

9 Topic Maps

Topic Maps – The Idea

- Topic Maps ...
 - ◆ are a formalism for the representation and exchange of knowledge
 - ◆ can represent meta knowledge about information resources (documents, images, web pages ...)
 - ◆ represent knowledge in the form of semantic networks
- Objective of topic maps
 - ◆ enhance searching and browsing for information
 - ◆ increase quality and speed of finding information
 - ◆ integration various information sources

Basic Constructs of Topic Maps

Topic maps represent domain knowledge using...

Topics

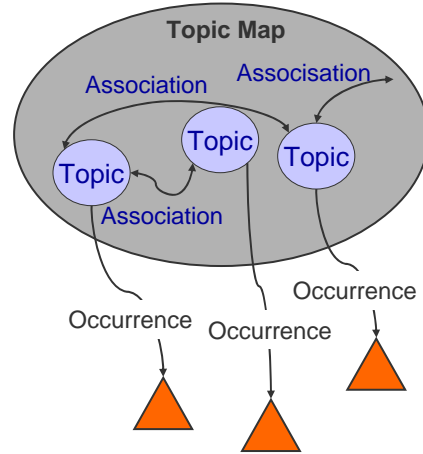
representing any concept like people, things, organizations, events

Associations

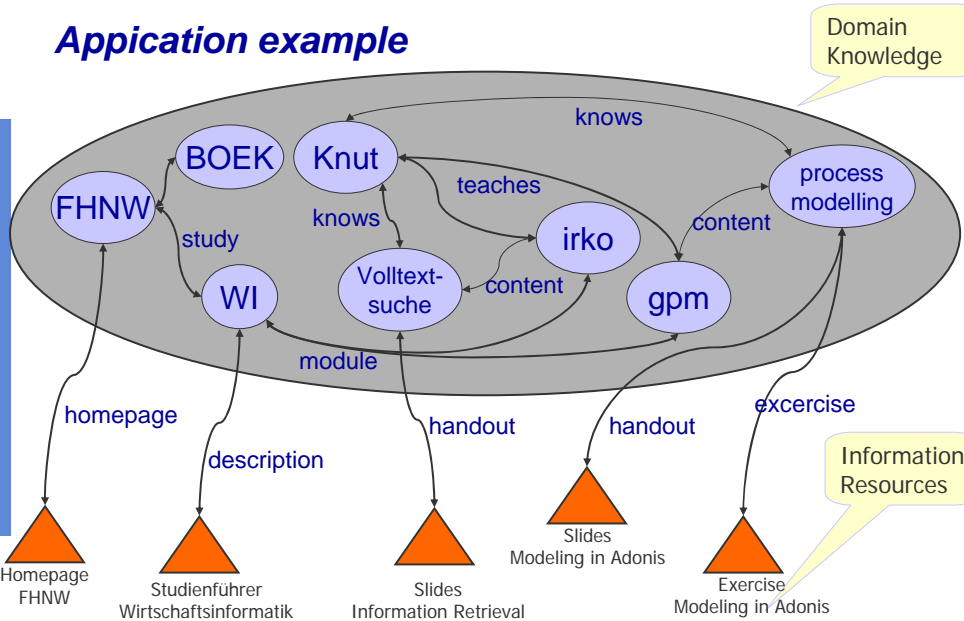
representing relationships between topics

Occurrences

representing relationships between topics and information resources



Application example



Topic Map Standardisation

There are two syntax standards for the representation and interchange of topic maps

- ISO 13250 Topic Maps (HyTime)
 - ◆ ISO/IEC 13250 is a standard from ISO based on SGML
 - ◆ siehe <http://www.ornl.gov/sgml/sc34/document/0058.htm>
- XML Topic Maps (XTM)
 - ◆ XTM is a realisation of the ISO 13250 topic map paradigm in XML
 - ◆ XTM is provided by TopicMaps.org, an independent consortium of parties that develops the applicability of the topic map paradigm to the World Wide Web by leveraging the XML family of specifications
 - ◆ siehe <http://www.topicmaps.org/xtm/1.0>



XTM – XML Topic Maps

- The XTM standard is made up of the following concepts:
 - ◆ Topic Maps
 - ◆ Topics
 - ◆ Topic Names
 - ◆ Associations
 - ◆ Occurrences
 - ◆ Public Subject Descriptor
 - ◆ Scopes

The description in these slides are based on XTM 1.0 (<http://www.topicmaps.org/xtm/1.0>)
In the meantime there is a proposal for XTM 2.0 (siehe <http://www.isotopicmaps.org/sam/sam-xtm/>),
which has not yet been accepted and thus there are no systems available for it.



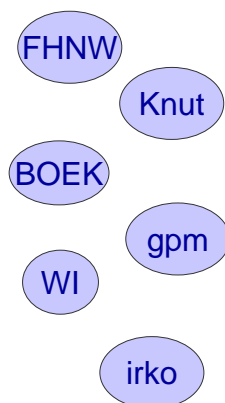
Topic Maps

```
<topicMap id="fhso-topicmap"
  xmlns="http://www.topicmaps.org/xtm/1.0/"
  xmlns:xlink="http://www.w3.org/1999/xlink">
  <topic> ...
</topic>
  <topic> ...
</topic>
  <association> ...
</association>
  <mergeMap> ...
</mergeMap>
</topicMap>
```

- A topic map is a collection of topic map nodes
 - ◆ topics
 - ◆ associations
 - ◆ mergeMaps
- `<topicMap>` is the root element of a topic map and can contain `<topic>`, `<association>` and `<mergeMap>` elements



Topics, Subjects, Reification



- Topics represent elementary subjects
- A subject can be everything we can talk or think about, e.g.
 - ◆ person, sentences, country, feeling, object, word, number, relationship, ...
- To make assertions about a subject in a topic map one has to create a topic for this subject. This is called *reification*.



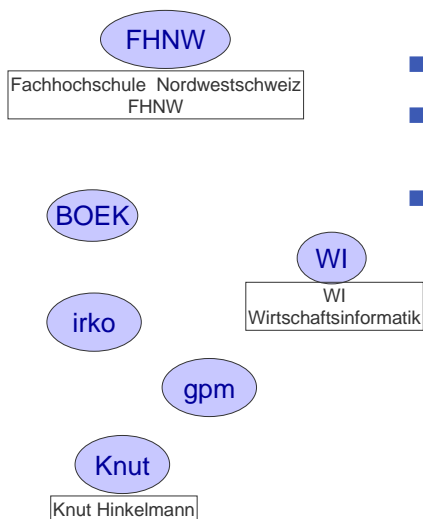
Topics

```
<topic id="FHNW" />  
<topic id="WI" />  
<topic id="BOEK" />  
<topic id="Knut" />  
<topic id="gpm" />  
<topic id="irko" />  
<topic id="kgnklasghsag" />
```

- Topics are represented by the XML element `topic`
- Each topic has an ID
- The ID is unique within a topic map document
- The ID need not to be a meaningful name. For reasons of readability, a meaningful ID should be used.



Topic Names



- Topics have names
- In contrast to an ID the name should have a meaning for people.
- A topic may have zero or more names, e.g.
 - ◆ short name
 - ◆ abbreviation
 - ◆ name that is used for sorting
 - ◆ print name



Topic Names

```
<topic id="FHNW">
  <baseName>
    <baseNameString>
      Fachhochschule Nordwestschweiz
    </baseNameString>
  </baseName>
</topic>
```

- Each name may exist in multiple forms.
- A name always has exactly one base form, known as the base name
- A *base name* is the base form of a topic name; it is always a string

In XTM 2.0 `baseName` will be `name` and `baseNameString` is dropped

Variant Names

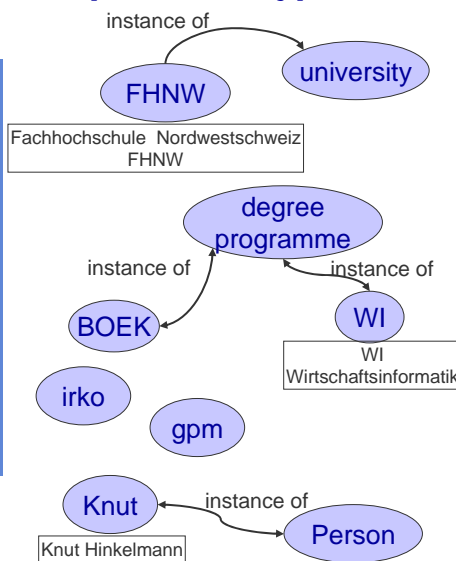
- A name may have one or more variants for use in specific processing contexts:
 - ◆ A *variant name* is an alternative form of a base name, that is optimized for a particular computational purpose, such as sorting or display
 - ◆ It may be any kind of a resource, including a string.
 - ◆ An application chooses among variant names by evaluating their parameters.
- Parameters are information, in the form of a set of topics, that expresses the appropriate processing context for a variant name

Topic Names: Example

```
<topic id="FHNW">
  <baseName>
    <baseNameString>
      Fachhochschule Nordwestschweiz
    </baseNameString>
    <variant>
      <parameters> <topicRef xlink:href="#display"/> </parameters>
      <variantName>
        <resourceData> FHNW </resourceData>
      </variantName>
    </variant>
  </baseName>
</topic>
```

In XTM 2.0 baseName will be name the elements baseNameString and variantName will be dropped

Topics and Types



- **Types:** A topic can have one or multiple types
- The type relation is called *instance of*
- *Types of topics are themselves topics* and therefore must be represented using a topic element (reification)
- **Type hierarchy:** Types can be organized hierarchically

Topic Types

```
<topic id="FHNW">
  <instanceOf>
    <topicRef xlink:href="#university"/>
  </instanceOf>
  <baseName>
    <baseNameString>
      FHNW
    </baseNameString>
  </baseName>
</topic>
```

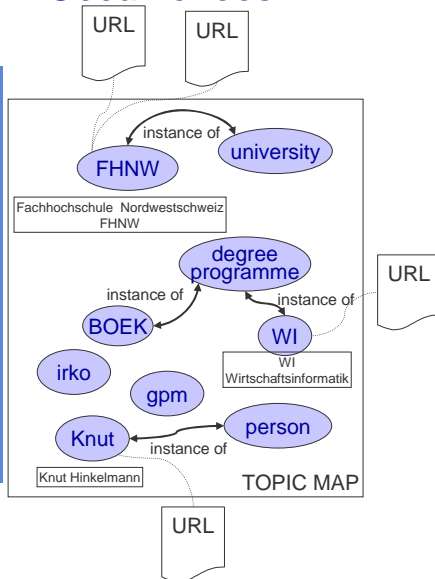
- A topic can have an arbitrary number of types
- types are enclosed in the element <instanceOf>
- Each type is a topic itself
 - ◆ in the same map


```
<topicRef
  xlink:href="#university"/>
```

 (university is defined as a topic)
 - ◆ or in another document, e.g.


```
<topicRef xlink:href=
  "univ.xtm#university"/>
```

Occurrences



- An occurrence is any information that is specified as being relevant to a given subject.
- Occurrences must be resources that are either
 - ◆ addressable by reference using a URI (a “resource reference”), or
 - ◆ placed inline as character data (“resource data”).
- A topic can have multiple occurrences
- Occurrences may have types (instanceOf)
- Occurrence types are themselves topics

Occurrences

```
<topic id="FHNW">
  <baseName>
    <baseNameString>
      FHNW
    </baseNameString>
  </baseName>
  <occurrence>
    <resourceData>
      One of 8 universities of applied
      sciences in Switzerland
    </resourceData>
  </occurrence>

  <occurrence>
    <resourceRef
      xlink:href="http://www.fhnw.ch"/>
    </resourceRef>
  </occurrence>
</topic>
```

Occurrences are either:

- ◆ internal:
 - placed inline as character data `<resourceData>`
- ◆ external resource:
 - addressable by reference using a URI `<resourceRef>`

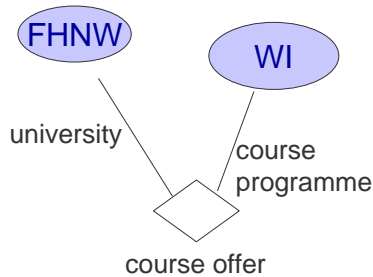
Occurrence Types

```
<topic id="FHNW">
  ..
  <occurrence>
    <instanceOf>
      <topicRef xlink:href="#homepage"/>
    </instanceOf>

    <resourceRef
      xlink:href="http://www.fhnw.ch"/>
    </resourceRef>
  </occurrence>
</topic>
```

Every Occurrence can have a type, given the the element `<instanceOf>`

Associations



In this example
"course offer" is the
association type
"university" and "course
programme" are roles

- An association is a relationship between one or more topics
- Each association has an *association type*
- There is no directionality inherent in an association
 - ◆ If *A* is related to *B*, then *B* must also be related to *A*.
- The issue is rather,
 - ◆ what is the *type* of the relationship, and
 - ◆ what *roles* are played by its members



Associations (cont.)

```
<association>
  <instanceOf>
    <topicRef xlink:href="#course-offer"/>
  </instanceOf>
  <member>
    <topicRef xlink:href="#FHNW"/>
  </member>
  <member>
    <topicRef xlink:href="#WI"/>
  </member>
</association>
```

- Associations are represented using the element `<association>`
- Each association is an instance of a single class of association (also known as an *association type*).
- This association type is given the the element `instanceOf`.
- The topics in the relationship are given by the element `<member>`
- An association can have arbitrary number of members



Associations (cont.)

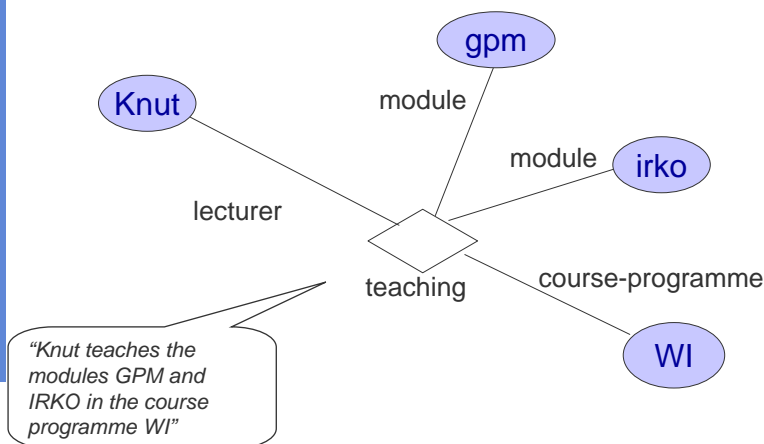
```
<association>
  <instanceOf>
    <topicRef xlink:href="#course-offer"/>
  </instanceOf>
  <member>
    <roleSpec>
      <topicRef xlink:href="#university"/>
    </roleSpec>
    <topicRef xlink:href="#FHNW"/>
  </member>
  <member>
    <roleSpec>
      <topicRef xlink:href="#course-programme"/>
    </roleSpec>
    <topicRef xlink:href="#WI"/>
  </member>
</association>
```

- Each topic in the relationship plays a role as a member of that association
- The role is given by the element `roleSpec`
- Role specifications are optional



Roles in an Association

An association can have arbitrary number of members
Multiple topics can have the same role



```

<association>
<instanceOf>
  <topicRef xlink:href="#teaching"/>
</instanceOf>
<member>
  <roleSpec>
    <topicRef xlink:href="#lecturer"/>
  </roleSpec>
  <topicRef xlink:href="#Knut"/>
</member>
<member>
  <roleSpec>
    <topicRef xlink:href="#course-programme"/>
  </roleSpec>
  <topicRef xlink:href="#wi"/>
</member>
<member>
  <roleSpec>
    <topicRef xlink:href="#module"/>
  </roleSpec>
  <topicRef xlink:href="#irko"/>
  <topicRef xlink:href="#gpm"/>
</member>
</association>

```

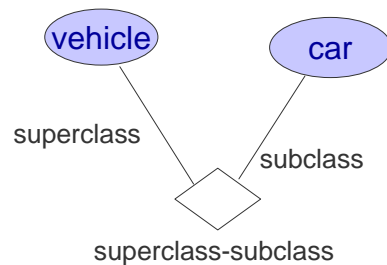
XTM representation of the association

“Knut teaches the modules GPM and IRKO in the course programme WI”

(see previous slide)



Superclass-subclass Relationship



- Type hierarchies are represented as superclass-subclass relationships
 - ◆ association type is Superclass-subclass
 - ◆ roles are superclass and subclass



Subject Indicators

- Subject identity is the means to established which subject is reified by a particular topic
- Subject identity can be established in one of two ways:
 - ◆ By addressing the subject directly. This is only possible when the subject is an addressable information resource.
 - ◆ By indicating the subject via a subject indicator
- A subject indicator is a resource that is intended by the topic map author to provide a positive, unambiguous indication of the identity of a subject.
- There are three ways of indicating a subject in a topic map:
 - ◆ Pointing via a <subjectIndicatorRef> element to a resource that **indicates** the subject;
 - ◆ Pointing via a <resourceRef> element to a resource that **is** the subject.
 - ◆ Pointing via a <topicRef> element to a <topic> element that shares the same subject;
- When two topics have the same subject identity, they are considered to be “about” the same thing



Subject Indicator with SubjectIndicatorRef

- The <subjectIndicatorRef> element provides a URI reference to a resource that acts as a subject indicator
- Example:

```
<topic id="FHNW">
  <baseName>
    <baseNameString>
      Fachhochschule Nordwestschweiz
    </baseNameString>
  </baseName>
  <subjectIdentity>
    <SubjectIndicatorRef
      xlink:href="www.so.ch/extappl/gs/daten/gs2005/100_126_050504.pdf"/>
    </ subjectIdentity >
</topic>
```

The <subjectIndicatorRef> references the „Staatsvertrag“, a document that unambiguously identifies the FHNW.



Subject Indicator with topicRef

Example:

```
<topic id="FHSO">
  <baseName>
    <baseNameString>
      Fachhochschule Solothurn
    </baseNameString>
  </baseName>
  <subjectIdentity>
    <topicRef xlink:href="#FHNW"/>
  </subjectIdentity >
</topic>

...

<topic id="FHNW">
...
</topic>
```

- <topicRef>s are identical to <subjectIndicatorRef>s, except for the additional constraint that they must point to <topic> elements.
- The target <topic> need not be in the same document.

The topic FHSO is identical to the topic FHNW, because its identity is defined by a topicRef-reference to FHNW



Subject Indicator with ResourceRef

- The <resourceRef> element provides a URI reference to a resource.
- The value auf the <resourceRef> element in a subject indicator **is** the subject represented by the topic.
- Example:

```
<topic id="FHNW-homepage">
  <baseName>
    <baseNameString>
      Homepage FHNW
    </baseNameString>
  </baseName>
  <subjectIdentity>
    <resourceRef xlink:href="http://www.fhnw.ch"/>
  </subjectIdentity >
</topic>
```

This topic is about the homepage of FHNW (and not about FHNW itself) referenced by resourceRef



Example: Two Topics with the same Subject Indicator

```
<topic id="FHNW-homepage">
  <baseName>
    <baseNameString> Homepage FHNW </baseNameString>
  </baseName>
  <subjectIdentity> <resourceRef xlink:href="http://www.fhnw.ch"/>
</subjectIdentity >
</topic>
```

```
<topic id="FHNW_im_WWW">
  <baseName>
    <baseNameString> Website of Fachhochschule Nordwestschweiz
  </baseNameString>
  </baseName>
  < subjectIdentity > <resourceRef xlink:href="http://www.fhnw.ch"/>
</subjectIdentity >
</topic>
```

Both topics have the same subject indicator and thus are about the same subject.



Published Subject Indicators

- A published subject indicator is any resource that has been published in order to provide a positive, unambiguous indication of the identity of a subject.
- Examples for published subject indicators are
 - ◆ ISBN for books
 - ◆ the social security number for people
- Authors are encouraged to always indicate the subject identity of their topics in the most robust manner possible, in particular through the use of standardized ontologies expressed as published subject indicators.



Topic Characteristic

- Anything that may be asserted about a topic is known as a characteristic of that topic.
- Characteristics can be one of the following:
 - ◆ a topic name,
 - ◆ a topic occurrence, or
 - ◆ a role played by a topic as a member of an association
- The assignment of such characteristics is considered to be valid within a certain scope, or context.



Scopes

- Scope specifies the extent of the validity of a topic characteristic assignment. It establishes the context
 - ◆ in which a name or an occurrence is assigned to a given topic, and
 - ◆ the context in which topics are related through associations.
- Every characteristic has a scope, which may be specified either
 - ◆ explicitly, as a set of topics, or
 - ◆ implicitly, in which case it is known as the unconstrained scope.

Assignments made in the unconstrained scope are always valid.

- The interpretation of a characteristic's scope and its effect on processing is left to the application

An exception is the *topic naming constraint*: topics having the same base name in the same scope implicitly refer to the same subject and therefore should be merged.



Scopes (cont.)

- Scope of a characteristic is given in the element `<scope>`
- The specification of the scope is given by
 - ◆ `<topicRef>`,
 - ◆ `<subjectIndicatorRef>` or
 - ◆ `<resourceRef>`
- elements (cf. subjectIndicators)
- Scope are defined topics
- If no scope is given, the characteristics are always valid: unconstrained scope



Scope – Example: German and English Names

```
<topic id="FHNW">
  <baseName>
    <scope>
      <topicRef xlink:href="#DE" />
    </scope>
    <baseNameString>
      Fachhochschule Nordwestschweiz
    </baseNameString>
  </baseName>
  <baseName>
    <scope>
      <topicRef xlink:href="#EN" />
    </scope>
    <baseNameString>
      University of Applied Sciences Northwestern Switzerland
    </baseNameString>
  </baseName>
  ...
</topic>
```

The Topics DE and EN must be defined in the topic map.



Merging

- The term *merging* covers two distinct processes:
 - ◆ merging two *topic maps*, either as a result of explicit `<mergeMap>` directives, or for any application-specific reasons.
 - ◆ merging *two topics*.
- The rules governing all forms of merging can be briefly (and incompletely) stated as follows:
 - ◆ When two topic maps are merged,
 - any topics that the application determines to have the same subject are merged, and
 - any duplicate associations are removed.
 - ◆ When two topics are merged, the result is a single topic whose characteristics are the union of the characteristics of the original topics, with duplicates removed.

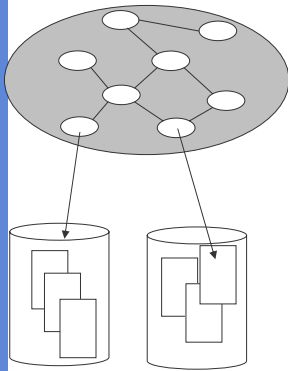


Determination of Subject Identity in Merging

- Two topics are always deemed to have the same subject if:
 - ◆ they have one or more subject indicators in common,
 - ◆ they reify the same addressable subject, or
 - ◆ they have the same base name in the same scope.



Topic Map Applications



- Web Portals
 - ◆ Use of topic maps for navigation in portals:
Generation of HTML pages from topic maps
- Knowledge-based Intranets
 - ◆ Semantic Networks for the representation of subject of an organisation (staff, products, processes, roles) and relationships and links to relevant documents
- Content-Management Systems
 - ◆ Organisation of content with a Topic Map
- Enterprise Application Integration
 - ◆ Integration of information from various sources



Example Application for Topic Maps: Online Product Catalogue

- Content: Products and accessories
 - ◆ different types of resources (occurrences) like images, description, technical specification
- Associations: Relationships of
 - ◆ Products with suitable accessories
 - ◆ products and product components
 - ◆ products and their application domains
 - ◆ accessories and application domains
- Navigation: Links from
 - ◆ Links from products to suitable and available accessories
 - „suitable graphic cards for notebooks“
 - ◆ Components to Products: "PC with internal DVD drive"
 - ◆ application to accessories: „Accessories for use in a network“



Example Application for Topic Maps : Project Repository in an Intranet

Automatically generate an Intranet from a topic map

- Content: Projects, staff and used technologies
 - ◆ different types of resources (Occurrences) like project description, competence profiles, manuals
- Associations:
 - ◆ Technologies used in Projects
 - ◆ members/managers of a project
 - ◆ clients of projects
 - ◆ competences of employees
 - ◆ ...
- Navigation: Links from
 - ◆ projects to used technologies
 - ◆ technologies to applications in projects
 - ◆ ...

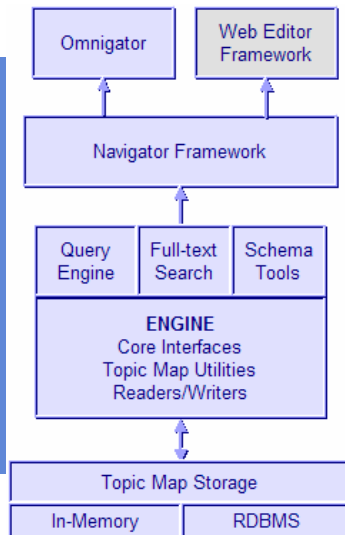


Topic Map Tools

- There are a number of commercial and free topic map tools, e.g.
- Free / Open Source:
 - ◆ SemanText (<http://www.semantext.com>)
 - ◆ TM4J (<http://tm4j.sourceforge.net> or <http://tm4j.org>)
 - ◆ Nexist (<http://nexist.sourceforge.net>)
 - ◆ Techquila Open source topic map (<http://www.techquila.com>)
- Commercial Systems from
 - ◆ Ontopia (www.ontopia.net)
 - ◆ Empolis (www.empolis.com)
 - ◆ Mondeca (www.mondeca.com)
 - ◆ Infloom (www.infloom.com)
 - ◆ NetworkedPlanet (<http://www.networkedplanet.com/>)



Exempel: Die Ontopia Knowledge Suite



Set of products for the development of topic map applications

Topic Map Engine: Java SDK with functionalities for software using topic maps: loading, storing of topic maps, access to topic maps

Navigator Framework for the development of web applications based on topic maps

Web Editor Framework for client-specific web editors

Vizigator for graphical visualisation of topic maps

Omnigator: demo tool for browsing in topic maps



Omnigator

- Omnigator is a generic browser implemented using the Navigator Framework from Ontopia
 - ◆ Objective: Instruction tool
 - understanding the concept of topic maps
 - easy creation and test of topic maps
- Functionalities of Omnigator
 - ◆ display of topics, associations, occurrences etc.
 - ◆ plugins for search and visualisation
- The Omnigator interface is not suitable for end users
 - ◆ end users do not need to know that an application has been implemented with topic maps
 - ◆ end users do not need to know about "Topic", "Association" or "Scope".



A page in Omnigator

The screenshot shows the Omnigator application interface. At the top, there is a navigation bar with 'Home', 'Manage', 'Website', 'Support', and 'About'. Below this is a search bar with 'omnigator' and a 'Find' button. The main content area is titled 'Puccini, Giacomo' and is divided into several sections:

- Untyped Names (2):**
 - Puccini, Giacomo
 - Puccini - Scope: Short name
- Normal form (1):**
 - Giacomo Puccini
- Associations (18):**
 - Born in: Lucca
 - Composed:
 - La Bohème
 - Edgar
 - La fanciulla del West
 - Gianni Schicchi
 - Madama Butterfly
 - Manon Lescaut
 - La rondine
- Subject Identifiers (1):**
 - http://en.wikipedia.org/wiki/Puccini
- Internal Occurrences (4):**
 - Bibliography:
 - Budden, Julian: "Puccini: His Life and Works", Oxford University Press (Oxford, 2002)
 - Sadie, Stanley (ed): "Puccini and His Operas", Macmillan (London, 2000)
 - Date of birth: 1858-12-22
 - Date of death: 1924-11-29
- External Occurrences (10):**
 - Article:
 - file://C:/topicmaps/opera/occurs/sn/puccini.htm - Scope: Offline; Store Norske Leksikon
 - http://www.ontopia.net/topicmaps/examples/opera/occurs/sn/puccini.htm - Scope: N
 - Gallery:
 - file://C:/topicmaps/opera/occurs/puccini-gallery.htm - Scope: Offline
 - Illustration:
 - file://C:/Programme/ontopia/oks-professional-3.1.0/apache-tomcat/webapps/omnigator
 - Sound clip:
 - http://www.puccini.it/files/vocepucc.wav - Scope: Centro studi Giacomo Puccini; Italian;
 - Web page:
 - file://C:/topicmaps/opera/occurs/hnh-puccini.htm - Scope: Naxos; Offline

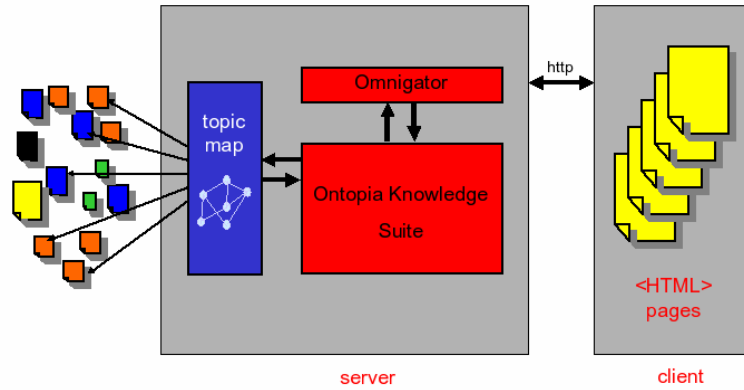
Application-specific page of the same content

The screenshot shows an application-specific page titled 'The Italian Opera Topic Map'. It features a navigation menu on the left with categories: Operas, Composers (highlighted), Librettists, Writers, Theatres, Cities and Regions, and Countries. The main content area is for 'Puccini, Giacomo' and includes:

- A portrait of Giacomo Puccini.
- Biographical information: Italian composer. Born 1858 (22 Dec) in Lucca. Died 1924 (29 Nov) in Brussels.
- Operas:**
 - La Bohème (Florence) 1896 (1 Feb)
 - Edgar 1869 (21 Apr)
 - La Fanciulla del West 1910 (10 Dec)
 - Gianni Schicchi 1918 (14 Dec)
 - Madama Butterfly 1904 (17 Feb)
 - Manon Lescaut 1893 (1 Feb)
 - La rondine 1917 (27 May)
 - Scenes from the Marriage of Figaro 1885 (14 Dec)
 - Il Tabarro 1918 (14 Dec)
 - Traviata 1875 (14 Jan)
 - Il Trittico
 - Turandot 1926 (25 Apr)
 - La Villi 1884 (31 May)
- Other sites:**
 - http://www.italianopera.com/puccini/gallery.htm
 - http://www.hnh.com/composers/puccini.htm
 - file://C:/topicmaps/opera/occurs/sn/puccini.htm

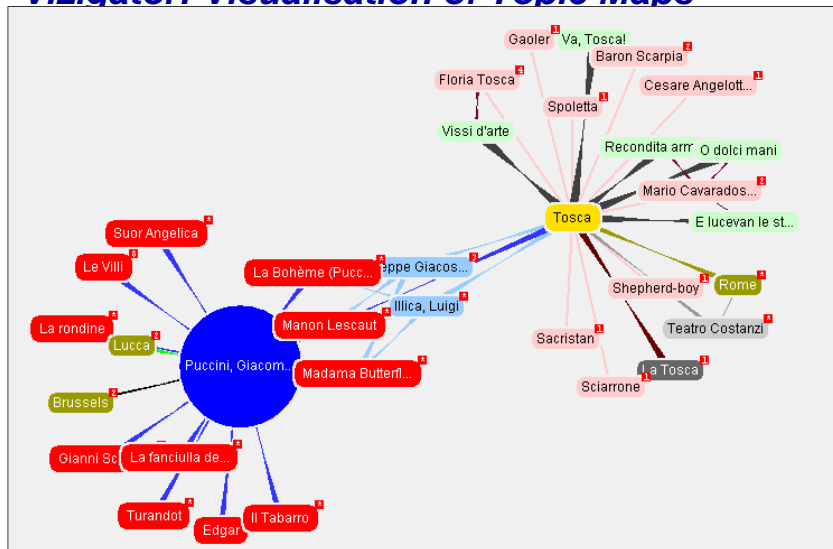
At the bottom, there is a search bar for 'Search Operabase for performances of operas by Giacomo Puccini' and a navigation bar with 'City', 'Country', 'Title', 'Date', and 'Search' buttons.

Architecture of Omnigator



- Omnigator loads topic maps and generates HTML pages
- Every click on a links sends a request to the server application, which extracts the requested information from the topic map and generates an HTML page

Vizigator: Visualisation of Topic Maps



Fulltext Search in Topic Maps

- Fulltext search in Omnigator takes advantage of knowledge represented in a topic map
- If multiple topics satisfy the query, the user is supported in finding the one with the correct meaning.

„There are various topics for Tosca. Which one do you mean:

- the opera
- the play
- the person Floria Tosca
- the Aria Va Tosca!

| Topic | Type | Match in | Score |
|--------------|-----------|----------|-------|
| Tosca | Opera | basename | 100% |
| La Tosca | Play | variant | 62% |
| Floria Tosca | Character | basename | 62% |
| Va, Tosca! | Aria | basename | 62% |

Querying Topic Maps

- Targeted search for topics, associations and occurrences satisfying specific conditions
- Query languages corresponds to "SQL for Topic Maps"

All opera composed by Verdi

```

Query:
using op for i"http://pai.ontopia.net/music/#"
using lit for i"http://pai.ontopia.net/literature/#"
op:composed-by(4OPERA : lit:work, verdi : op:composer)
order by #OPERA
    
```

| OPERA |
|-------------------------|
| Aida |
| Alzira |
| Aroldo |
| Attila |
| Un ballo in maschera |
| La battaglia di Legnano |
| Il corsaro |

Execution time 20 millisecs
Result rows 28