SCIENTIFIC EXCHANGE INITIATIVE
HUNGARY – BADEN-WÜRTTEMBERG

FACHHOCHSCHULE STUTTGART
HOCHSCHULE DER MEDIEN

Budapest, March 29-30, 2004

Web Databases and Open-Source Technologies

Prof. Dr. Wolf-Fritz Riekert
Fachhochschule Stuttgart – Hochschule der Medien (HdM)
University of Applied Sciences Stuttgart – School of Media

mailto:riekert@hdm-stuttgart.de
http://v.hdm-stuttgart.de/~riekert

PROF. DR. WOLF-FRITZ RIEKERT CURRICULUM VITAE

FACHHOCHSCHULE STUTTGART
HOCHSCHULE DER MEDIEN

1977	**University of Stuttgart**: **Diploma** in Mathematics
1977 –	**Informatik GmbH, Stuttgart**: Software Developer and Team Leader
1984 –	**University of Stuttgart**, Institute for Informatics, Research Scientist (Knowledge-Based Man-Machine Communication) **Doctorate** in Computer Science (1986)
1987 –	**Siemens AG Munich:** Software Developer and Leader of the AI Programming Environment project (Siemens Common Lisp, Prolog)
1988 –	**Siemens AG Munich:** Assigned to **Research Institute for Applied Knowledge Processing (FAW) Ulm**, Project Leader (Geographic Information Systems, Remote Sensing, Object-Oriented Databases)
1993 –	**FAW Ulm**: Head of Environmental Information Systems Unit
1998 – today	**University of Applied Sciences Stuttgart - School of Media** (Hochschule der Medien Stuttgart): Professor in Information Technology (Computer Networks, Databases, Web Applications)
Offices	**German Informatics Society GI**: Vice Chair of the Special Interest Group Computer Science in Environment Protection **European Commission:** Expert for the Information Society Technologies Programme (project reviews, proposal evaluations)

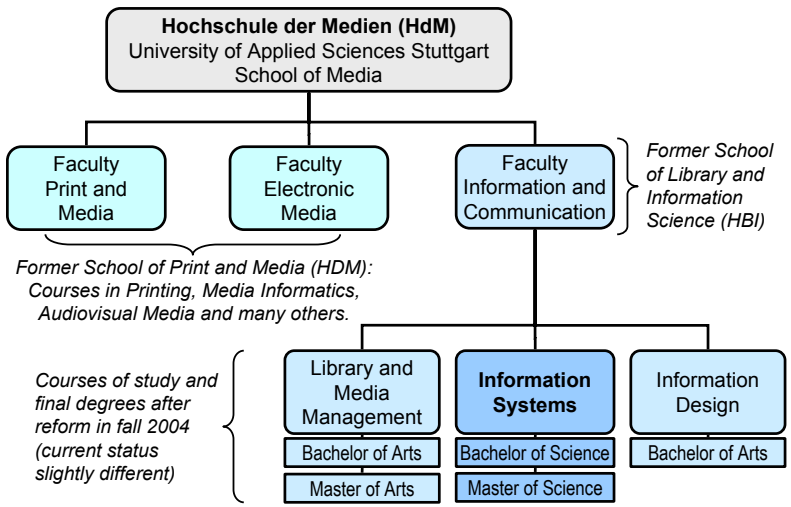
OVERVIEW

FACHHOCHSCHULE STUTTGART
HOCHSCHULE DER MEDIEN

- Information Systems at HdM Stuttgart
- Open Source in Education and Practice
- Service-oriented Software Architecture
- Web Database Applications (ISIQUA, IFAK)
- Peer-to-Peer Applications (PEERLINK)
- Catalog and Metainformation Systems
- Thesauri
- A Thesaurus Web Service (SWD Web Service)
- Outlook

INFORMATION SYSTEMS AT THE HOCHSCHULE DER MEDIEN (HdM)

FACHHOCHSCHULE STUTTGART
HOCHSCHULE DER MEDIEN



COURSE OF STUDY INFORMATION SYSTEMS

FACHHOCHSCHULE STUTTGART
HOCHSCHULE DER MEDIEN

Name:	Information Systems (IS) / Wirtschaftsinformatik
Degrees:	Bachelor of Science (BSc, after 3 years) Master of Science (MSc, additional 2 years)
Admissions/year:	~80 students (BSc), ~20 students (MSc)
Professorships:	12
Subjects:	Business Administration Corporate Application Systems Information Technology Information and Knowledge Management Communication and Media Electives Internship (5 months, part of BSc curriculum) Bachelor Thesis / Master Thesis

OPEN AND FREE SOFTWARE IN EDUCATION AND PRACTICE

FACHHOCHSCHULE STUTTGART
HOCHSCHULE DER MEDIEN

- **Open Source and Free Software**: an inexpensive option
  - ⇨ Low initial investment
  - ⇨ Joint software development in open source communities
- Especially suited for **education purposes**
  - ⇨ Free of charge for students and academic institutions
  - ⇨ Absence of sophisticated development environments as an advantage: Basic principles become more evident
- Increasing importance in **professional environments**:
  - ⇨ Attractive solutions for companies, especially SMEs
  - ⇨ Increasing Linux Server Market:
    2003: $ 1 billion = + 63% (Windows: $ 4 billion = + 16%)
    IBM, Novell expect 50% Linux share for 2006/2007
  - ⇨ Linux Client systems: Administrations (City of Munich), Banking and Insurances (3270 terminal emulations)

OPEN SOURCE AT HdM: LAMP

FACHHOCHSCHULE STUTTGART
HOCHSCHULE DER MEDIEN

LAMP (= **Linux** + Apache + MySQL + PHP/Perl/Python)

- **Apache**: open source web server
- **MySQL**: open source relational database system with increasing functionality
- PHP, Perl, Python: powerful scripting languages with large software libraries (e.g., PEAR, CPAN,...)
  - ⇨ Here **PHP** is used in most cases
- Platform for database-driven web applications
- Powerful applications possible
- Also installable on windows systems („**WAMP**“)
- Easy to learn, install, and handle
  - ⇨ High acceptance by students

OPEN SOURCE AT HdM JAVA-BASED DEVELOPMENT

FACHHOCHSCHULE STUTTGART
HOCHSCHULE DER MEDIEN

Sun‘s Java programming language is not open source, but open source development is possible with Java:

- Free download (http://java.sun.com)
- TOMCAT: open source Java application server (part of the APACHE project)
- Open source Java software development environments
  - ⇨ ECLIPSE (IBM)
  - ⇨ NETBEANS (Sun)
- Stable, secure, and professional software development possible
- Java system development more complex than LAMP development, requires more training

OPEN SOURCE AT HdM: OTHER POSSIBLE COMPONENTS

FACHHOCHSCHULE STUTTGART
HOCHSCHULE DER MEDIEN

Extensible Markup Language (XML)

- developed by the World Wide Web Consortium (W3C), all specifications are disclosed to the public
- A „metalanguage“ to create specific document types
- Most XML tools available as open source, e.g. as part of the Apache project

Web Services

- Applications may use remote applications as network services via the Internet
- Web Services support available for Java and LAMP environments as open source software

SERVICE-ORIENTED PARADIGM

FACHHOCHSCHULE STUTTGART
HOCHSCHULE DER MEDIEN

Most of the applications presented here follow a **Service-Oriented Paradigm**:

- Data
- Documents
- Functionality

are made available as **network services**.

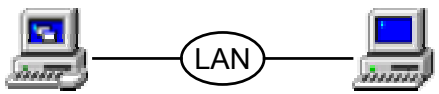
These services can be used

- directly by the users through a web browser in the form of a **web application**.
- by another service in the form of a **web service**.

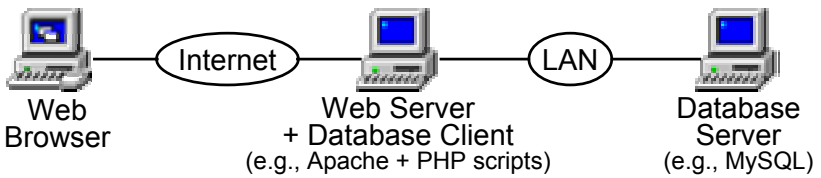
TWO TIERS VERSUS THREE TIERS

FACHHOCHSCHULE STUTTGART
HOCHSCHULE DER MEDIEN

Classical Client/Server model: 2-Tier-Architecture



Typical for Internet applications: 3 or more tiers

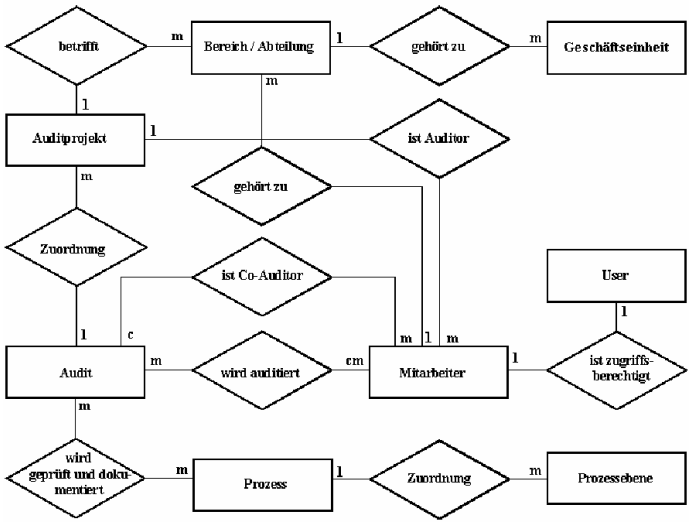


EXAMPLE: ISIQUA ISO 9001 QUALITY AUDITS

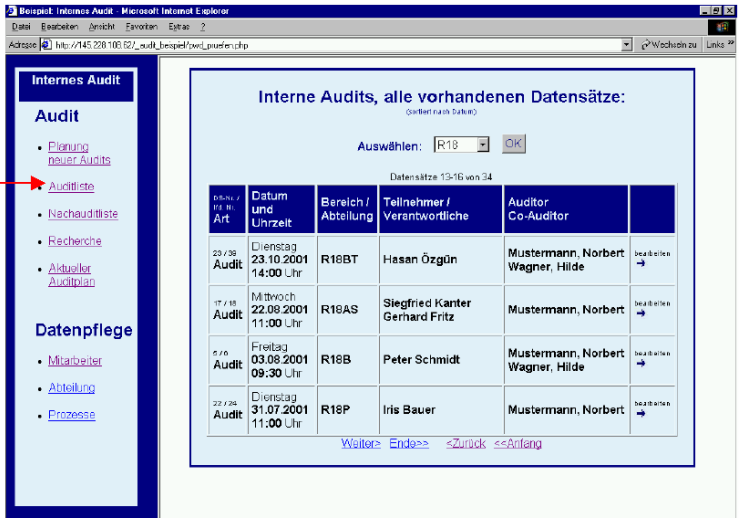
FACHHOCHSCHULE STUTTGART
HOCHSCHULE DER MEDIEN

- Purpose: **Management of internal ISO 9001 quality audits**
  - ⇨ Planning and scheduling of audit sessions
  - ⇨ Information platform about on-going audits
  - ⇨ Documentation, archival of reports
- User: **Marketing Service Süd-West**, a Bertelsmann company
- Developer: **Gina Frank**, M.Sc.
  Master Thesis in Information Systems
  at HdM Stuttgart, 2002, supervisor: W.-F. Riekert
  (http://v.hdm-stuttgart.de/~riekert/theses/master-frankg.pdf)
- Approach: Development as **LAMP system**

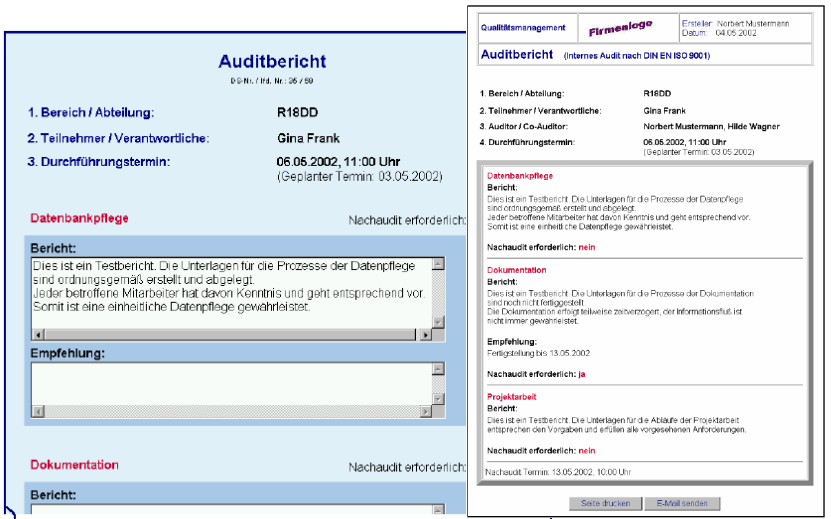
ISIQUA: ENTITY-RELATIONSHIP MODEL



ISIQUA: QUALITY AUDIT MANAGEMENT



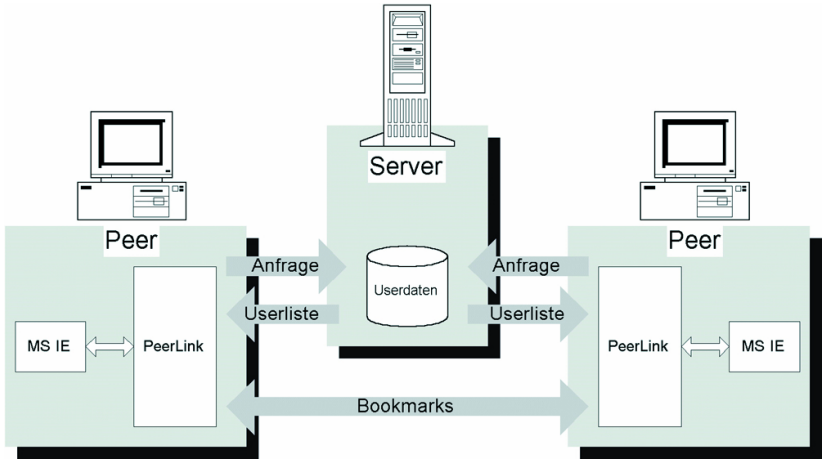
ISIQUA: DATABASE DRIVEN QUALITY REPORT MANAGEMENT



EXAMPLE PEERLINK: P2P BOOKMARK SHARING

- Purpose: Useful demonstration of a **Kazaa-like peer-to-peer application**
  - ⇨ **Bookmarks** (favorite URLs) can be shared directly between **Peers**
  - ⇨ Central user registry on a central **Server**, only used to get information about online users
- Developer: **Stefan Weisenbacher**, Diplom-Informationswirt
  Diploma Thesis in Information Systems
  at HdM Stuttgart, 2003, supervisor: W.-F. Riekert
  (v.hdm-stuttgart.de/~riekert/theses/dipl-weisenbacher-s.pdf)
- Approach: Development as **Java application**

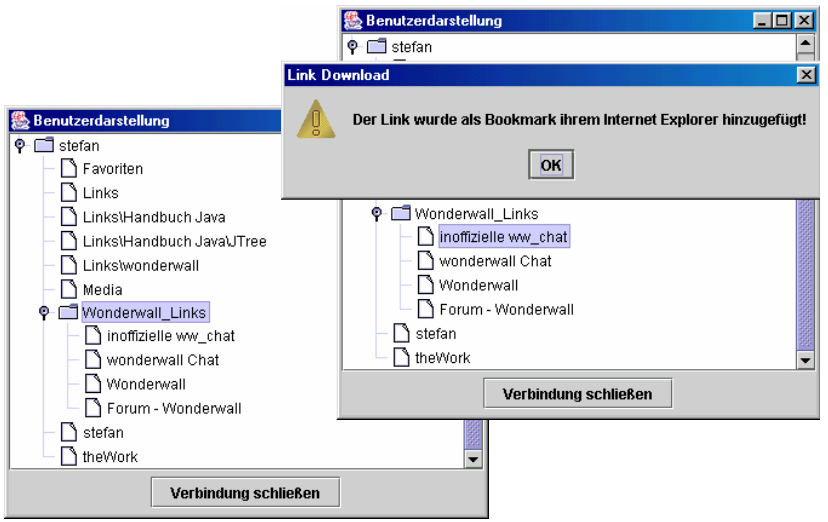
PEERLINK ARCHITECTURE



PEERLINK IMPLEMENTATION

- Peerlink: **Java** application installed at each **Peer**
- **MySQL** database on a central **Server** contains user registry
  - ⇨ Connection between Peer (Peerlink application) and Server (MySQL database) via **JDBC** (Java Database Connectivity, allows remote execution of SQL queries)
  - ⇨ Peerlink functions for **registration, logon, logoff**
- Peer-to-peer communication:
  - ⇨ TCP Socket communication between Peers using an **HTTP-like protocol** (predefined Java classes for HTTP communication can be used)
  - ⇨ Allows for **browsing** in foreign bookmark folders and **downloading bookmarks**
- Interface to **Internet Explorer** bookmark files for reading and creating bookmarks

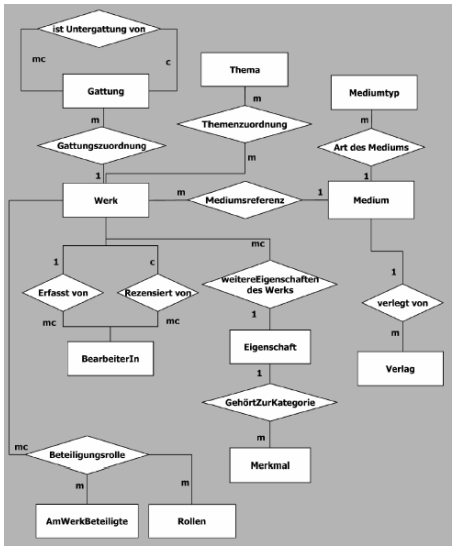
PEERLINK: SHARING BOOKMARKS



EXAMPLE: IFAK MEDIA RECOMMENDATIONS

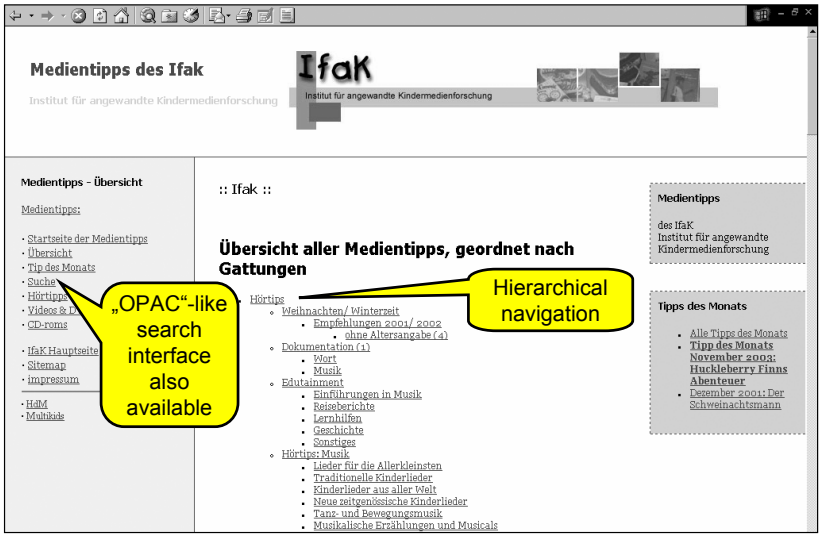
- Purpose: **Provide Recommendations/Reviews for Media products** (audiobooks, movies, computer games for kids)
  - ⇨ Web portal for children and parents
  - ⇨ Authoring system for reviewers
- User: **Institute for Applied Children Media Research** (IFAK – Institut für angewandte Kindermedienforschung)
- Developer: **Stephan Kimmerle**, Diplom-Informationswirt
  Diploma Thesis in Information Systems
  at HdM Stuttgart, 2004, supervisor: W.-F. Riekert
  (http://v.hdm-stuttgart.de/~riekert/theses/dipl-kimmerle.pdf)
- Approach: Development as **LAMP system**

IFAK: ENTITY RELATIONSHIP MODEL

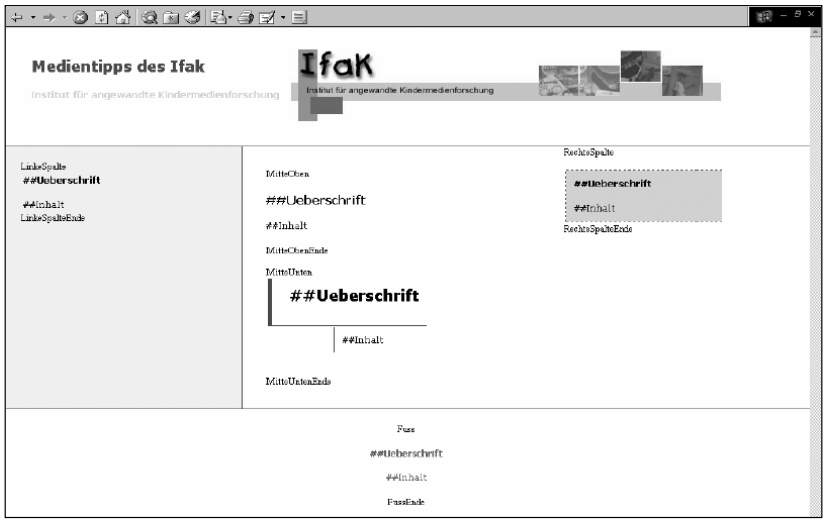


- Content is represented in a relational database
- Consistent presentation style (against predecessor system based on raw HTML pages)
- Various kinds of presentation possible:
  - ⇨ Hierarchy
  - ⇨ News
  - ⇨ Search results

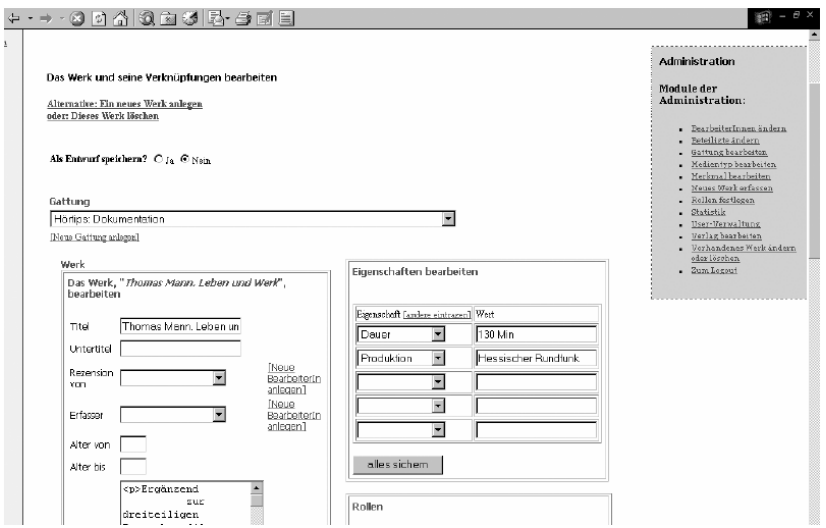
IFAK USER INTERFACE: HIERARCHICAL PRESENTATION



IFAK IMPLEMENTATION: TEMPLATE-BASED PAGE DESIGN



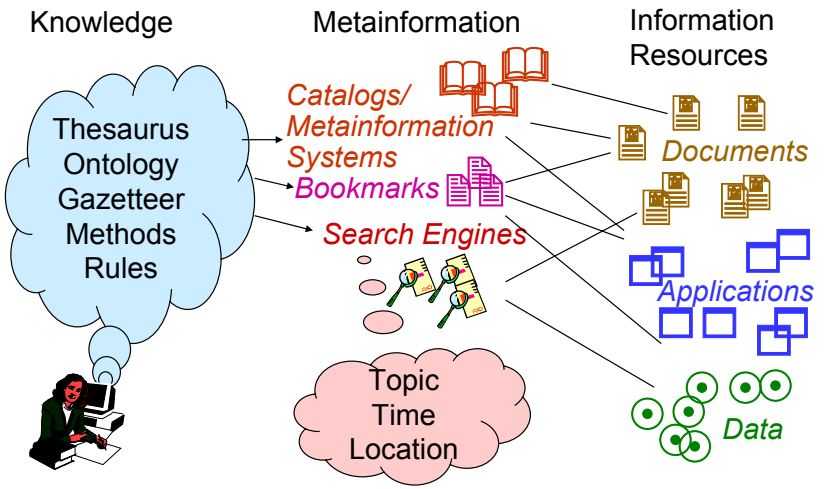
IFAK: INTERFACE FOR REVIEWERS



EXCURSUS ON CATALOG & METAINFORMATION SYSTEMS

- IFAK and PEERLINK are examples for catalog systems
- IFAK contains information about information and media products
- PEERLINK contains bookmarks, i.e. information about Internet resources
- Both contain information about information, i.e., metainformation
- Metainformation is of crucial importance for the retrieval of information in the internet:
  - ⇨ Information Catalogs / Metainformation Systems
  - ⇨ Bookmark lists
  - ⇨ Search Engines

INFORMATION RETRIEVAL



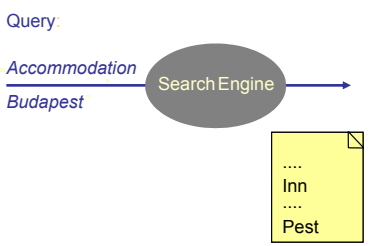
SEARCH ENGINES

Search engines are based on a full text index which intentionally covers the whole Web

- Retrieval via Web browser (string search)
- Index maintained by "robots" "crawling" along hyperlinks
- No additional efforts required from information suppliers

But:

- Search terms are interpreted only textually
- No semantic interpretation
- Full text index can only be used for textual resources



METAINFORMATION SYSTEMS

Metainformation systems support semantic criteria for indexing and retrieval:

- Thematic references (e.g., "Accommodation")
- Spatial references (e.g., "Budapest")
- Temporal references (e.g., "March 29, 2004")

Indexing (i.e., entering the metainformation) is done manually by the system administrator or information suppliers:

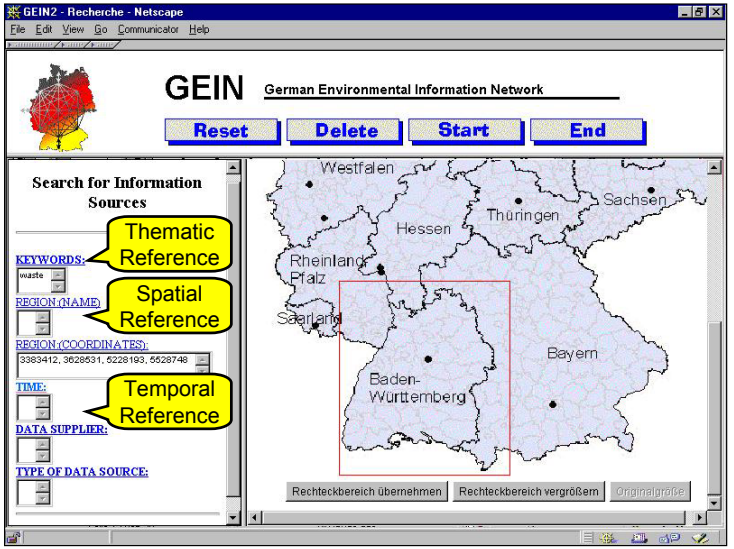
- Higher information quality (compared to search engines)
- Higher workload imposed on system administrator or information suppliers

Example: German Environmental Information Network (GEIN), the author participated in the prototype development

EXAMPLE: GEIN PROTOTYPE A METAINFORMATION SYSTEM

- Purpose: **Metainformation System for Environmental Information Resources**
- User: **German Federal Environment Agency** (UBA – Umweltbundesamt), **Ministry of Environment and Traffic Baden-Württemberg**
- Developer: **Research Institute for Applied Knbowledge Processing FAW Ulm**, (Forschungsinstitut für anwendungsorientierte Wissensverarbeitung), W.-F. Riekert, Ch. Fuchs, G. Klingler, 1998 (http://v.hdm-stuttgart.de/~riekert/papers/99nuernb.pdf)
- Approach: Partially proprietary, by using PERL, Java, C++, NCSA Web Server, ORACLE database

GEIN PROTOTYPE: A METAINFORMATION SYSTEM



SPECIFICATION AND PROCESSING OF SEMANTIC CRITERIA

Requirements

- Vocabulary for the specification of thematic, spatial and temporal references of information resources
- Techniques for the automated processing of thematic, spatial and temporal references

Approach

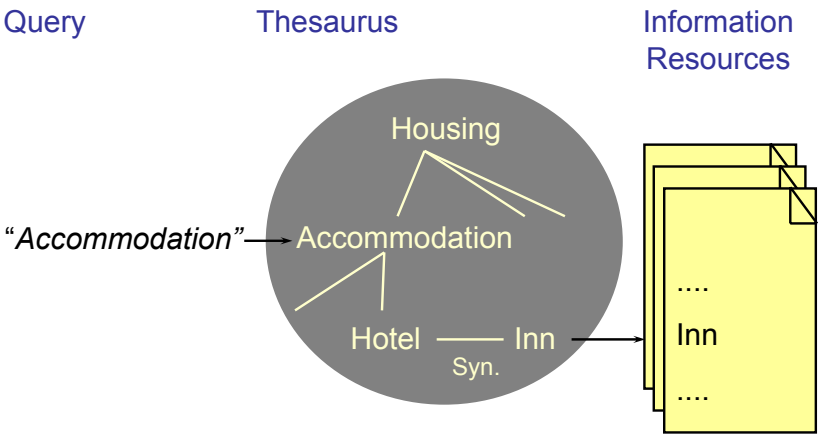
- Thesaurus to support specification and processing of thematic references
- analogously: „Gazetteer“ to support specification and processing of spatial references
- Handling of temporal references: requires some basic temporal reasoning faciulities

THESAURUS

A Thesaurus is a structured collection of terms with the following properties:

- Terms provide a controlled vocabulary for the specification of thematic references,
- Terms can be used for both indexing and retrieval.
- Terms are more than simple keywords.
- Terms form a semantic network established by:
  - ⇨ synonym relationship (inn - hotel)
  - ⇨ generalization hierarchy of broader / narrower terms (accommodation - hotel)
  - ⇨ linkage via related terms (accommodation - tourism)

THESAURUS-SUPPORTED QUERY PROCESSING



BLACK BOX SEARCH PROBLEM: A THESAURUS CAN HELP

Problem: Information resources are searched for by using a form in most metainformation systems ("black box search")

- It is not clear which level of detail is required while specifying a query
  - ⇨ Many casual users dislike form-based search interfaces

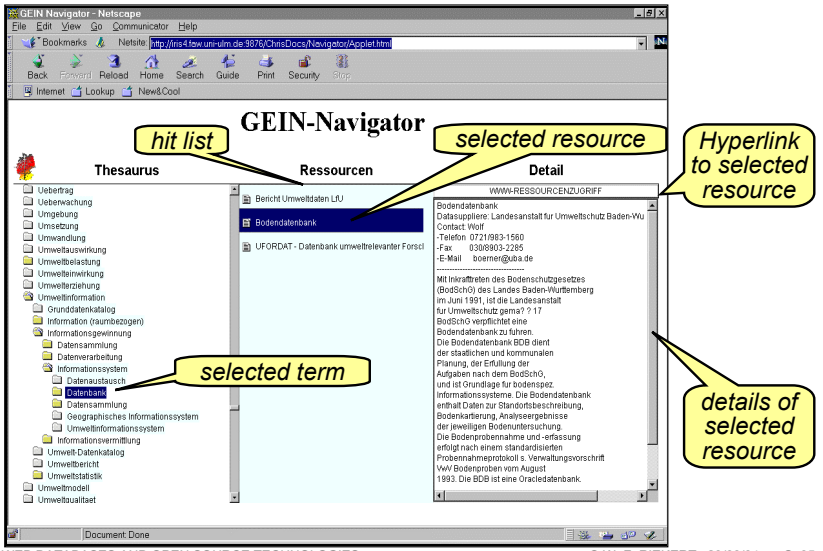
Requirement: Hierarchical directories to access the information resources

- However: Manual maintenance of hierarchical directories very time-consuming

Solution: Use a thesaurus for the automated generation of a hierarchical directory

Example: GEIN Navigator (prototype developed at FAW Ulm)

PROTOTYPICAL GENERATION OF A HIERARCHICAL DIRECTORY



A PROCEDURE TO GENERATE A HIERARCHICAL DIRECTORY

- Create a "weeded" thesaurus consisting of all relevant terms, i.e.:
  - ⇨ take all terms used as an index for existing information resources,
  - ⇨ add recursively all broader terms,
  - ⇨ disregard all other terms
- Display thesaurus in a hierarchical presentation (Windows Explorer-like), starting from "toplevel terms"
- Special highlighting indicates which terms
  - ⇨ directly lead to hits,
  - ⇨ possess narrower terms leading to hits
- Provide navigation paths to the metainformation records and from there to the original information resources

METAINFORMATION SYSTEMS VS. SEARCH ENGINES

FACHHOCHSCHULE STUTTGART
HOCHSCHULE DER MEDIEN

Metainformation system:

- Easy retrieval by using semantical criteria
- But: Indexing very expensive for administrators or information suppliers

Search engine:

- Indexing very easy, no work imposed on suppliers
- But: only textual processing of search criteria

Synthesis:

- Combination of the advantages of search engines and metainformation systems: Thesaurus-based preprocessor for search engines

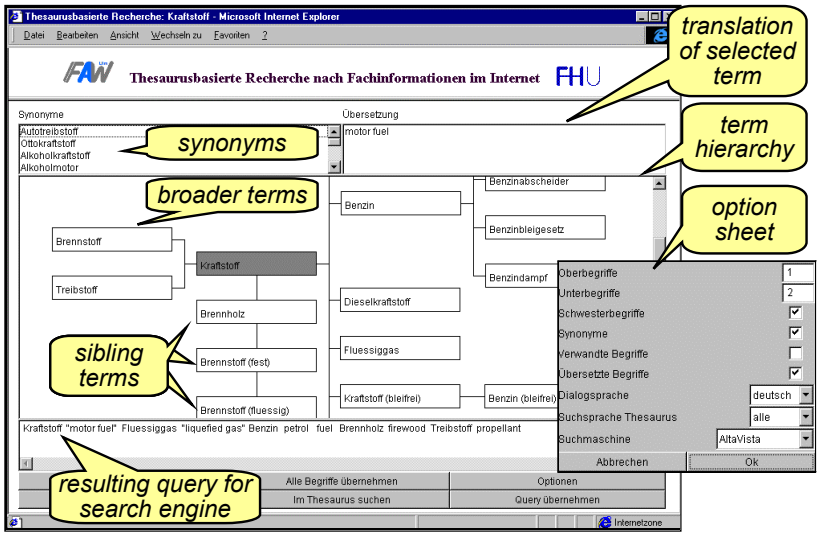
COMBINE THE ADVANTAGES

FACHHOCHSCHULE STUTTGART
HOCHSCHULE DER MEDIEN

	Indexing inexpensive	Semantic processing of search terms
search engine	✘	—
metainformation system	—	✘
search engine with thesaurus-based preprocessor	✘	✘

THESAURUS-BASED PREPROCESSOR FOR SEARCH ENGINES

FACHHOCHSCHULE STUTTGART
HOCHSCHULE DER MEDIEN

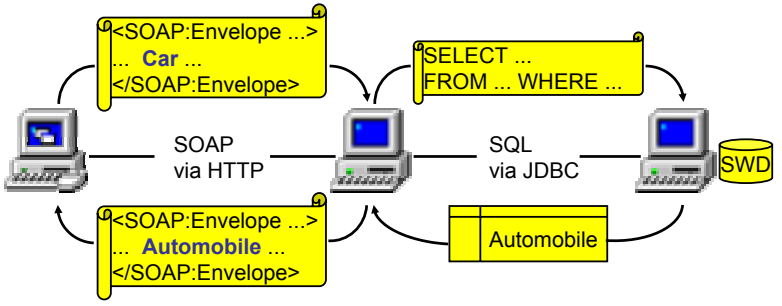


EXAMPLE: SWD WEBSERVICE A THESAURUS WEBSERVICE

FACHHOCHSCHULE STUTTGART
HOCHSCHULE DER MEDIEN

- Purpose: Make the **SWD thesaurus** available to other applications, particularly catalog systems, as a **webservice**
  - ⇨ SWD ("Schlagwortnormdatei"), a thesaurus used in German libraries for indexing and retrieval purposes
  - ⇨ SWD is copyrighted, the service approach avoids deliverance of the full data corpus
  - ⇨ Prototype system to explore webservice potential
- User: **Library Service Centre Baden-Württemberg** (BSZ – Bibliotheksservice-Zentrum Baden-Württemberg)
- Developer: **Wolfgang Habel**, M.A.
  Master Thesis in Library and Media Management at HdM Stuttgart, 2003, supervisor: W.-F. Riekert (http://v.hdm-stuttgart.de/~riekert/theses/master-habel.pdf)
- Approach: Development in **Java** (Jakarta / AXIS) using the Simple Object Access Protocol (**SOAP**)

SWD WEBSERVICE: HOW IT WORKS



Webservice Client, e.g., OPAC Application

SWD Webservice implemented in Java/Axis on Tomcat Webserver

SWD Thesaurus on MYSQL Database Server

CONCLUSION AND OUTLOOK

- Open Source provides powerful tools for software development
- Strong support for service-oriented software systems
  - ⇨ Web applications
  - ⇨ Web services
- Inexpensive approach, especially suited for academic projects
- Results nevertheless of high interest for industrial scenarios
- Open source community is supranational
  - ⇨ Favors joint projects, e.g. between Hungary and Baden-Württemberg
- A lot of interesting things can be done together!