

Authoring and Annotation of Web Pages in CREAM



Siegfried Handschuh, Steffen Staab

Institute AIFB

Knowledge Management Group

University of Karlsruhe, Germany

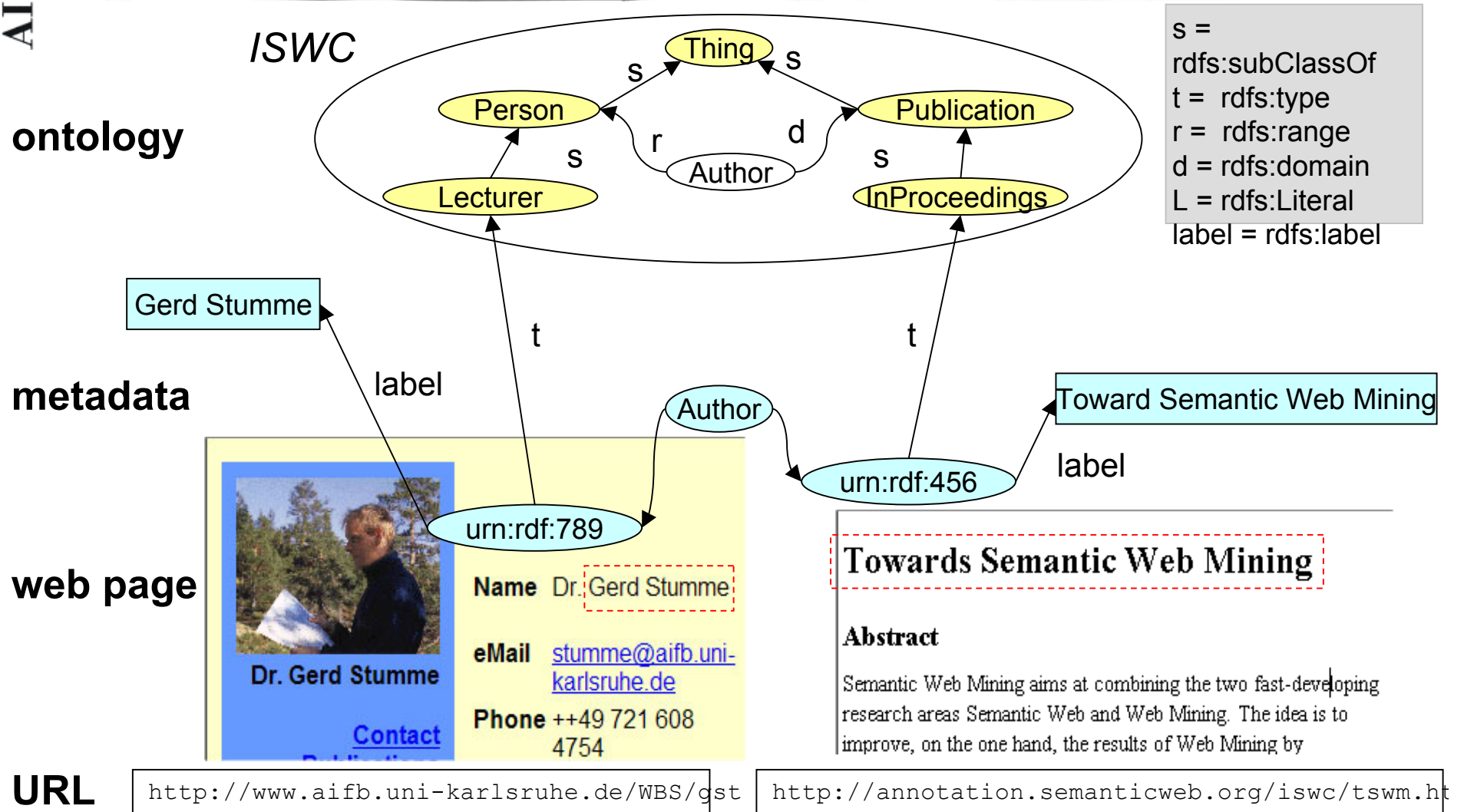
Overview

- Introduction
- Relational Metadata
- Modes of Interaction
- Meta Ontology
- Outlook & Summery

Introduction

- Creating **RE**lational **A**nnotation-based **M**etadata for the Semantic Web
- Annotation Frameworks
 - **F**irst create content, **s**econd annotate content
- **Propose**: Combine Annotation and Authoring
- Hide the **B**order between Annotation and Authoring
- Extend CREAM Framework and **OntoMat-*Annotizer***, requires
 - **M**odes of Interaction
 - **M**eta Ontology

Relational Metadata



Modes of Interaction - GUI

The screenshot shows an 'Ontology Browser' window with a tree view on the left and a document viewer on the right. The tree view includes categories like 'Formal_Language', 'Method', 'Organization', 'Person', 'Project', 'Publication', and 'Book'. The 'Towards Semantic Web Mining' document is selected. A table below shows attributes and values for this document, including 'title' and 'author (Person)'. The 'author' list includes Bettina Berendt, Gerd Stumme, and Andreas Hotho. Three yellow callout boxes provide definitions:

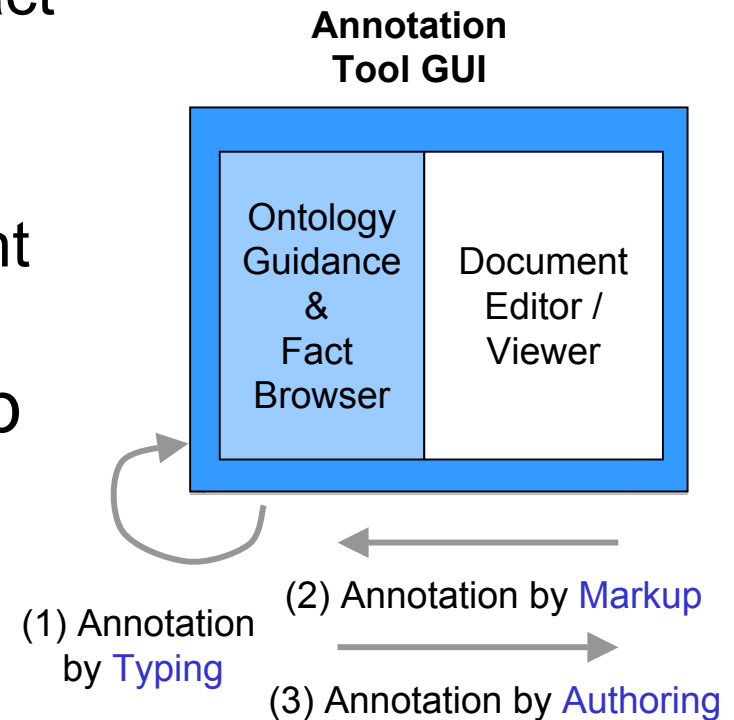
- Concepts**: Points to the 'Person' class in the tree view.
- Guidance & Instances of Concepts**: Points to the 'author (Person)' instance in the table.
- Attribute Instances = instance of a property to a datatype instance**: Points to the 'title' attribute in the table.
- Relationship Instances = instance of a property to a class instance**: Points to the 'author' relationship in the table.

A separate callout box on the right is labeled 'Document Viewer / Editor'.

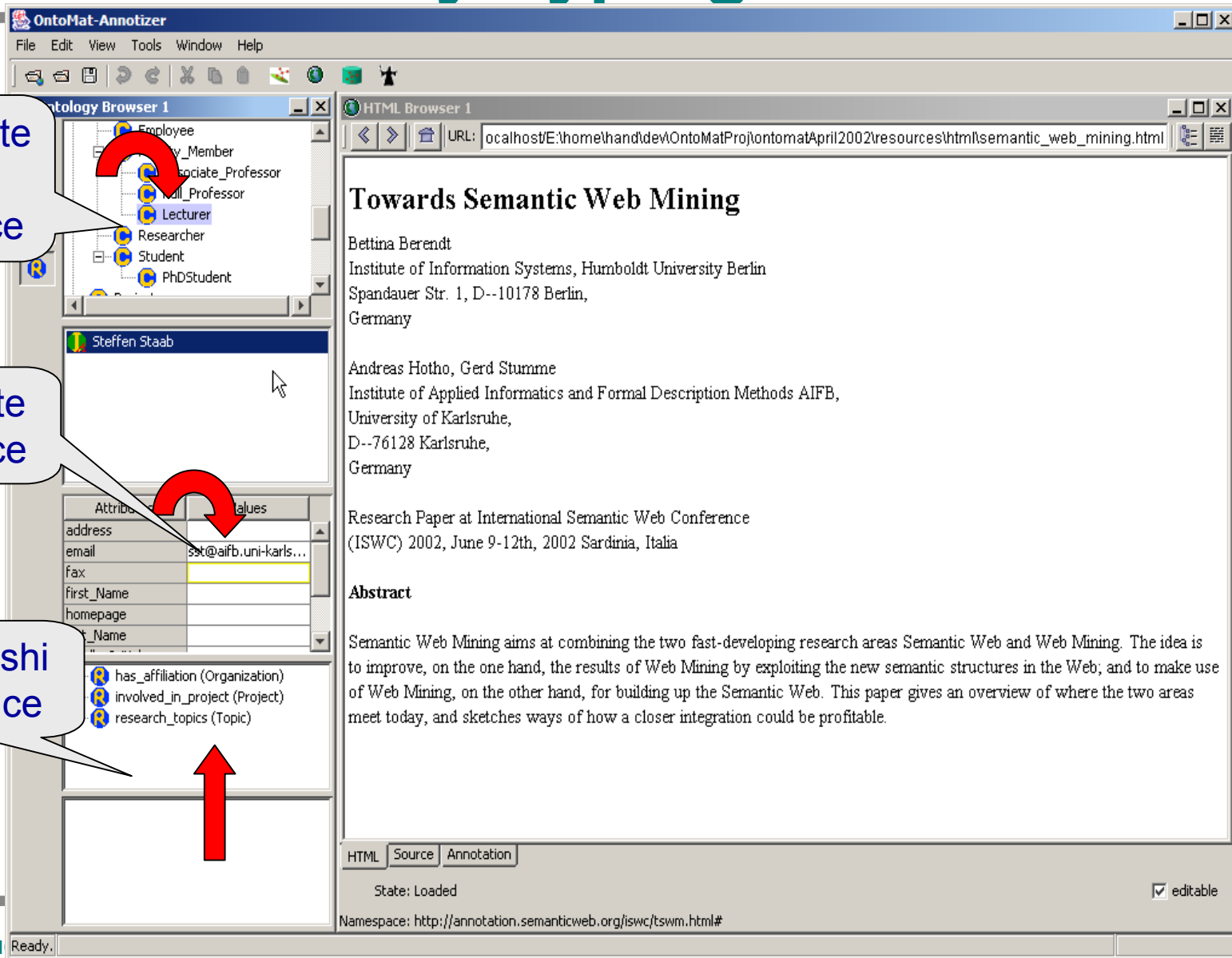
Modes of Interaction



1. Annotation by **Typing** Statements
 - Exclusively with Ontology & Fact Browser
2. Annotation by **Markup**
 - Reuse Data from the Document Editor in the Fact Browser
3. Annotation by **Authoring** Web Pages
 - Reuse Data from the Fact Browser in the Document Editor



Annotation by Typing Statements



The screenshot shows the OntoMat-Annotizer application with three main components:

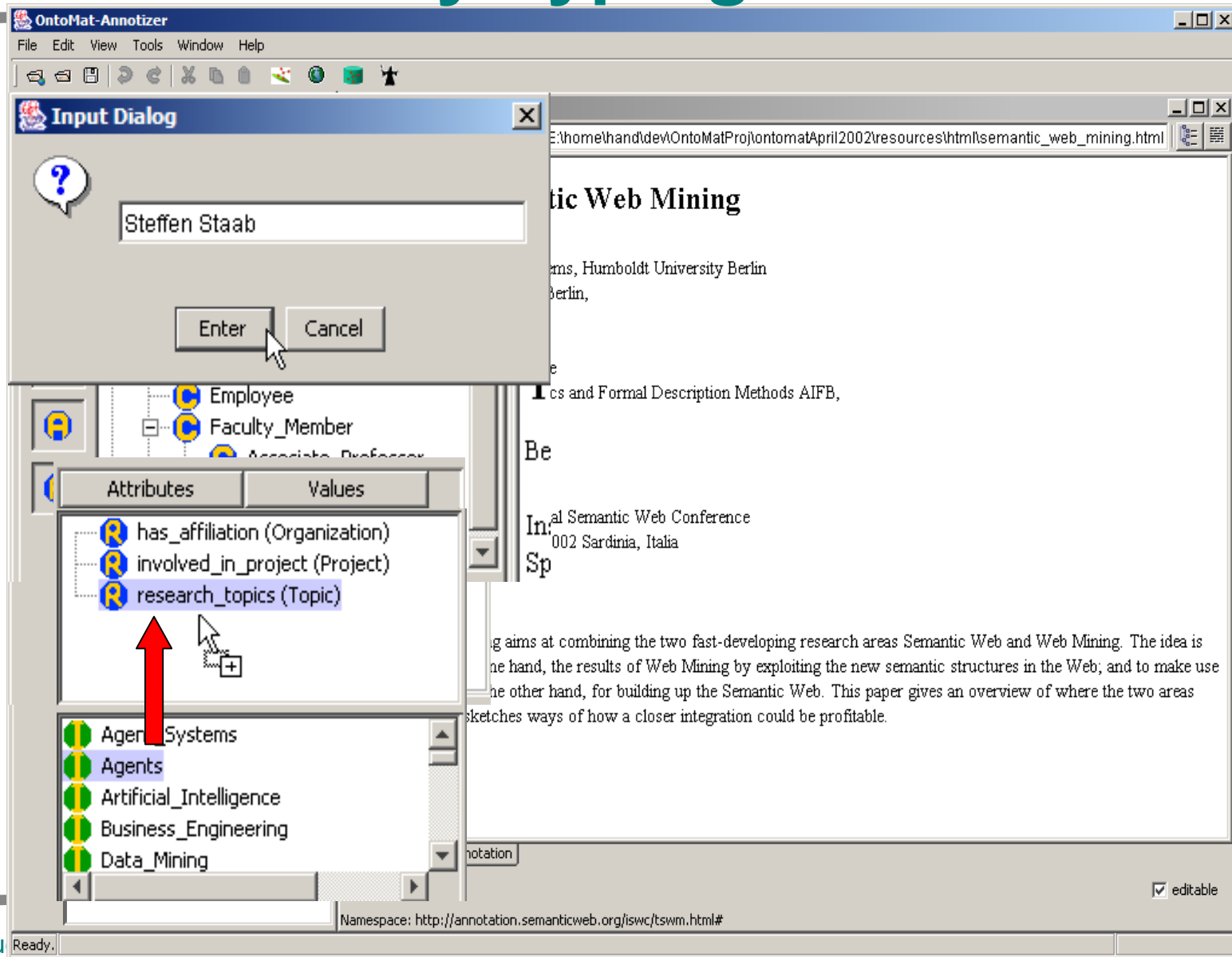
- Ontology Browser 1:** Displays a hierarchy of classes including Employee, Member, Associate_Professor, Full_Professor, Lecturer, Researcher, Student, and PhDStudent. A red arrow points from the 'Generate Class Instance' callout to the 'Employee' class.
- HTML Browser 1:** Displays a webpage titled 'Towards Semantic Web Mining' by Bettina Berendt, with contact information for the Institute of Information Systems at Humboldt University Berlin and the Institute of Applied Informatics at the University of Karlsruhe.
- Attribute Instance Table:** A table with columns 'Attribute' and 'Values'. The 'email' row contains 'sst@aifb.uni-karls...'. A red arrow points from the 'Attribute Instance' callout to this row.
- Relationship Instance:** A list of relationships including 'has_affiliation (Organization)', 'involved_in_project (Project)', and 'research_topics (Topic)'. A red arrow points from the 'Relationship Instance' callout to this list.

Generate Class Instance

Attribute Instance

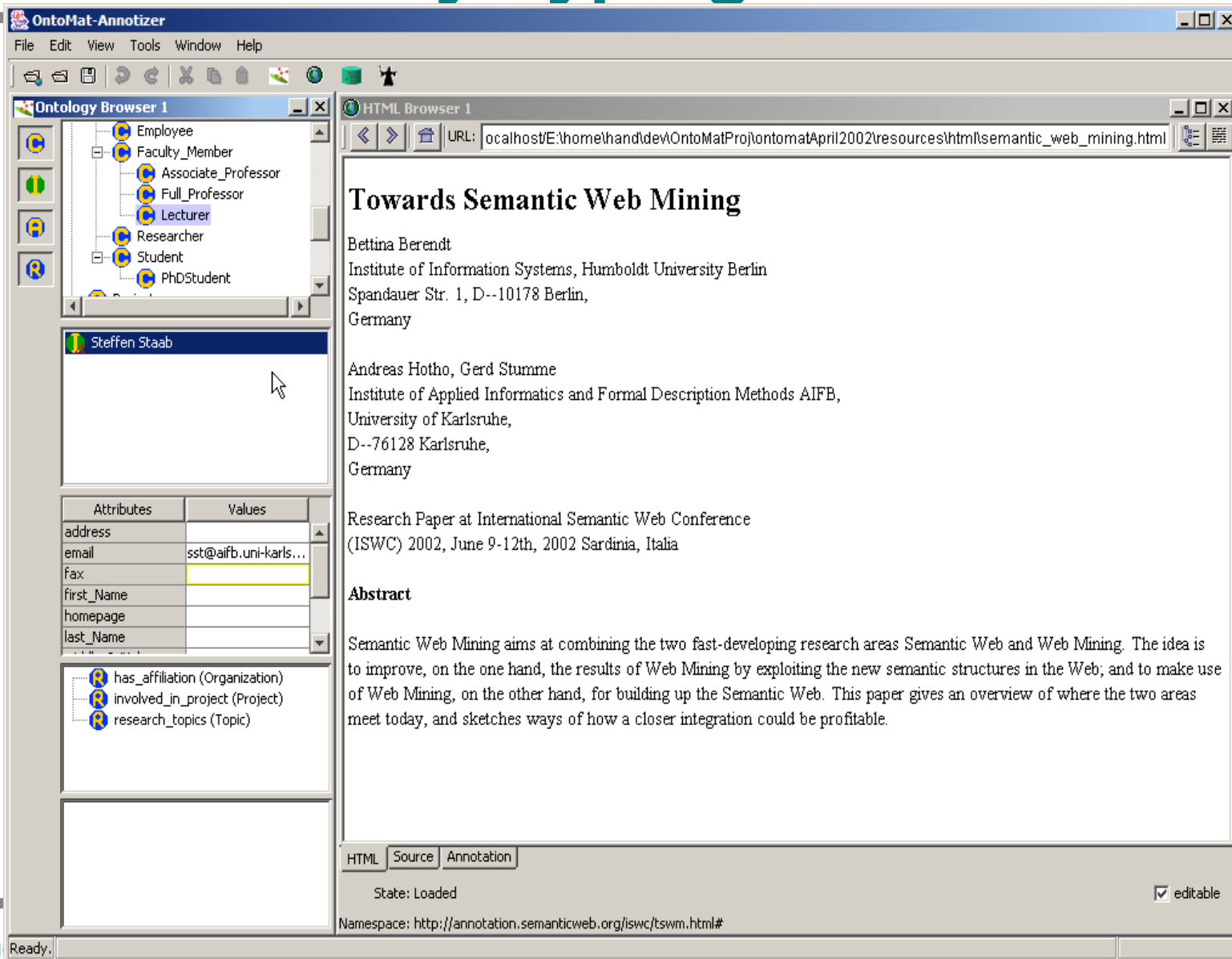
Relationship Instance

Annotation by Typing Statements



The screenshot shows the OntoMat-Annotizer application. The main window displays a web page titled "Semantic Web Mining" with text about combining Semantic Web and Web Mining. An "Input Dialog" is open over the page, containing a text field with "Steffen Staab" and "Enter" and "Cancel" buttons. Below the dialog, a list of ontology classes is shown: Employee, Faculty_Member, Associate_Professor, Agent_Systems, Agents, Artificial_Intelligence, Business_Engineering, and Data_Mining. A sub-list of properties is also visible: has_affiliation (Organization), involved_in_project (Project), and research_topics (Topic). A red arrow points to the "research_topics (Topic)" property. The status bar at the bottom shows "Ready." and "Namespace: http://annotation.semanticweb.org/iswc/tswm.html#".

Annotation by Typing Statements



Ontology Browser 1

- Employee
 - Faculty_Member
 - Associate_Professor
 - Full_Professor
 - Lecturer
 - Researcher
 - Student
 - PhDStudent

Steffen Staab

Attributes	Values
address	
email	sst@aifb.uni-karls...
fax	
first_Name	
homepage	
last_Name	

- has_affiliation (Organization)
- involved_in_project (Project)
- research_topics (Topic)

HTML Browser 1

URL: localhost/E:/home/handdev/OntoMatProj/ontomat/April2002/resources/html/semantic_web_mining.html

Towards Semantic Web Mining

Bettina Berendt
Institute of Information Systems, Humboldt University Berlin
Spandauer Str. 1, D--10178 Berlin,
Germany

Andreas Hotho, Gerd Stumme
Institute of Applied Informatics and Formal Description Methods AIFB,
University of Karlsruhe,
D--76128 Karlsruhe,
Germany

Research Paper at International Semantic Web Conference
(ISWC) 2002, June 9-12th, 2002 Sardinia, Italia

Abstract

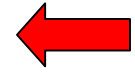
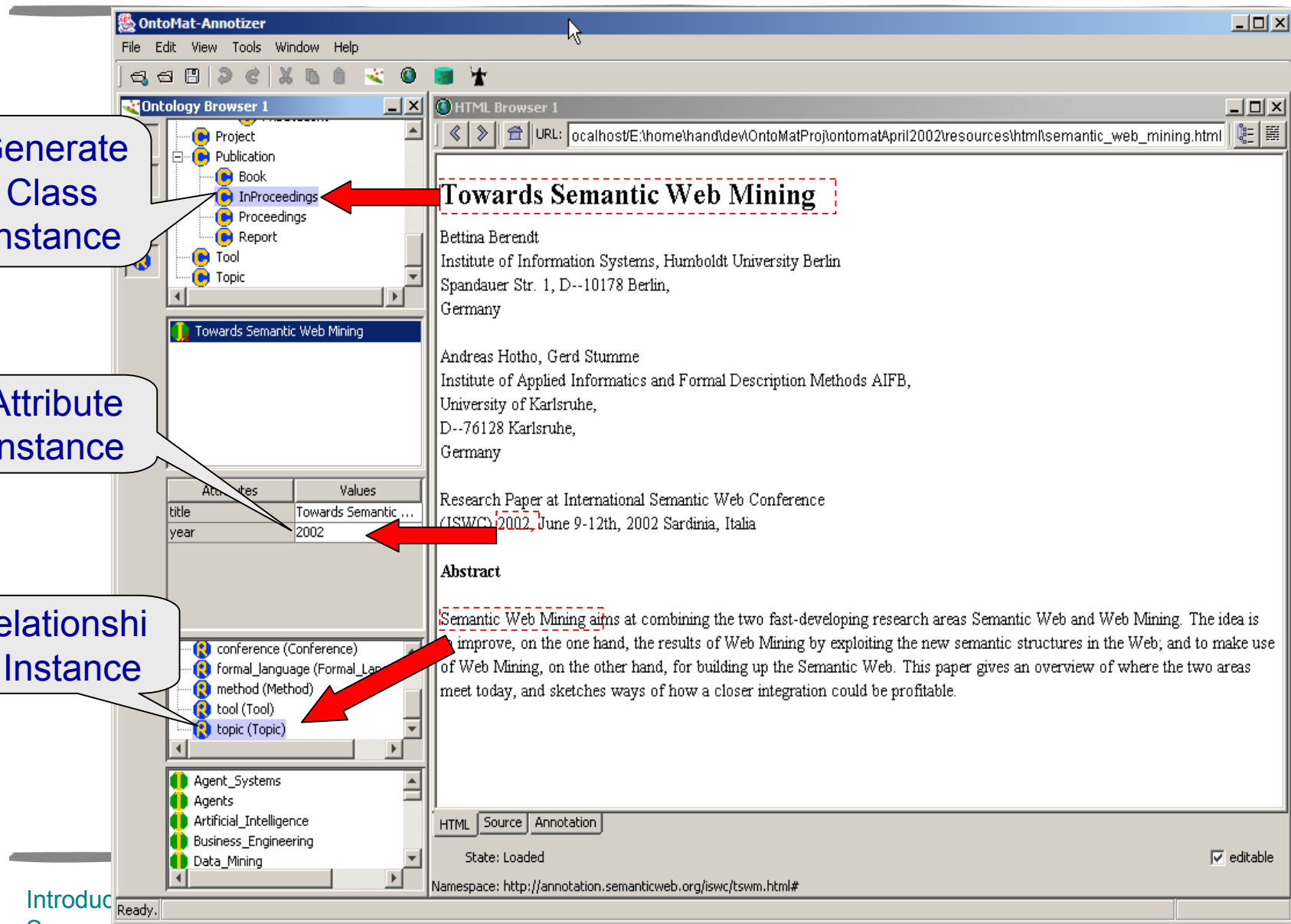
Semantic Web Mining aims at combining the two fast-developing research areas Semantic Web and Web Mining. The idea is to improve, on the one hand, the results of Web Mining by exploiting the new semantic structures in the Web; and to make use of Web Mining, on the other hand, for building up the Semantic Web. This paper gives an overview of where the two areas meet today, and sketches ways of how a closer integration could be profitable.

HTML | Source | Annotation

State: Loaded editable

Namespace: http://annotation.semanticweb.org/iswc/tswm.html#

Annotation by Markup

The screenshot shows the OntoMat-Annotizer application with three main panes:

- Ontology Browser 1:** A tree view of an ontology. The 'InProceedings' class is highlighted, with a red arrow pointing to it from the 'Generate Class Instance' callout.
- HTML Browser 1:** Displays the rendered HTML of a document titled 'Towards Semantic Web Mining'. The title is enclosed in a red dashed box, with a red arrow pointing to it from the 'Generate Class Instance' callout. Below the title, the authors and their affiliations are listed. The abstract text is also enclosed in a red dashed box, with a red arrow pointing to it from the 'Relationship Instance' callout.
- Attributes Table:** A table with two columns: 'Attributes' and 'Values'. It contains the following data:

Attributes	Values
title	Towards Semantic ...
year	2002

 A red arrow points from the 'year' value to the 'Attribute Instance' callout.

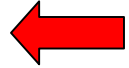
At the bottom of the interface, there are tabs for 'HTML', 'Source', and 'Annotation'. The 'HTML' tab is active. The status bar shows 'State: Loaded' and a checked 'editable' checkbox. The namespace is defined as 'http://annotation.semanticweb.org/iswc/tswm.html#'. A red arrow points from the top right of the image towards the application window.

Generate
Class
Instance

Attribute
Instance

Relationship
Instance

Markup Class Instances



OntoMat-Annotizer

File Edit View Tools Window Help

Ontology Browser 1

- Organization
- Person
- Project
- Publication
 - Book
 - InProceedings
 - Proceedings
 - Report
- Tool
- Topic

Towards Semantic Web Mining

Attributes	Values
title	
year	

algorithm (Algorithm)

application (Application)

application_domain (Application_...

author (Person)

conference (Conference)

HTML Browser

URL: localhost/E:/home/hand/dev/OntoMatProj/ontomat/April2002/resources/html/semantic_web_mining.html

Towards Semantic Web Mining

Bettina Berendt

Institute of Information Systems, Humboldt University Berlin
Spandauer Str. 1, D--10178 Berlin,
Germany
<http://www.wiwi.hu-berlin.de/~berendt>
berendt@wiwi.hu-berlin.de

Andreas Hotho, Gerd Stumme
Institute of Applied Informatics and Formal Description Methods AIFB,
University of Karlsruhe,
D--76128 Karlsruhe,
Germany
<http://www.aifb.uni-karlsruhe.de/WBS>
([hotho](mailto:hotho@aifb.uni-karlsruhe.de), [stumme](mailto:stumme@aifb.uni-karlsruhe.de))@aifb.uni-karlsruhe.de

Abstract

Semantic Web Mining aims at combining the two fast-developing research areas Semantic Web and Web Mining. The idea is to improve, on the one hand, the results of Web Mining by exploiting the new semantic structures in the Web; and to make use of Web Mining, on the other hand, for building up the Semantic Web. This paper gives an overview of where the two areas meet today, and sketches ways of how a closer integration could be profitable.

Research Paper at International Semantic Web Conference (ISWC) 2002, June 9-12th, 2002 Sardinia, Italia

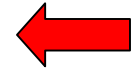
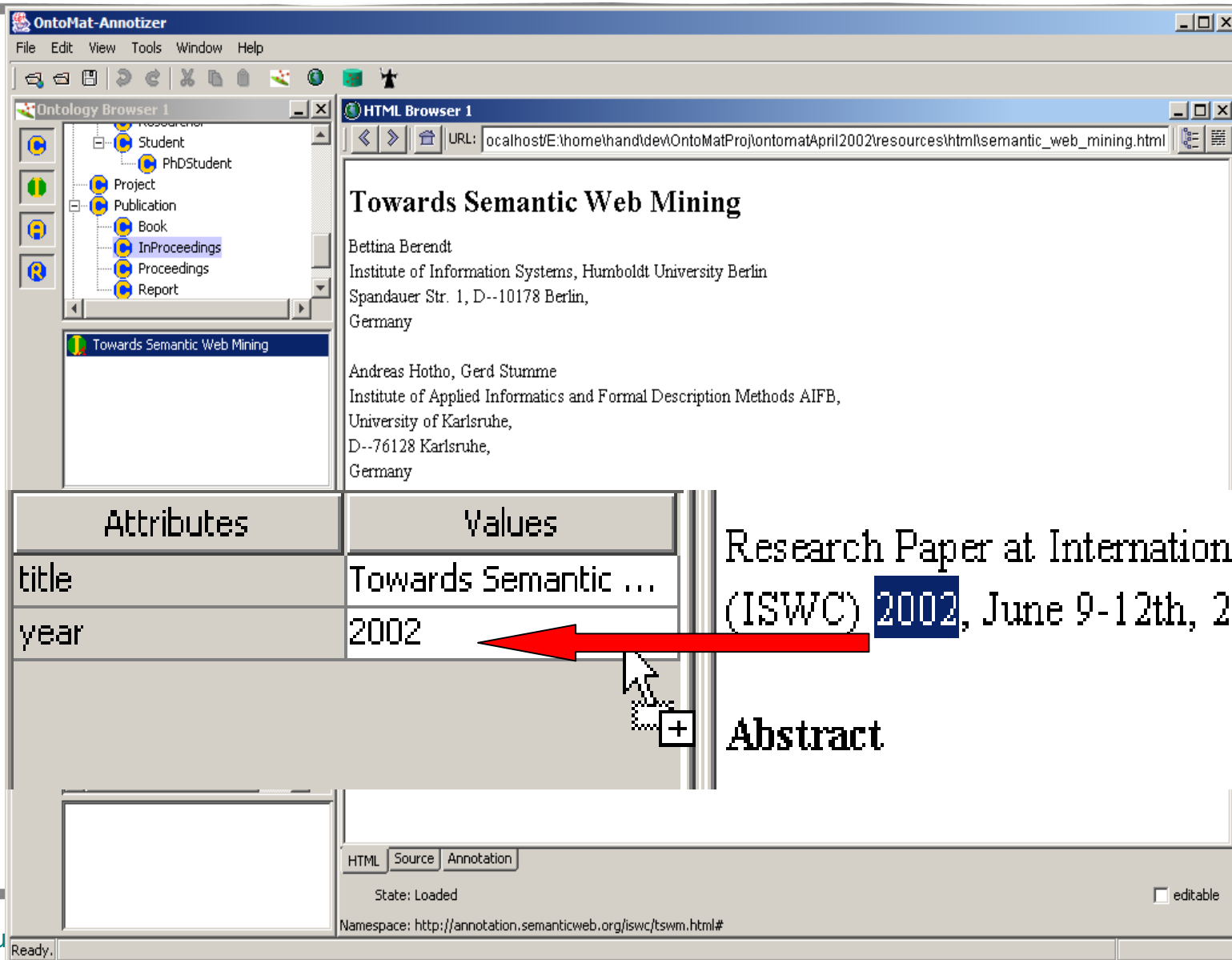
HTML Source Annotation

State: Loaded editable

Namespace: http://annotation.semanticweb.org/iswc/Towards_Semantic_Web_Mining.html#

Ready.

Markup Attribute Instances

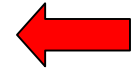
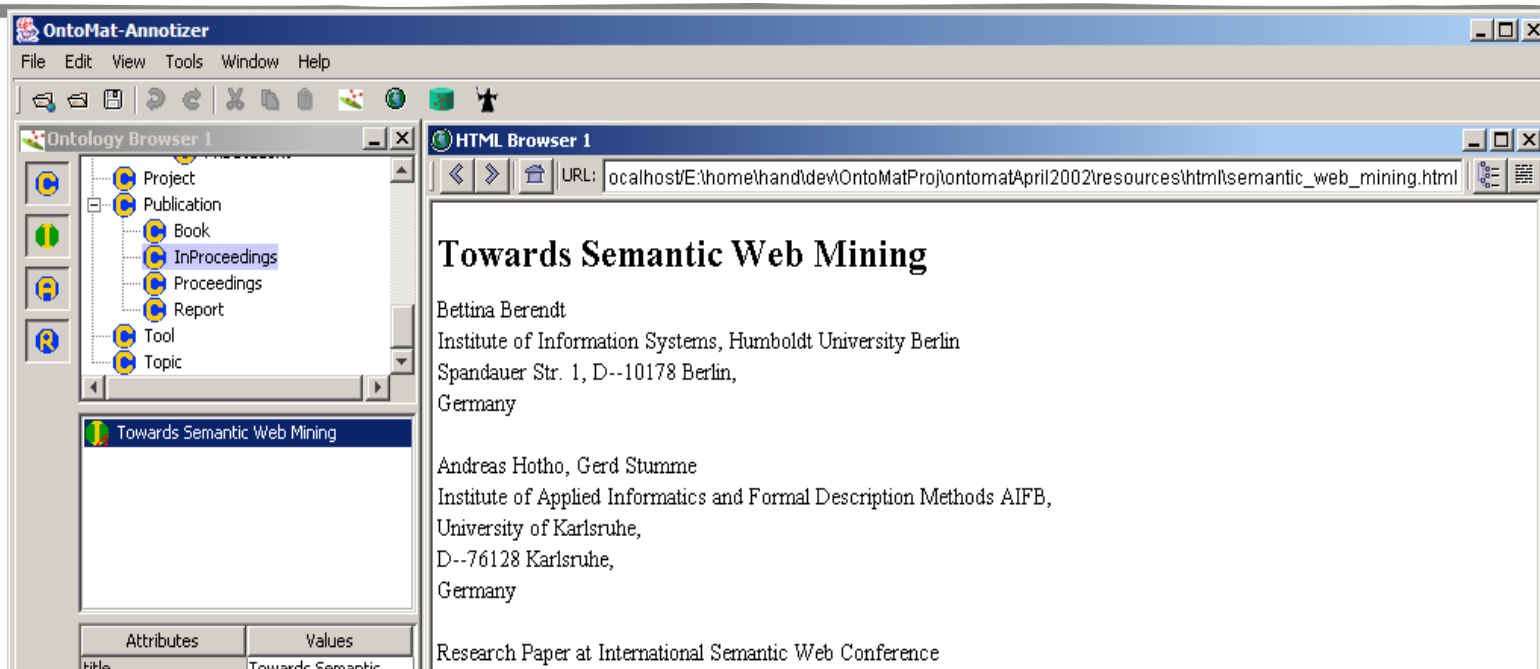
The screenshot shows the OntoMat-Annotizer application. On the left is the 'Ontology Browser 1' showing a tree structure with categories like Student, Project, Publication, Book, InProceedings, Proceedings, and Report. The 'Towards Semantic Web Mining' document is selected. On the right is the 'HTML Browser 1' displaying the document content, including the title 'Towards Semantic Web Mining' and authors 'Bettina Berendt' and 'Andreas Hotho, Gerd Stumme'. Below the HTML browser is a table of attribute instances:

Attributes	Values
title	Towards Semantic ...
year	2002

To the right of the table, the text 'Research Paper at International (ISWC) 2002, June 9-12th, 200' is displayed, with '2002' highlighted in blue. Below this text is a '+' icon and the word 'Abstract'. A red arrow points from the '2002' value in the table to the '2002' in the text. Another red arrow points from the top right of the slide towards the main title.

At the bottom of the application window, there are tabs for 'HTML', 'Source', and 'Annotation'. The status bar shows 'State: Loaded', 'Namespace: http://annotation.semanticweb.org/iswc/tswm.html#', and an 'editable' checkbox.

Markup Relationship Instances

OntoMat-Annotizer

File Edit View Tools Window Help

Ontology Browser 1

- Project
 - Publication
 - Book
 - InProceedings
 - Proceedings
 - Report
 - Tool
 - Topic

Towards Semantic Web Mining

Attributes	Values
file	Towards Semantic

HTML Browser 1

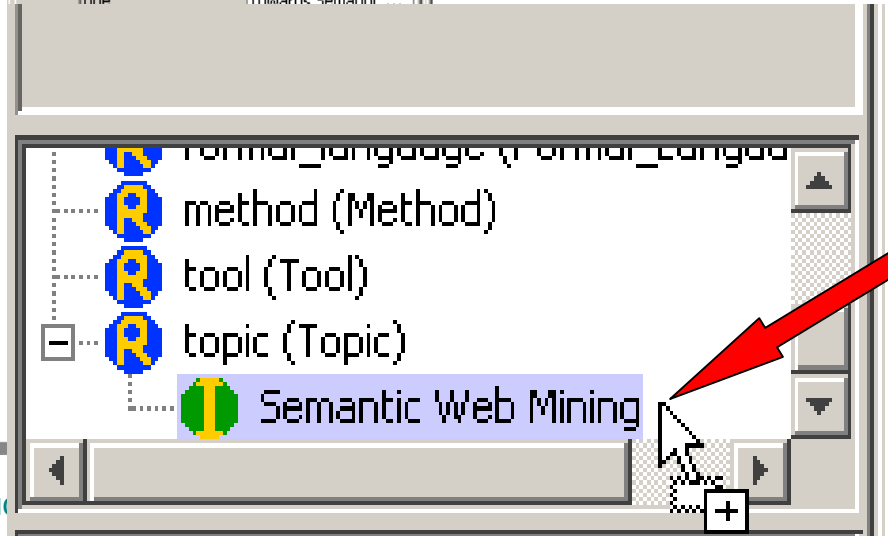
URL: localhost/E:/home/hand/dev/OntoMatProj/ontomat/April2002/resources/html/semantic_web_mining.html

Towards Semantic Web Mining

Bettina Berendt
 Institute of Information Systems, Humboldt University Berlin
 Spandauer Str. 1, D--10178 Berlin,
 Germany

Andreas Hotho, Gerd Stumme
 Institute of Applied Informatics and Formal Description Methods AIFB,
 University of Karlsruhe,
 D--76128 Karlsruhe,
 Germany

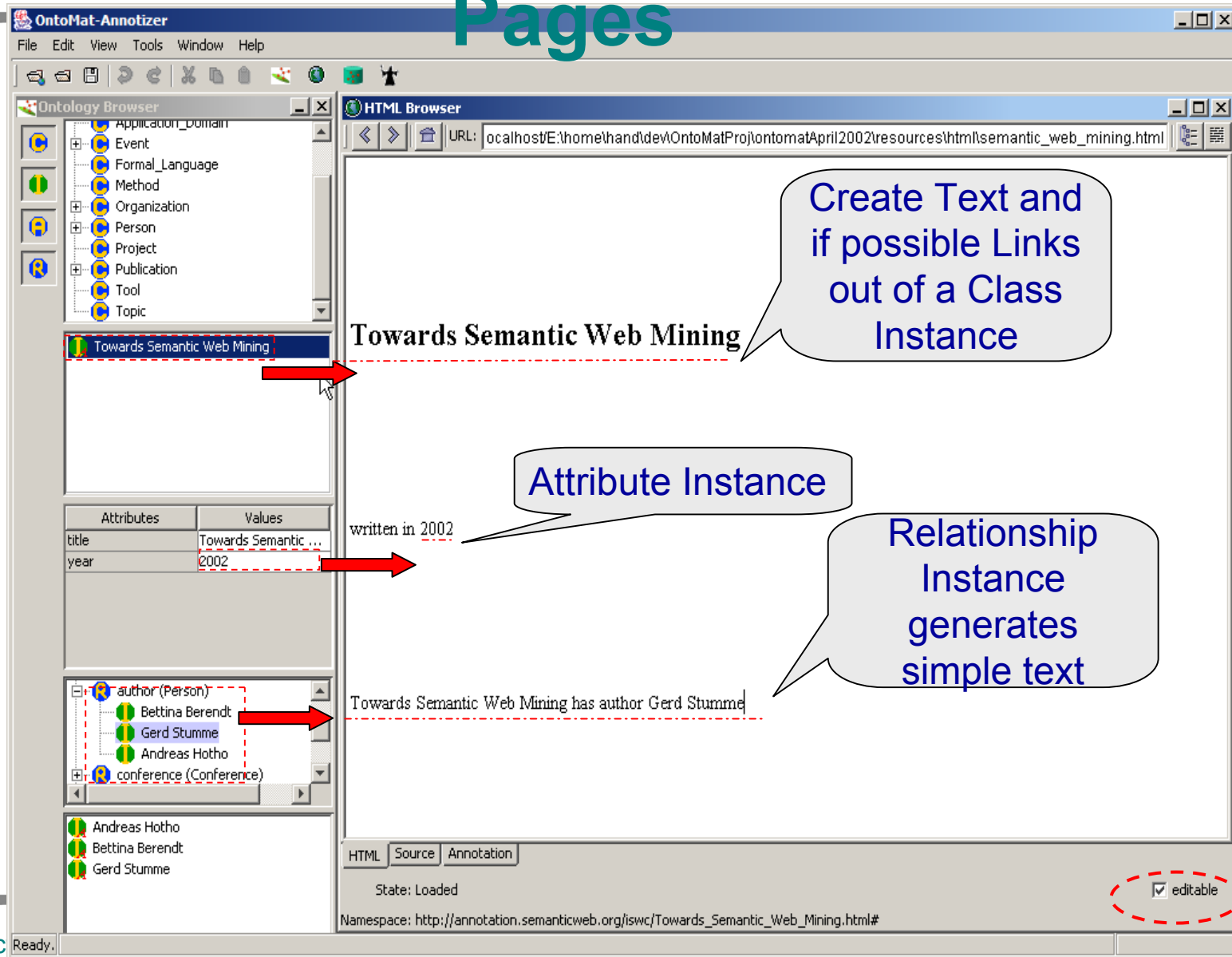
Research Paper at International Semantic Web Conference



- formal_language (Formal_Language)
- method (Method)
- tool (Tool)
- topic (Topic)
- Semantic Web Mining

Semantic Web Mining aims at to improve, on the one hand, the state of Web Mining, on the other hand to meet today, and sketches way:

Annotation by Authoring Web Pages

The screenshot shows the OntoMat-Annotizer application with the following components:

- Ontology Browser:** A tree view on the left showing classes like Application_Domain, Event, Formal_Language, Method, Organization, Person, Project, Publication, Tool, and Topic. The class "Towards Semantic Web Mining" is selected.
- HTML Browser:** A window showing the URL "localhost/E:/home/hand/dev/OntoMatProj/ontomat/April2002/resources/html/semantic_web_mining.html".
- Attributes Table:** A table with columns "Attributes" and "Values".

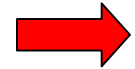
Attributes	Values
title	Towards Semantic ...
year	2002
- Instance Lists:** Below the table, there are lists for "author (Person)" and "conference (Conference)". The "author (Person)" list includes Bettina Berendt, Gerd Stumme, and Andreas Hotho.
- Annotation Area:** The main text area contains the text "Towards Semantic Web Mining" and "Towards Semantic Web Mining has author Gerd Stumme".
- Bottom Panel:** Includes tabs for "HTML", "Source", and "Annotation", a "State: Loaded" indicator, and a "Namespace" field. A red dashed circle highlights the "editable" checkbox.

Create Text and if possible Links out of a Class Instance

Attribute Instance

Relationship Instance generates simple text

Authoring Class Instances

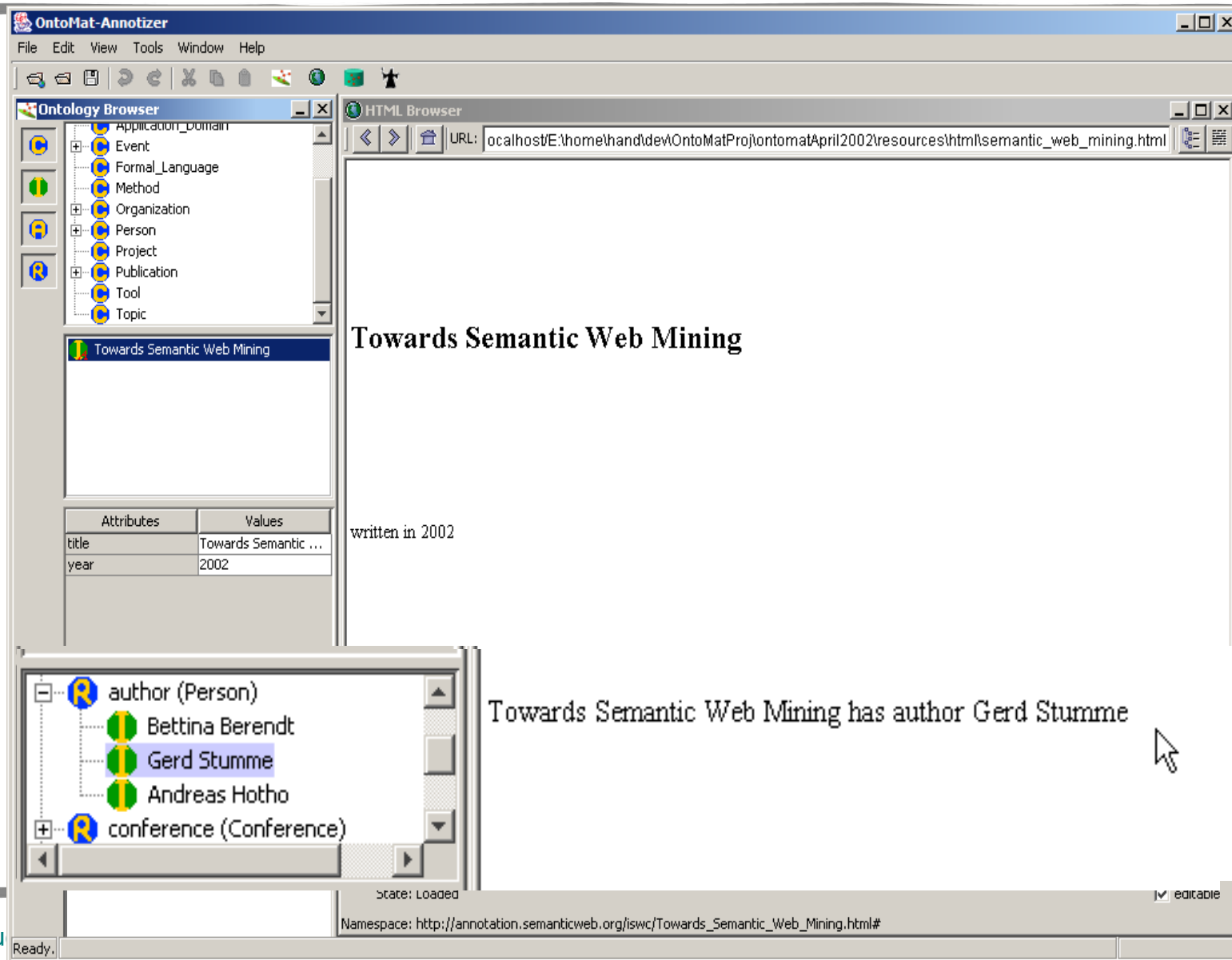


Towards Semantic Web Mining

Authoring Attribute Instances



Authoring Relationship Instances

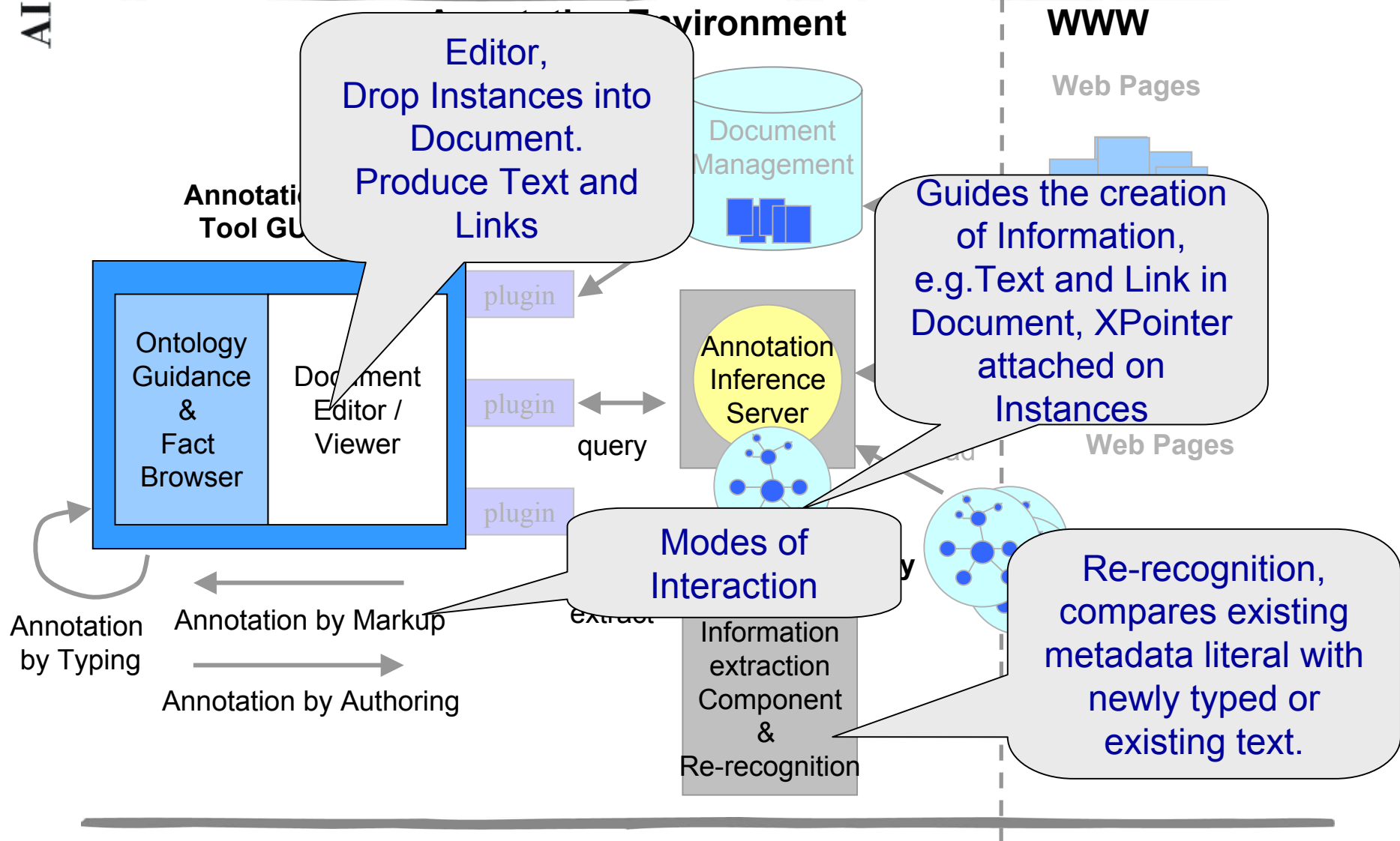


The screenshot shows the OntoMat-Annotizer application with the following components:

- Ontology Browser:** A tree view on the left showing a hierarchy of classes: Application_Domain, Event, Formal_Language, Method, Organization, Person, Project, Publication, Tool, and Topic. The 'Towards Semantic Web Mining' class is selected.
- HTML Browser:** A window on the right displaying the content of a web page. The URL is `localhost/E:\home\hand\dev\OntoMatProj\ontomat\April2002\resources\html\semantic_web_mining.html`. The page content includes the title "Towards Semantic Web Mining" and the text "written in 2002".
- Attributes Table:** A table below the ontology browser showing the attributes of the selected class:

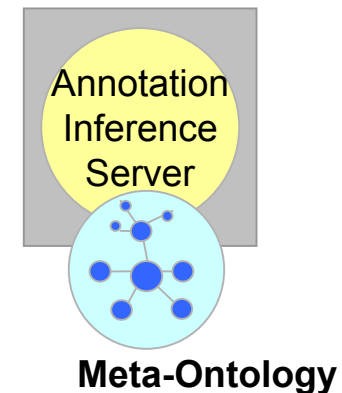
Attributes	Values
title	Towards Semantic ...
year	2002
- Relationship Instances:** A window at the bottom shows a list of instances:
 - author (Person):** A list of names: Bettina Berendt, Gerd Stumme (highlighted), and Andreas Hotho.
 - conference (Conference):** A list of conference names.
- Status Bar:** At the bottom, it shows "Ready.", "Namespace: http://annotation.semanticweb.org/iswc/Towards_Semantic_Web_Mining.html#", and a checked "editable" checkbox.

Design of CREAM

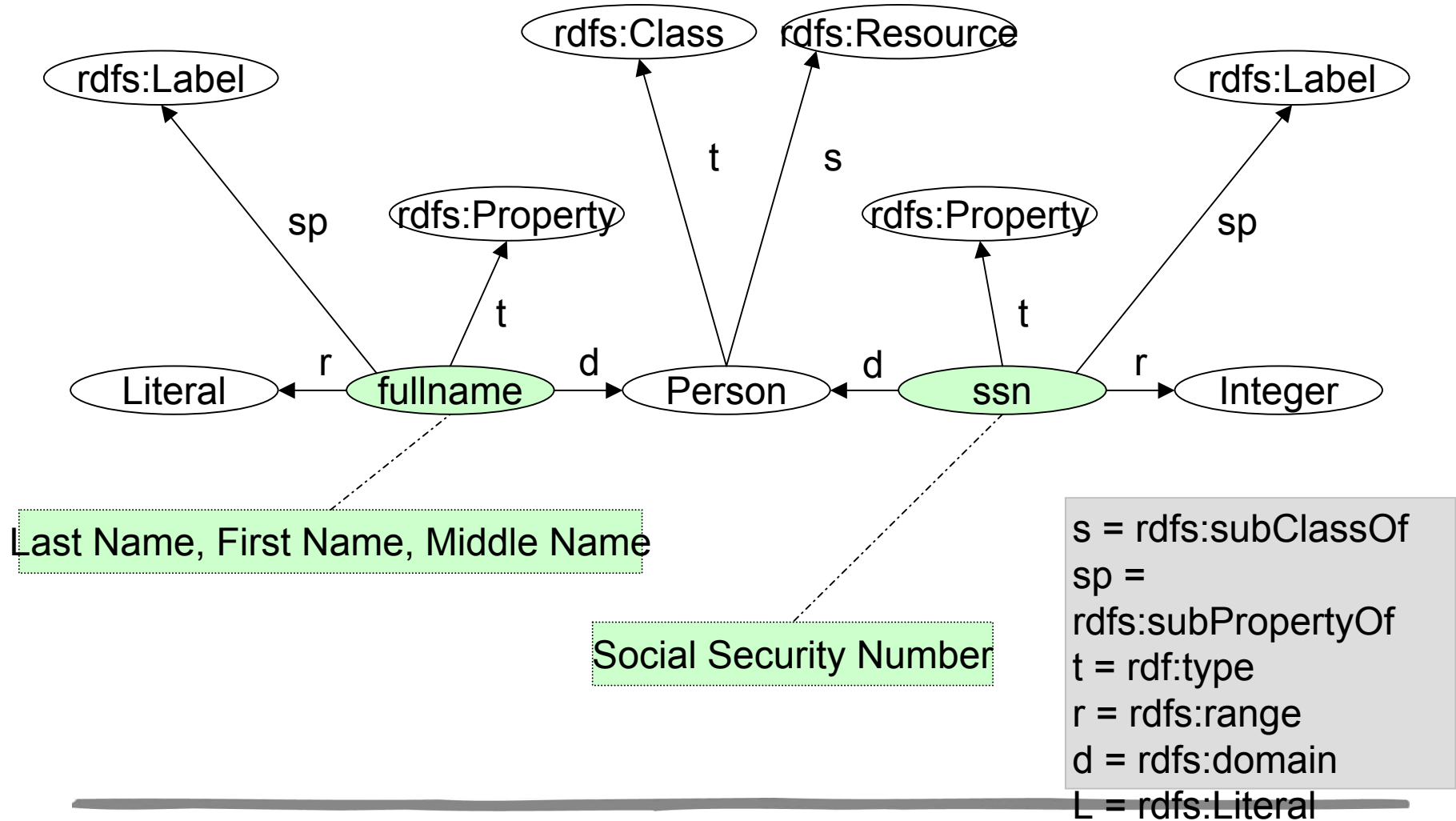


Meta Ontology

- **Modularization** of ontology development and use
- Describes how **classes**, **attributes** and **relationships** from the ontology should be used by the CREAM environment.
- The Meta Ontology supports the **modes of interaction**
- Meta Ontology **characterizations**:
 - Label
 - Default Pointing
 - Property Mode



Meta Ontology - Label



Meta Ontology - Label

- Labels are used at (at least) two points of interaction
 - Instance Generation
 - RDF-API for a new URN as ID
 - Assign piece of Text with Attribute recorded as rdfs:label, e.g. fullname or ssn
 - Content Generation
 - Text is produced by rdfs:label, e.g. fullname or ssn

→ The connection is **not objective**.

→ Their linkage depends on usage in a particular **scenario**.

Meta Ontology – Default Pointing

- Specify the default pointing behavior for class instances
- **Exploiting** the XPointer candidate recommendation:
CREAM:UniqueDPointer, CREAM:AutoDPointer and
CREAM:AutoUniqueDPointer
- **Instance-Generation** (Annotation by Markup):
 - CREAM:AutoDPointer or CREAM:AutoUniqueDPointer
- **Content-Generation** (Annotation by Authoring):
 - CREAM:UniqueDPointer or CREAM:AutoUniqueDPointer
- **Example:**
 - Person with properties
hasHomepage (CREAM:AutoUniqueDPointer) and fullname
(Label)

Meta Ontology – Property Mode

- Distinguishes between different **roles**
- **Reference:**
e.g. refer to the current U.S. president at
<http://www.whitehouse.gov>
- **Quotation:**
e.g. “Bill Clinton” as president of U.S. in 1999 at
<http://www.whitehouse.gov>
- **Unlinked Fact:**
e.g. a fact-attributes may be filled with “Spanish Civil War” for the reference pointing to the picture “Guernica”,
<http://www.grnica.swinternet.co.uk/guerni/ca.jpg>.

Meta-Ontology & Modes of Interaction



1. Annotation by Typing Statements

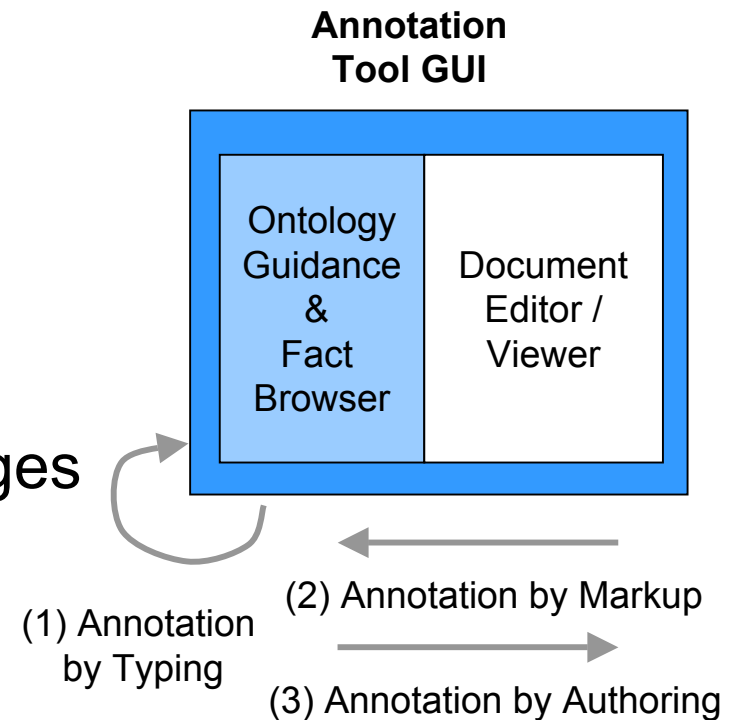
- rdfs:label
- default pointer: [AutoDPointer](#)
- property mode: [unlinked fact](#)

2. Annotation by Markup

- rdfs:label
- default pointer: [AutoDPointer](#)
- property mode: [reference](#), [quotation](#)

3. Annotation by Authoring Web Pages

- rdfs:label
- default pointer: [UniqueDPointer](#)
- property mode: [reference](#), [quotation](#), [unlinked fact](#)



Outlook – Future Topics



- Further Elaboration of
 - Template Generation
 - Provision of Metametadata (Author of Annotation)
- Annotation by Authoring – Creation of lists or tables from selected concepts
- Annotation by Authoring with the help of metadata re-recognition (needs information extraction)
- Creation of Relational Metadata in **PDF**, **SVG** and **SMIL**

Summary

- CREAM Framework supports metadata creation during Web page **authoring** as well as by **a-posteriori** annotation
 - Meta Ontology
 - Modes of Interaction
 - Reference implementation: **OntoMat-*Annotizer***
OntoMat + annotation plugins
 - A-Posteriori Version currently used for the ISWC annotation
 - Make the Semantic Web vision realistic!
<http://annotation.semanticweb.org/ontomat>
-