



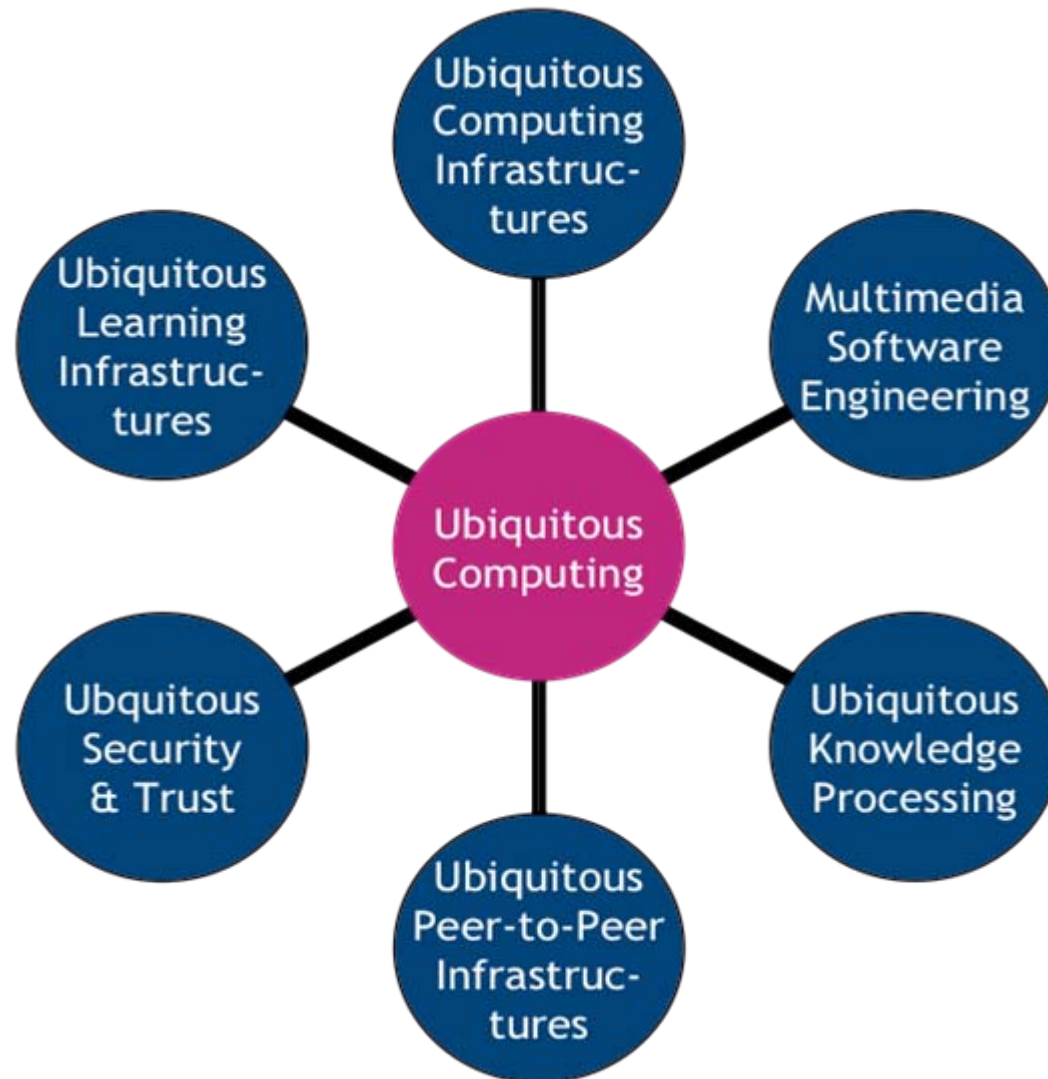
Darmstadt Knowledge Processing Repository Based on UIMA

Iryna Gurevych, Max Mühlhäuser, Christof Müller,
Jürgen Steimle, Markus Weimer, Torsten Zesch

Ubiquitous Knowledge Processing Group
Telecooperation, Computer Science Department
Darmstadt University of Technology

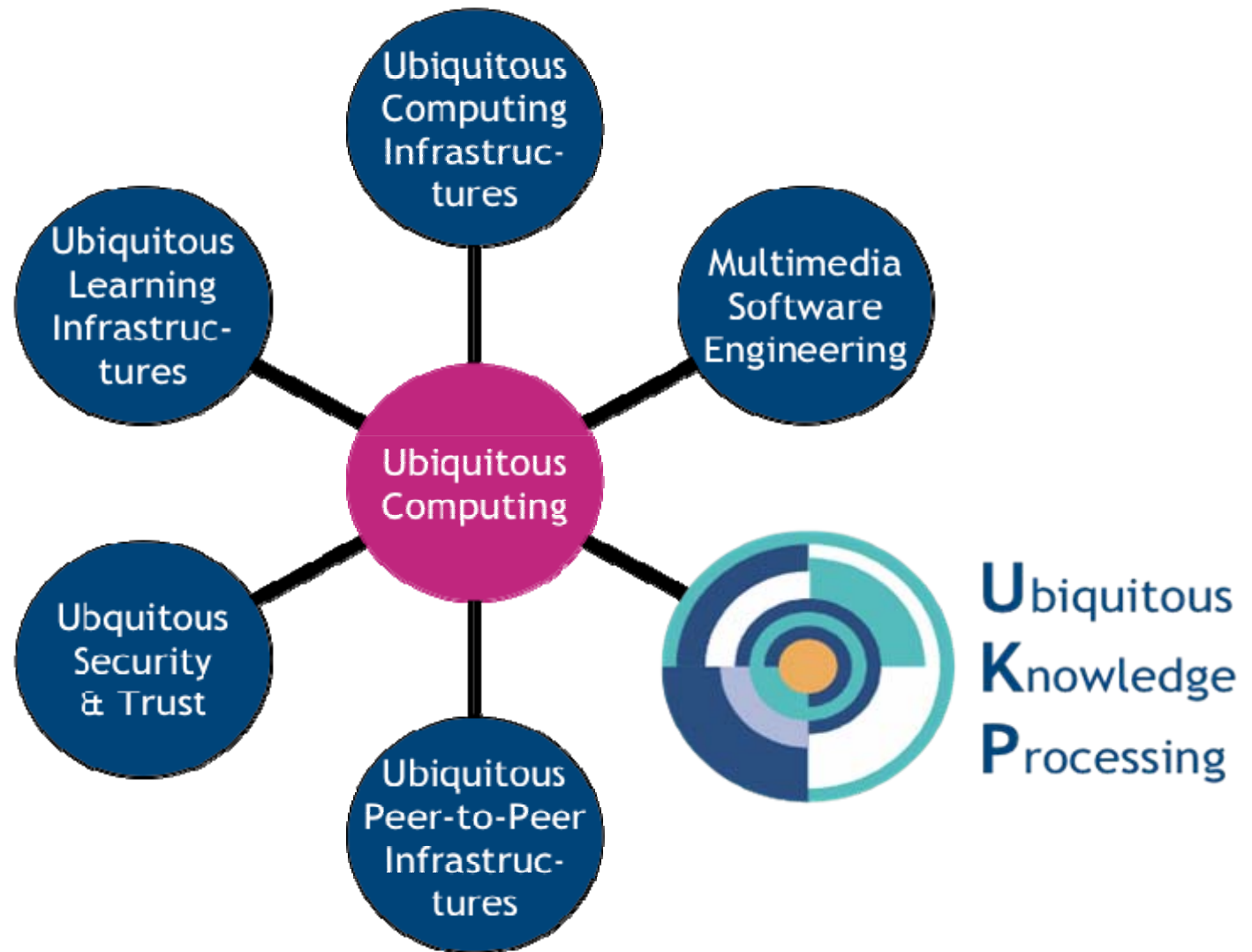


Telecooperation





Telecooperation



Ubiquitous Knowledge Processing



SIR

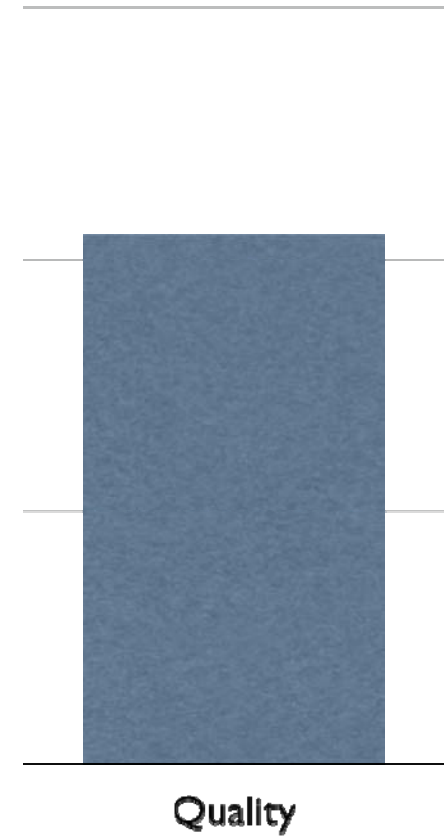
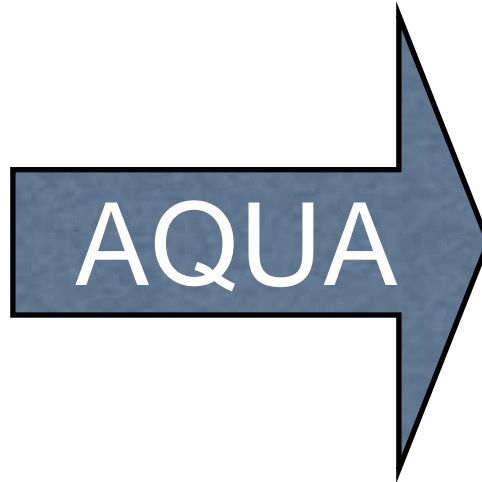
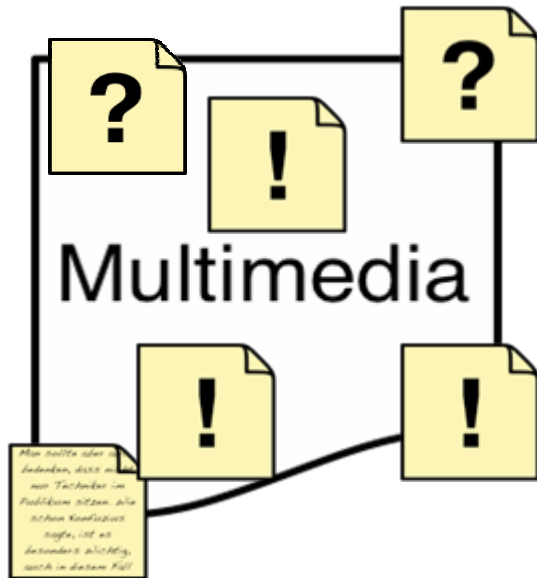
AQUA

THESEUS

Darmstadt Knowledge Processing Software
Repository



Automatic **Quality Assessment** and Feedback in eLearning 2.0 (AQUA)





User Generated Discourse in Web 2.0

The collage consists of several overlapping browser windows:

- YouTube:** Shows a video player for "Four Will Be Discovered" with a list of featured videos and a search bar.
- del.icio.us:** A bookmarking service interface with a search bar and a list of saved bookmarks.
- Flickr:** A photo sharing platform interface with a search bar and a list of photos.
- Wikipedia:** The main page of the German Wikipedia, featuring a search bar and a list of featured articles.
- Google Maps:** A map of Germany with a search bar and various map controls.
- Other sites:** Includes a mailing list page, a forum page, and a page for a seminar.



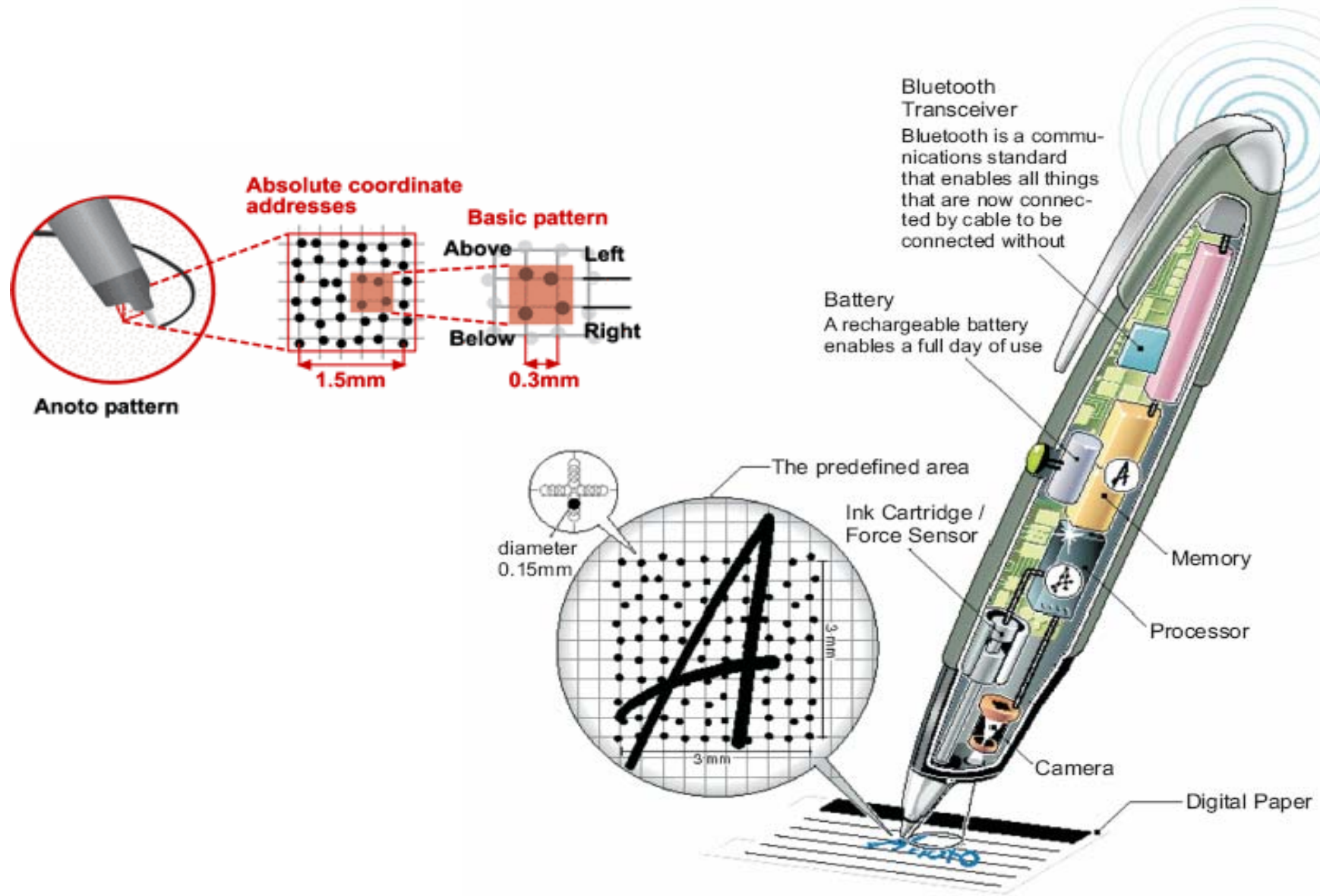
Moose is an interactive environment for exploring places based on the photos people take.

The screenshot shows the German Wikipedia main page with a search bar, a list of featured articles, and a sidebar with navigation options.

The screenshot shows an Amazon product page for "The Da Vinci Code (Widescreen Two-Disc Special Edition) (2006)". It includes a product image, price information, and a list of related products.

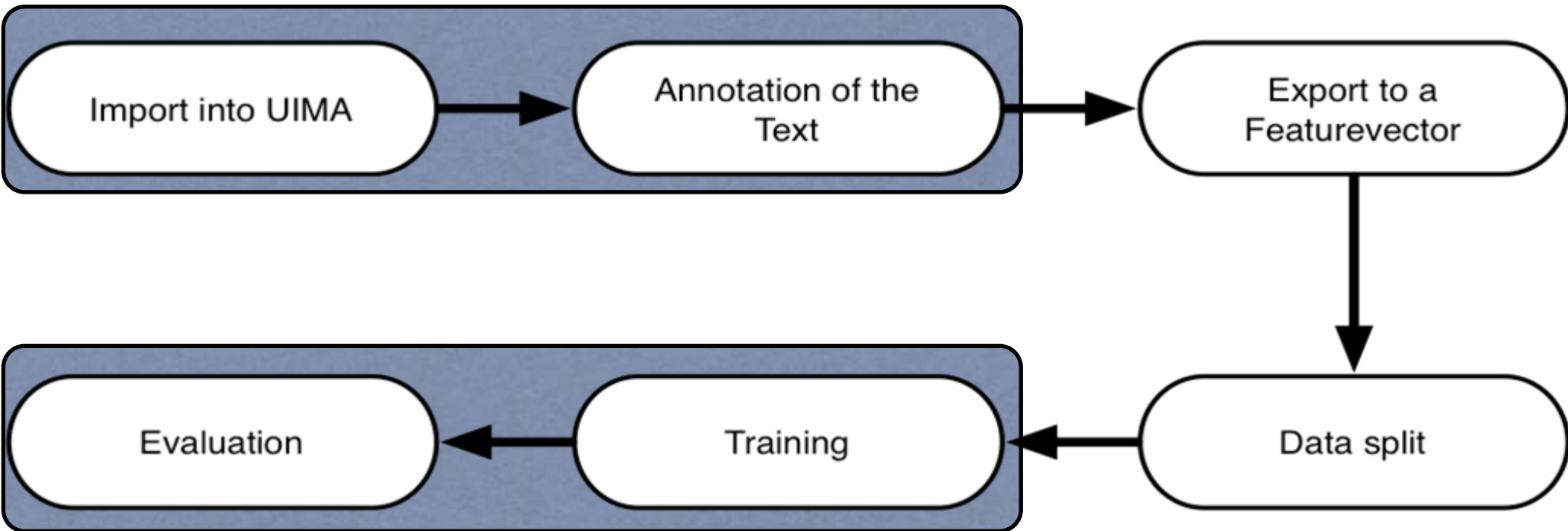


AQUA – Anoto pen



AQUA - System Architecture

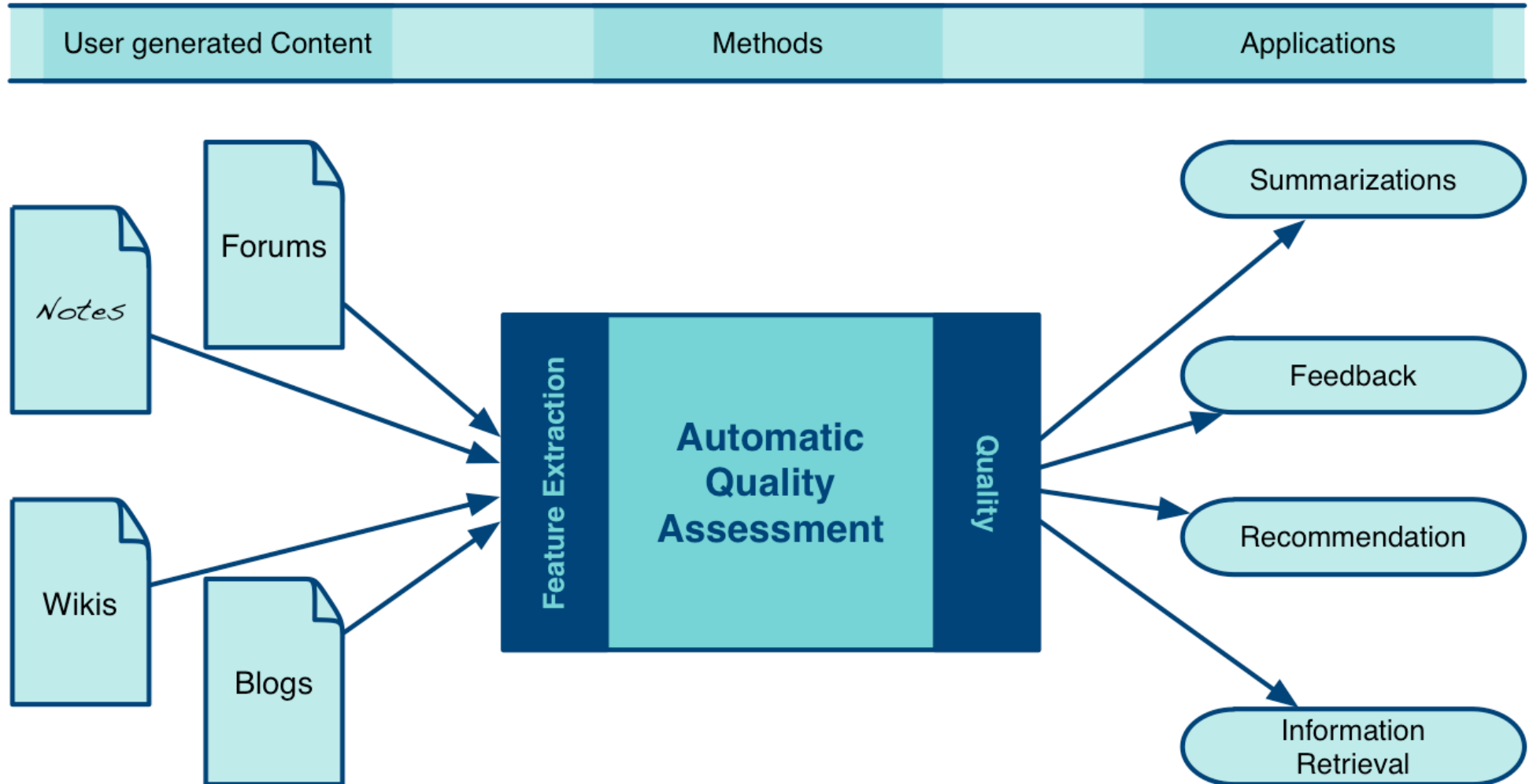
Natural Language Processing



Machine Learning

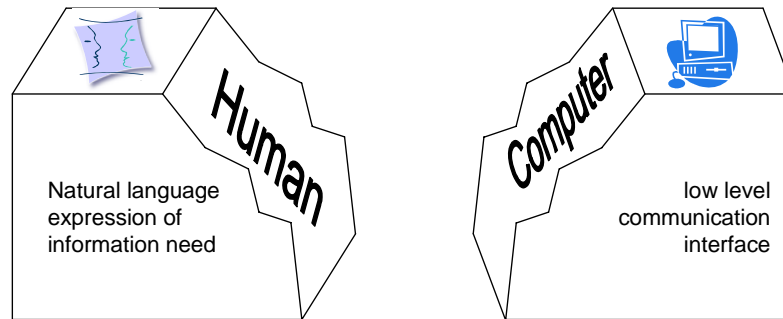


AQUA – System Architecture

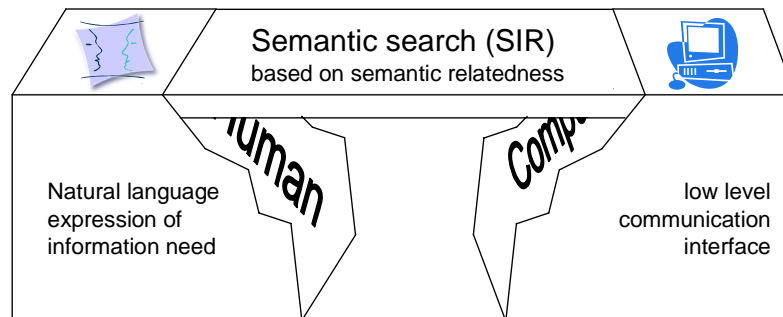
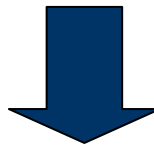


SIR (in cooperation with Prof. Hinrichs)

- **S**emantic **I**nformation **R**etrieval



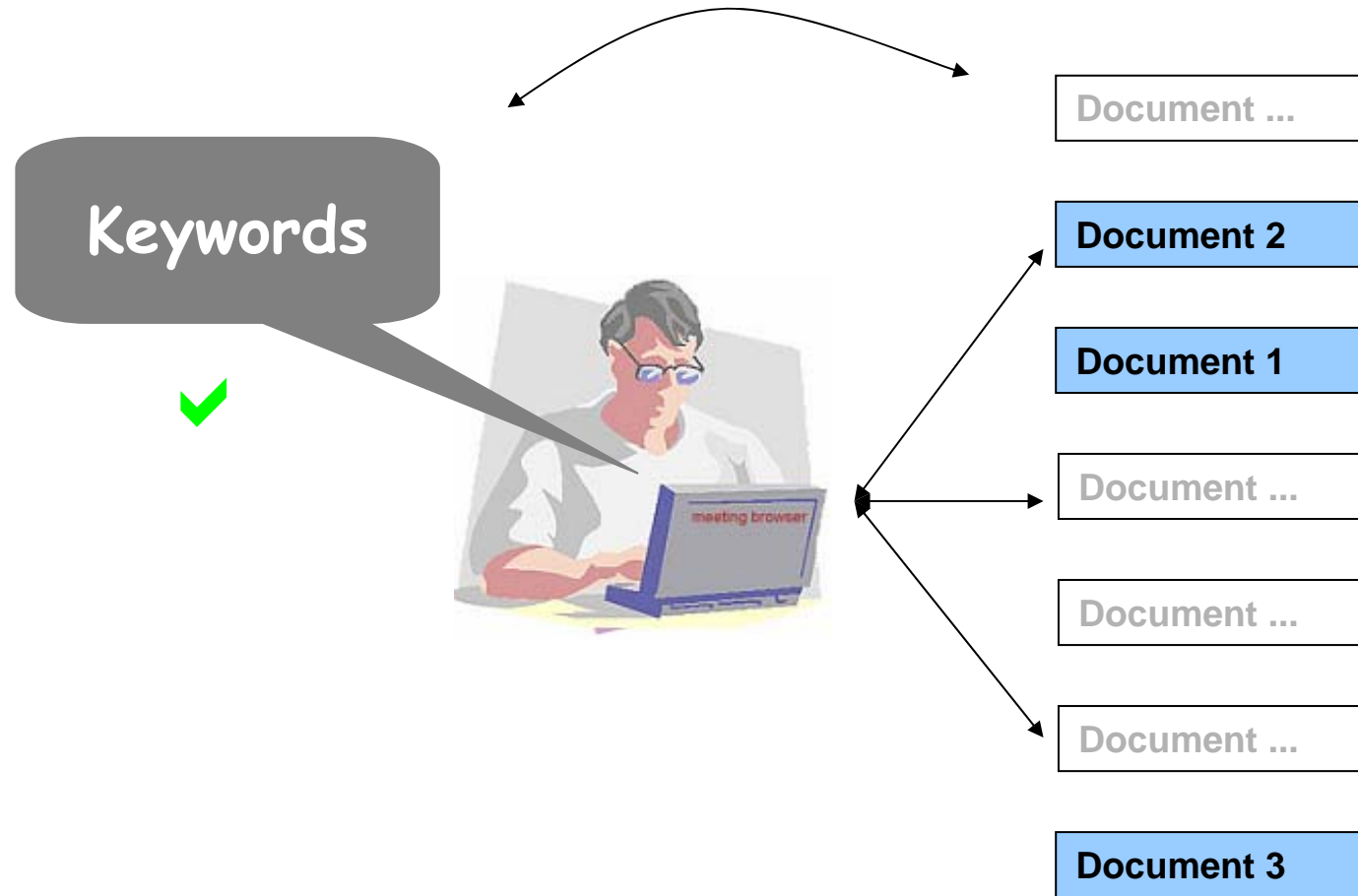
Bridge the human – computer gap





Information Retrieval (IR)

Boolean, Vector Space, ...





SIR-Project

Semantic Relatedness

baker, to program,
quality assurance

Essay

cake,
computer,
to read, ...



Profession ...

Profession 2

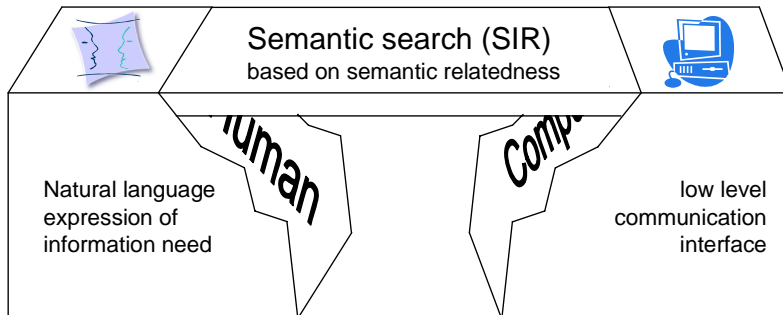
Profession 1

Profession ...

Profession ...

Profession ...

Profession 3





SIR Example

find good
index terms

Enter your essay here

Ich würde gerne mit Tieren arbeiten, sie behandeln, für sie sorgen, aber ich kann kein Blut sehen und ich habe zu viel Mitleid mit kranken Tieren.
Andererseits arbeite ich besonders gerne am Computer, kann programmieren in C, Python und VB und könnte mir daher auch in der Software-Entwicklung einen passenden Beruf vorstellen.

Load

Nouns

Semantic relatedness

Execute

IR Query

Compound Splitting
Negation Detection
WSD

IR Result

Save Results

Display Explanations

compute
semantic
relatedness

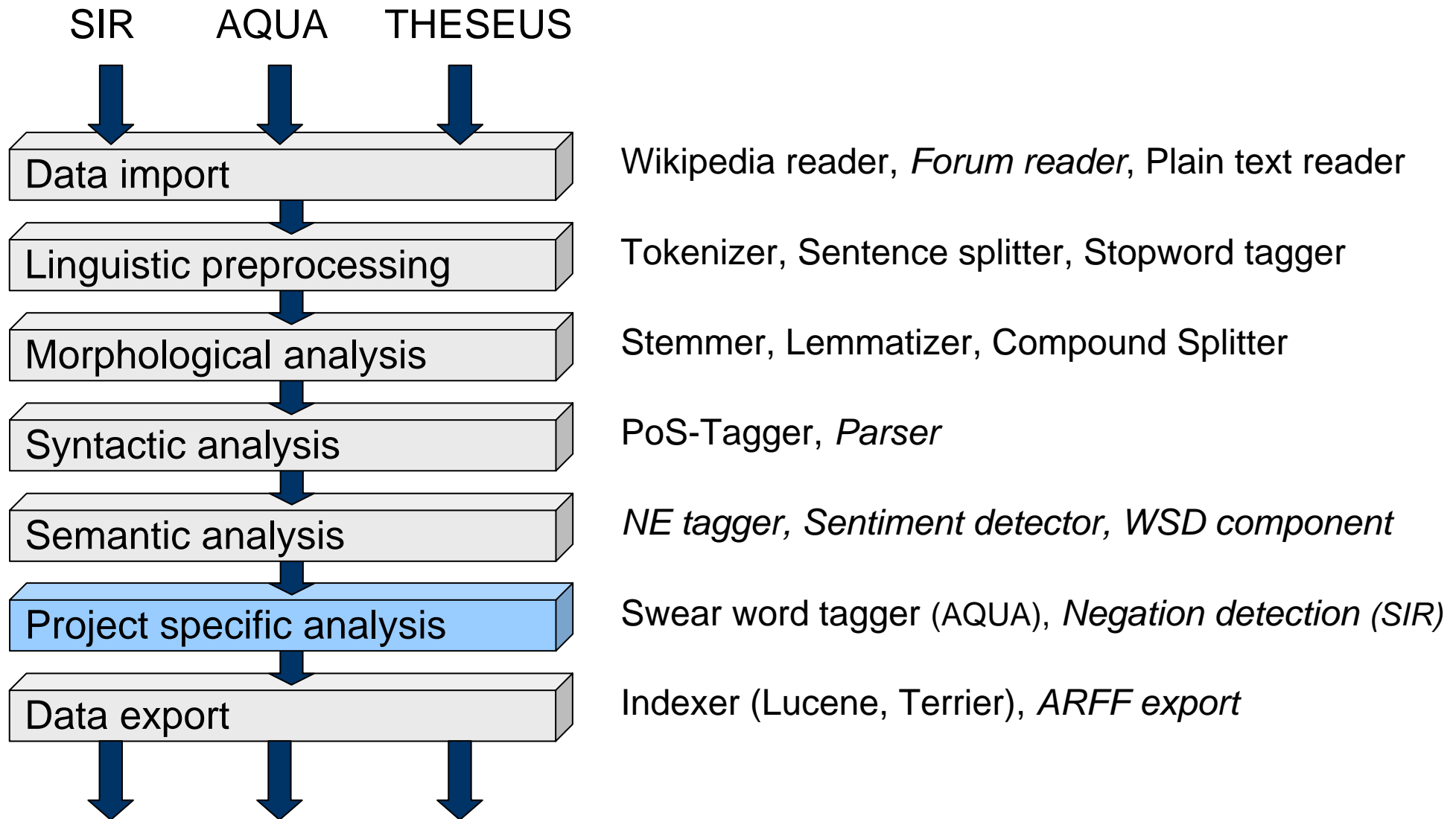


THESEUS - TEXO

- Large-scale BMBF-Project, industry (SAP, Siemens, etc.)
- Service Marketplaces in Web 2.0
 - Find services, both users and machines
- Problem:
 - Only keyword-based search
 - Lack of ontologies for semantic search
- Solution:
 - Use natural language descriptions of web services
 - Apply Semantic Information Retrieval
 - Community Mining for optimized service selection

→ Darmstadt Knowledge Processing Repository

UIMA components





Advantages of UIMA

- Components can be shared between projects
- Shared model of thinking
 - “Reader + Annotators + Consumer”
 - Configuration of components
- Descriptive component orchestration



Challenges

- Agree on a type system
 - No automatic type mapping
- Some rough edges in UIMA
 - No real plug'n'work with PEAR packages
 - Using constraints to align annotations seems to be slow



Wish list

- Automatic type matching
- Better tool support
 - Improving Eclipse plug-ins (robustness, features)
 - Refactoring of UIMA components
 - CPE runner ++ (automatic logging, performance monitor, etc.)
- Plug'n'work approach
- “Import by name” in CPEs
 - Or make `#{CPM_HOME}/path` also work for readers/consumers
- Construct XML descriptors from Java annotations
- More intuitive API



Thank you very much!

- Acknowledgements:
 - DFG for funding “Semantic Information Retrieval”
 - DFG for funding “Automatic Quality Assessment and Feedback in eLearning 2.0”

<http://www.ukp.tu-darmstadt.de/>