line, column, and/or matrix element can be accompanied by an open list of terms and keywords:

e.g.: 40 Information Systems and Technologies

Computer Systems Organization
 Information Systems Management
 Artificial Intelligence and Expert Systems
 Parallel Processing and Architecture
 Computer graphics
 Image Processing
 Speech Recognition
 Cognitive Science
 Information Retrieval

 Retrieval Languages
 GIS, Geographic Information System
 MIS, Management Information System

 Database Design

Any specific resource or question is represented by this matrix: each element carries a weight: 0% to 100% defining how relevant this item is in the description of knowledge provided or sought.

The algorithm matches the question with all available resource profiles (scalar matrix product), and the customer gets a short list of best fits i.e. potential resources he may contact. (for mathematicians: This mimics a Fuzzy Logic reasoning process where each knowledge item is represented by a set of linguistic variables.)

Resource profiles are provided by knowledge brokers of the WIP team in cooperation with the “human resources” (experts). Question profiles may be entered explicitly by the customer (a wizard type program helps), or are extracted automatically from the free text the customer sends as a question. In the case of non-anonymous customers (the normal case apart from the very first contact) a user profile is created, updated and used to clarify questions (it includes favorites, a trace of previous dialogues, information on profession and interests of the customer etc.).

The algorithm can accomplish what a competent knowledge broker is able to do. Certainly it will give up in complex cases – here a person, a “real” knowledge broker will be contacted automatically.

Implementation Technology

WIP is implemented using a Windows-NT-Server together with IIS (Internet Information Server) and ASP (Active Server Pages: Scripting with Visual Basic). A Microsoft SQL-Server acts as the database. Clients receive HTML-text (Explorer or Netscape browsers can be used).

WIP’s 3-tier architecture consists of the user-layer (ASP), middle-layer (ActiveX objects)

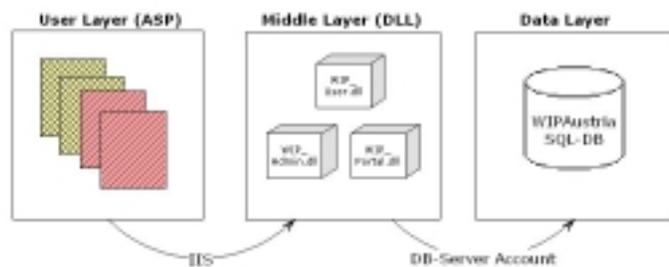


Fig. 1: WIP architecture



Fig.3: WIP homepage (including favourites and general news)

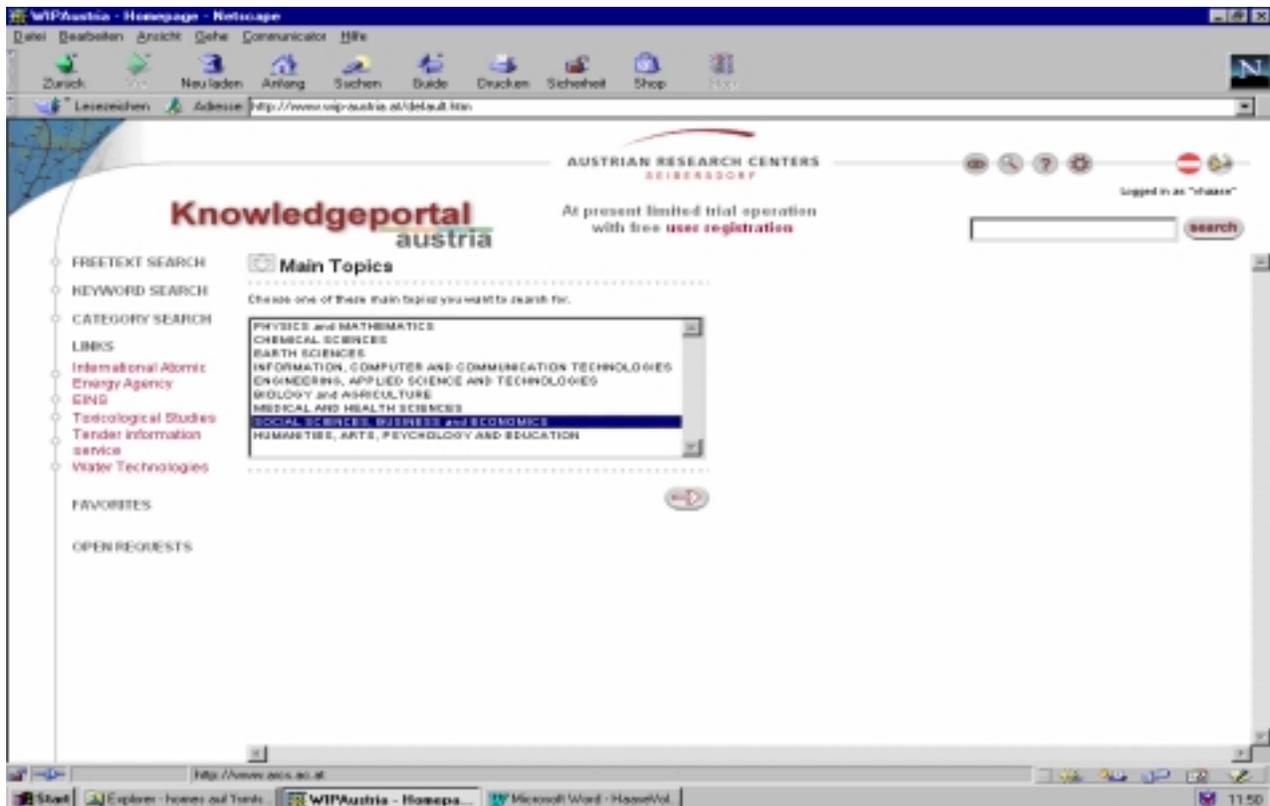


Fig.4: Category search (1st layer)

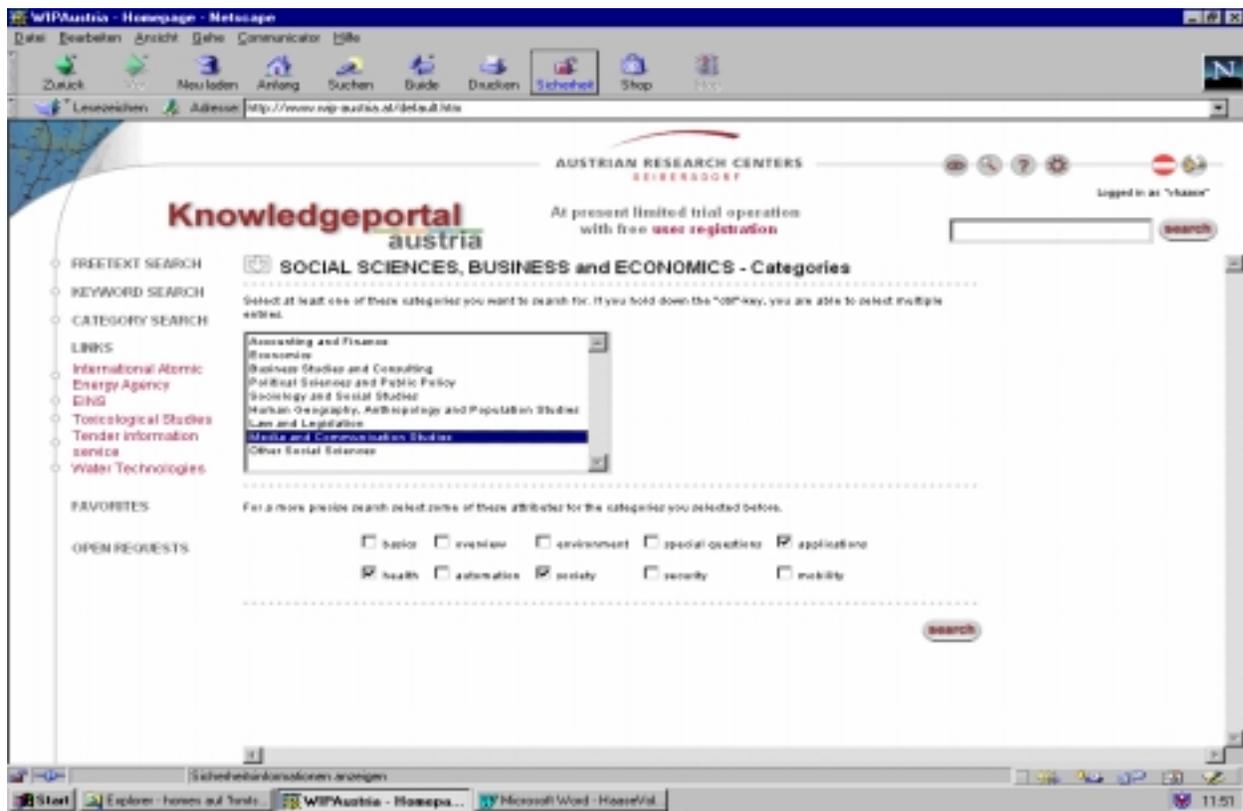


Fig.5: Category search (2nd layer) – user defines matrix of categories and

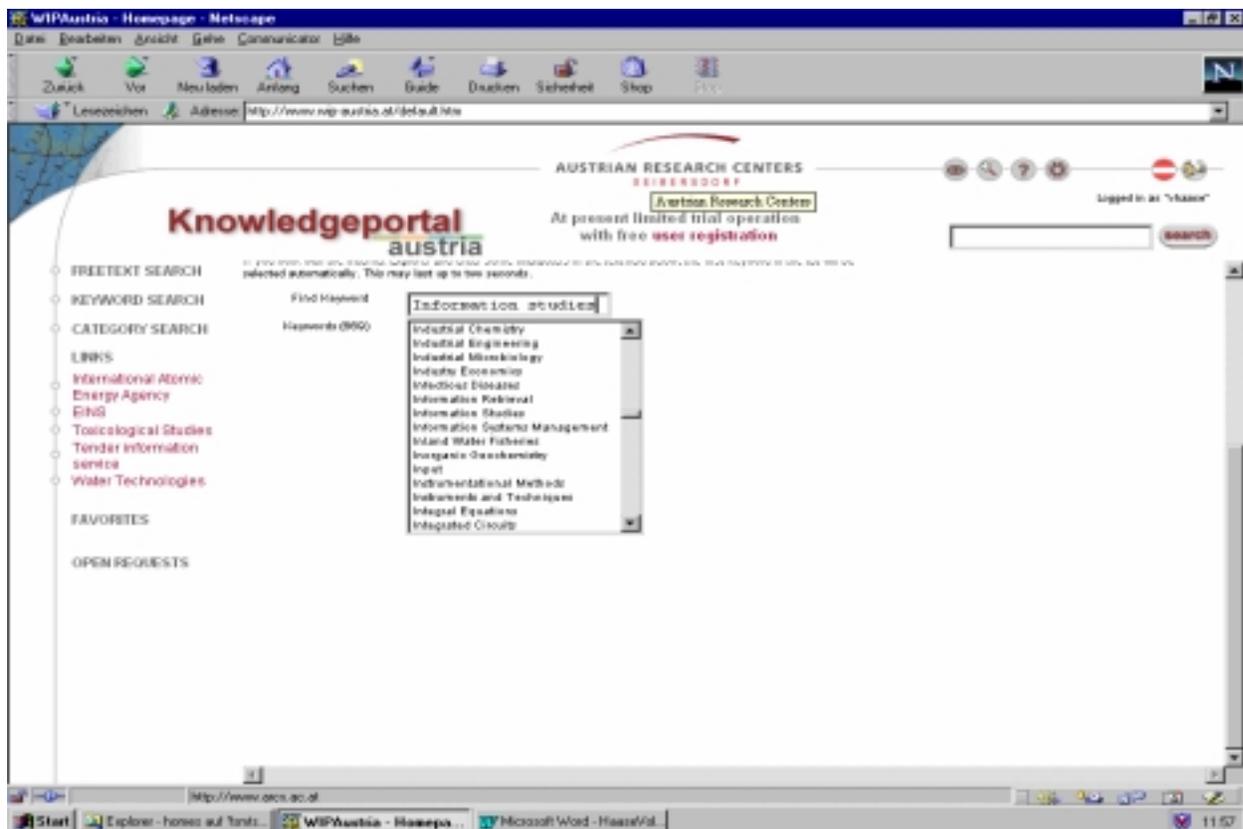


Fig.6: Keyword search: WIP offers a dictionary of keywords in its data-

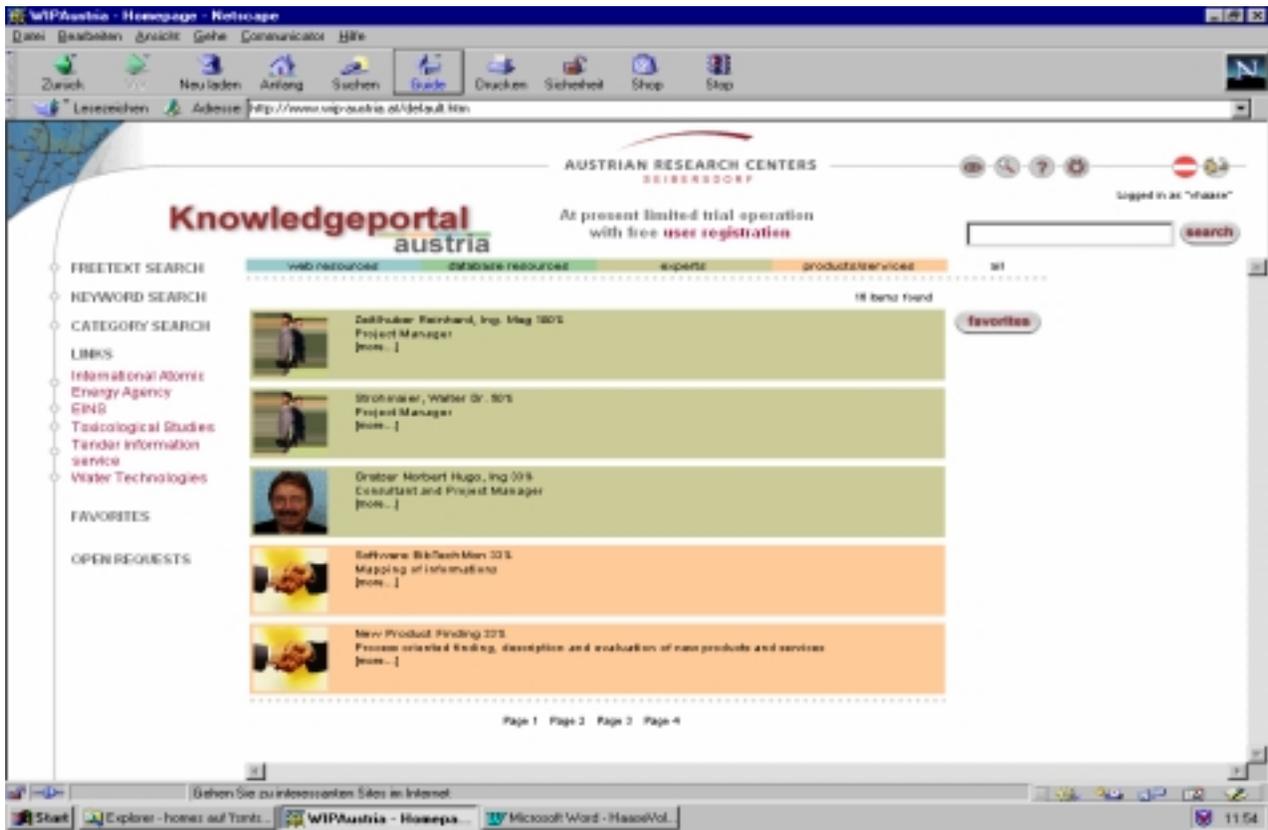


Fig.7: Results: Web-links, databases, experts, services



Fig.8: A typical expert profile



Fig.9: A typical service description

Conclusions and Future Prospects

In summer 2002 WIP should be operational as a portal that is open nationwide aiming at to provide (without charge) or to sell knowledge by a group of Austrian research institutions. Both German and English language interfaces will attract customers from abroad.

Lessons learned:

- a) it was easier (and cheaper) to build a proprietary system than to use or modify an existing one due to a clear concept and a competent programming team.
- b) it was difficult to provide the contents as this is “extra” work for scientists who are busy in their projects and not used to advertise and sell their knowledge.
- c) it is still most difficult to motivate Austrian research managers to promote the system. They expect immediate returns of investment which cannot be guaranteed.

Further steps are in two directions:

- a) technology: the matching algorithm will be refined into an automatic learning system improving itself at every transaction taking into account successes and failures.
- b) applications: extensions to allow to cash money will be provided; they include strategic classifications when what will be charged to whom, and some e-business module (from a third source)
- c) management: many other resources (documents and persons) will be represented by WIP. It will cover a high percentage of e-commerce of intangibles in Austria, and will serve educational purposes as well. It should also be licensed to other bodies who want to use it independently.

References

- Austrian Research Centers (2002). Knowledgeportal Austria. www.wip-austria.at
- EU (2002). e-Europe Benchmarking Report. http://europa.eu.int/comm/information_society/eeurope
- Heindl, E. (2001). Was ist ein Portal? <http://heindl.de/portale.htm>
- OCLC (2000). Dewey Decimal Classification. www.oclc.org/dewey
- Soroka, M. (2001). Assessing Consumer Health Web Sites. www.nyu.edu/wagner/Soroka.pdf
- Wittmann, S. (2001). Evaluating Websites. www.oakton.edu/~wittmann/find/eval.htm