4A Framework

Annotations Anywhere, Annotations Anytime

Jaroslav Dytrych

Natural Language Processing Group
Faculty of Information Technology
Brno University of Technology

April 28, 2011

Supervisor: Doc. RNDr. Pavel Smrž, Ph.D.

Outline

Outline

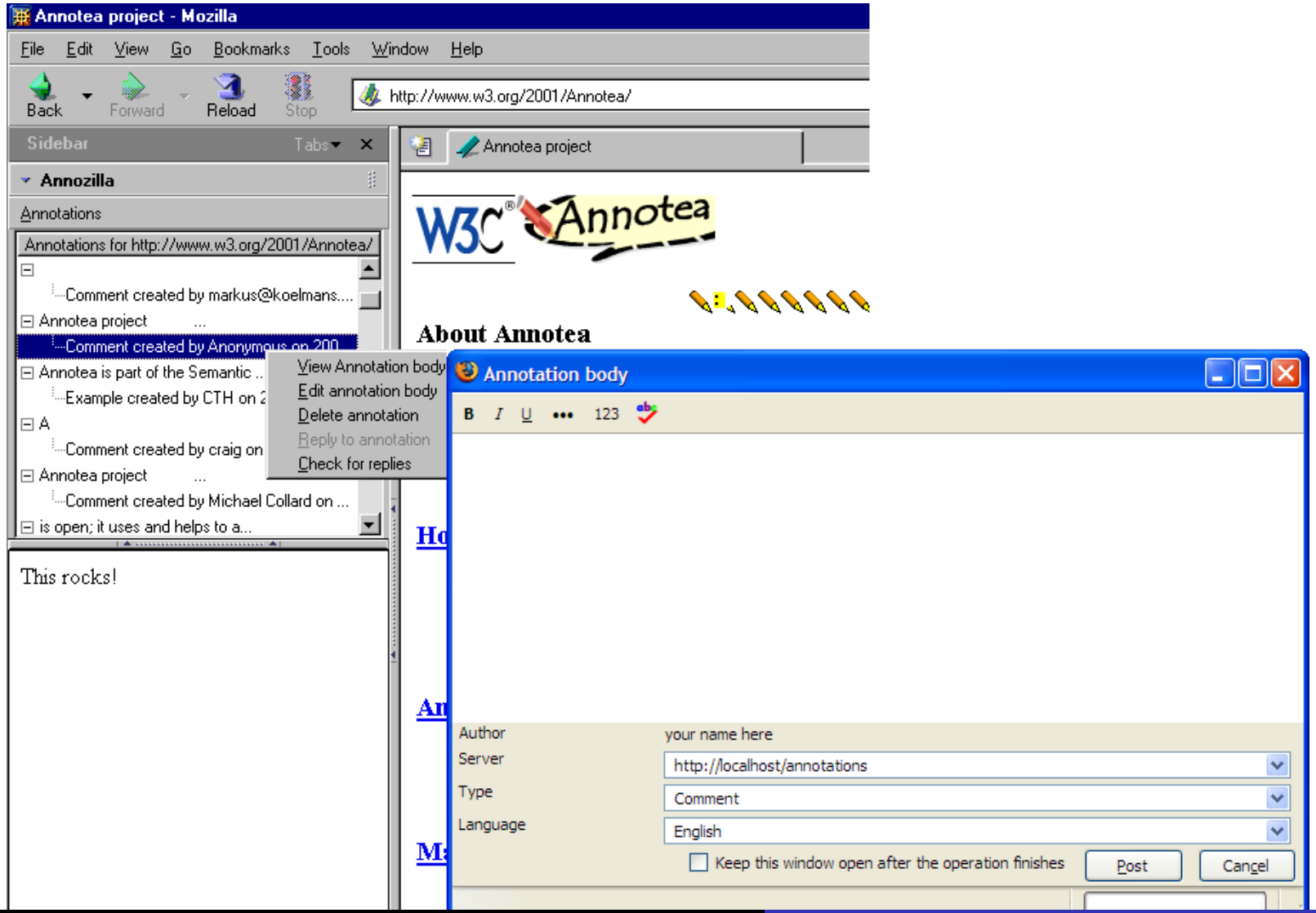
Motivation

- Success of knowledge-based systems (and the Semantic Web) depends on metadata
- There is a lot of unstructured data
- Automatic extraction techniques are unreliable
- Manual annotation is tedious
- Structure of knowledge is known in a limited number of (simple) cases only
- Annotations can be used for knowledge structuring

Desiderata

- Creating complex annotations should be as simple as tagging
- Users should be able to annotate in the applications they know and use for reading and content creation
- Automatic methods should facilitate the work by suggesting annotations
- Whenever possible, annotations should be anchored in text
- Knowledge structuring should build on popularity of social tagging

Annozilla


Annotea project - Mozilla
File Edit View Go Bookmarks Tools Window Help
Back Forward Reload Stop
http://www.w3.org/2001/Annotea/
Sidebar
Tabs
Annozilla
Annotations
Annotations for http://www.w3.org/2001/Annotea/
Comment created by markus@koelmans....
Annotea project ...
Comment created by Anonymous on 200
Annotea is part of the Semantic ..
Example created by CTH on 2
A
Comment created by craig on
Annotea project ...
Comment created by Michael Collard on ...
is open; it uses and helps to a...
This rocks!
View Annotation body
Edit annotation body
Delete annotation
Reply to annotation
Check for replies
Annotea project
W3C Annotea
About Annotea
Annotation body
B I U ••• 123
Author your name here
Server http://localhost/annotations
Type Comment
Language English
Keep this window open after the operation finishes
Post
Cancel

Sources of inspiration

Annotating

- Anozilla, SharedCopy, Stickis ...
- PREP Editor, Bundle Editor, FAST ...
- DiLAS

Annotation exchange

- Annotea (W3C)
- API for project InsightLink (IBM)

Real-time collaboration

- Google Wave
- Novell Pulse
- Google Docs
- Microsoft Office Live

Shortcomings of existing solutions

- It is hard to share and reuse annotations
- The content being annotated cannot be edited
- There is no structuring of annotations
- Fixed set of annotation types
- Limited integration of advanced information extraction tools

Annotations Anywhere, Annotations Anytime

- Simultaneous editing and annotating in heterogeneous environments
- Working on the same text in various formats
- Real-time collaboration
- Simple and structured annotations
- Annotation suggestions and automatic update of annotations
- Documents can change

Outline

Annotation format

- Based on RDF
- Annotations and tags together
- Simple and structured annotations
- Relations among annotations
- Dynamic structures
- Robust positioning

Annotation format

Encoded information:

- type of annotation
- date and time of creation
- author
- URI of a server copy of the annotated document
- textual fragments (XPath, offset, length and textual content),
- content of annotation
- attributes

Attributes:

- structuring of annotations, dynamically added
- simple data types, links to annotatoins or nested annotations

Exchange protocol

- Based on XML
- Transport in various protocols on lower level
- Messages can be combined to make communication efficient
- Two-way real-time asynchronous communication
- Synchronization and actualization of documents
- Subscribtion to annotations from defined sources
- Exchange of annotations and types of annotations
- Annotation suggestions

Exchange protocol

- session management
- users and groups
- subscription to annotations
- synchronization of annotated documents
- exchange of annotation types
- exchange of annotations
- suggested annotations
- settings
- errors and warnings
- explicit confirmation

Outline

Proof-of-Concept Implementation

Server

- core of the system functionality
- modular design
- universal

Clients

- addons and plugins for various applications (document viewers, editors, e-mail clients, etc.)
- varying complexity

4A Server

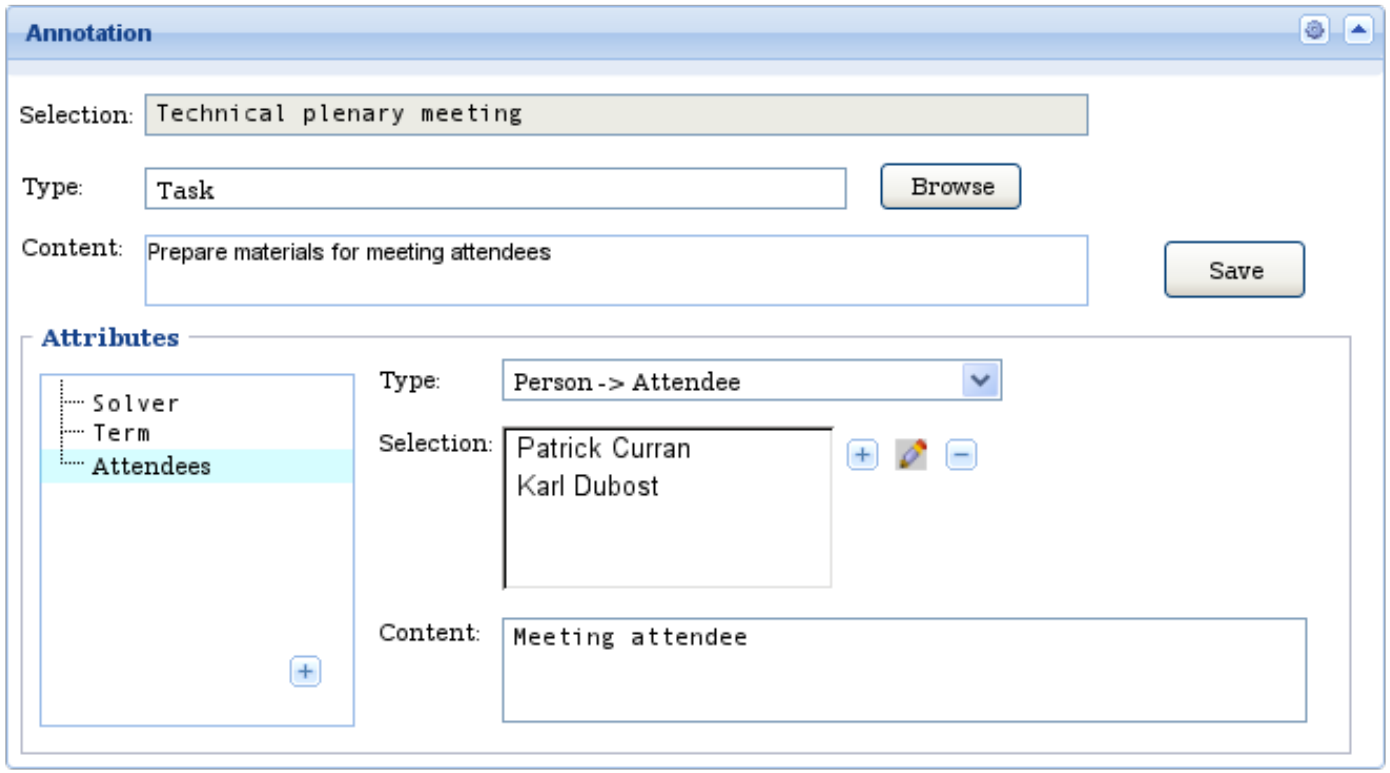
- implemented in Java
- communication with clients over Comet (Grizzly Framework)
- ORM EclipseLink (JPA 2.0)
- MySQL database
- annotation suggestions from the KiWi project
- other modules under development

4A Clients

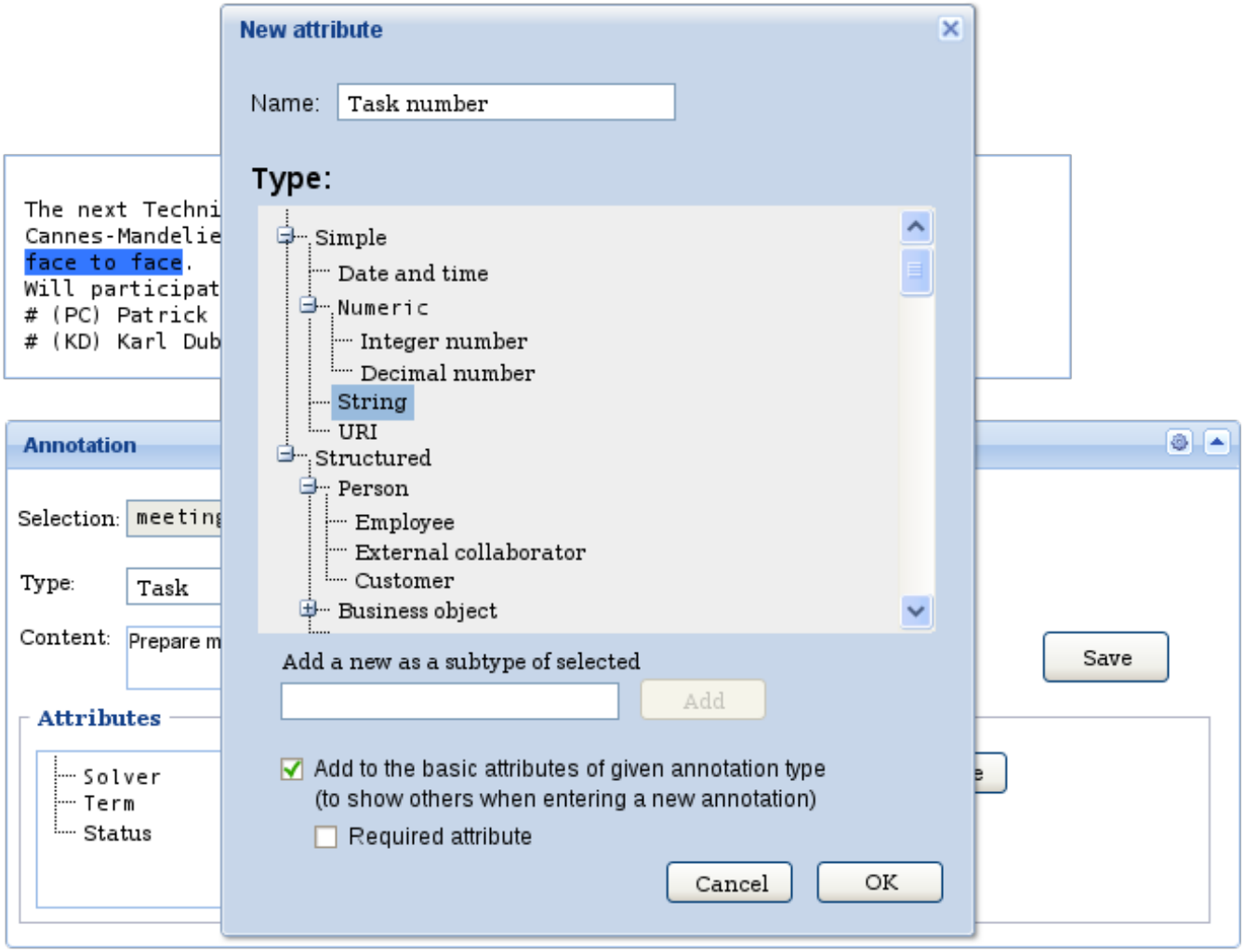
- specification of required and optional functionality forms a part of the framework
- implemented by several authors
- the first functional client – a plugin for TinyMCE
- addon for Mozilla Firefox being developed
- plugin for Microsoft Internet Explorer being developed

Structured annotation

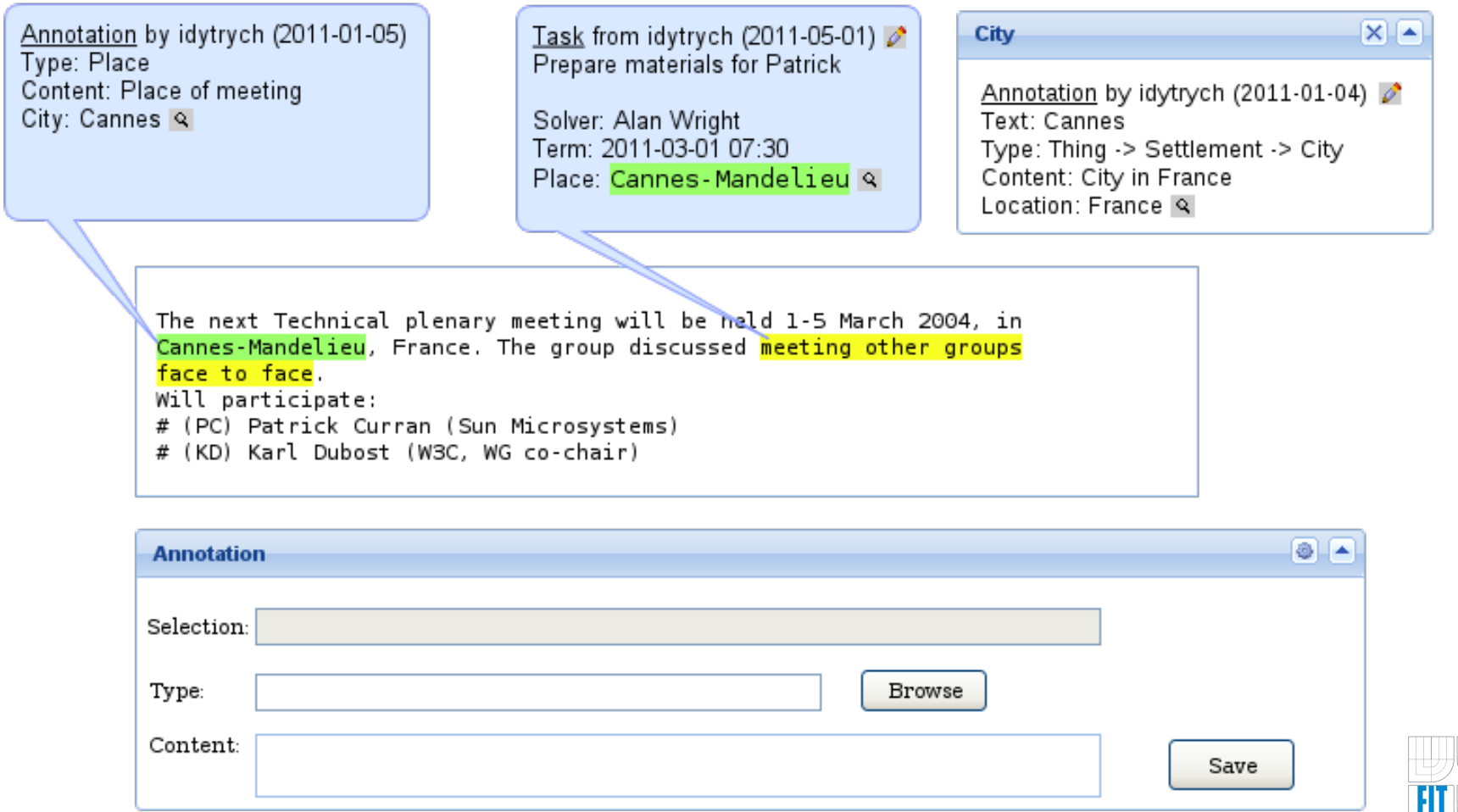
The next Technical plenary meeting will be held 1-5 March 2004, in Cannes-Mandelieu, France. The group discussed meeting other groups face to face.
Will participate:
(PC) Patrick Curran (Sun Microsystems)
(KD) Karl Dubost (W3C, WG co-chair)



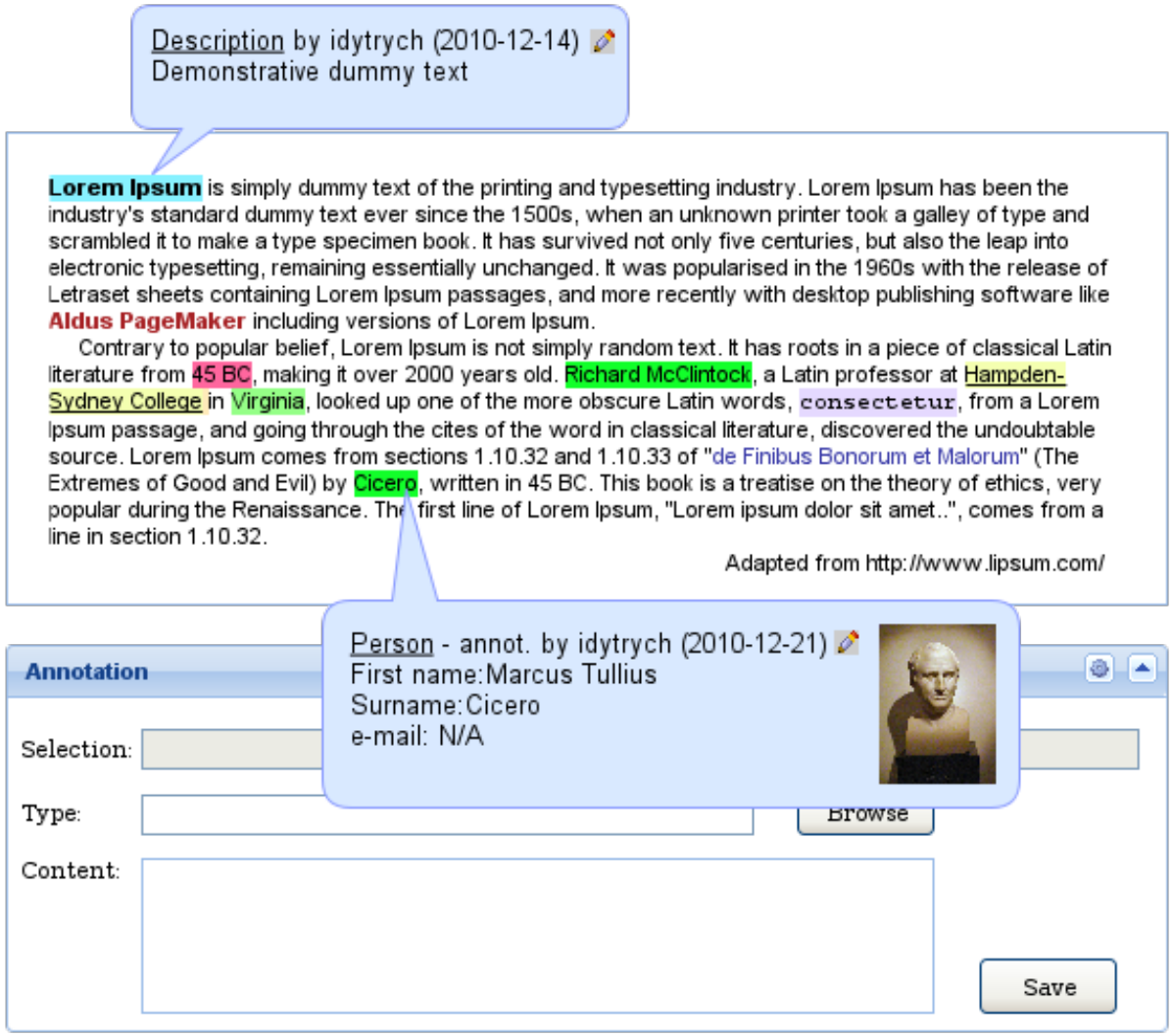
Adding of attribute



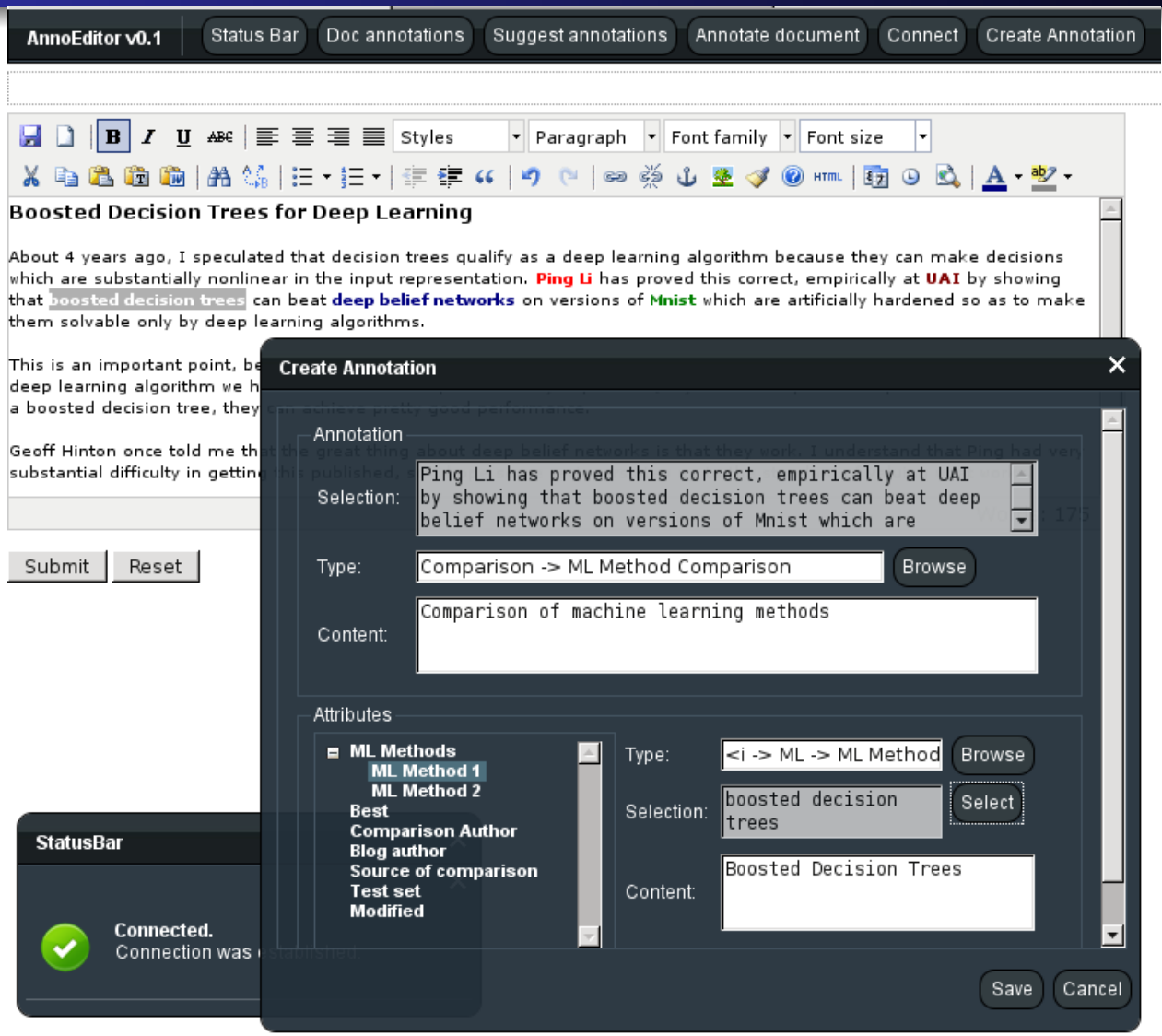
Visualization



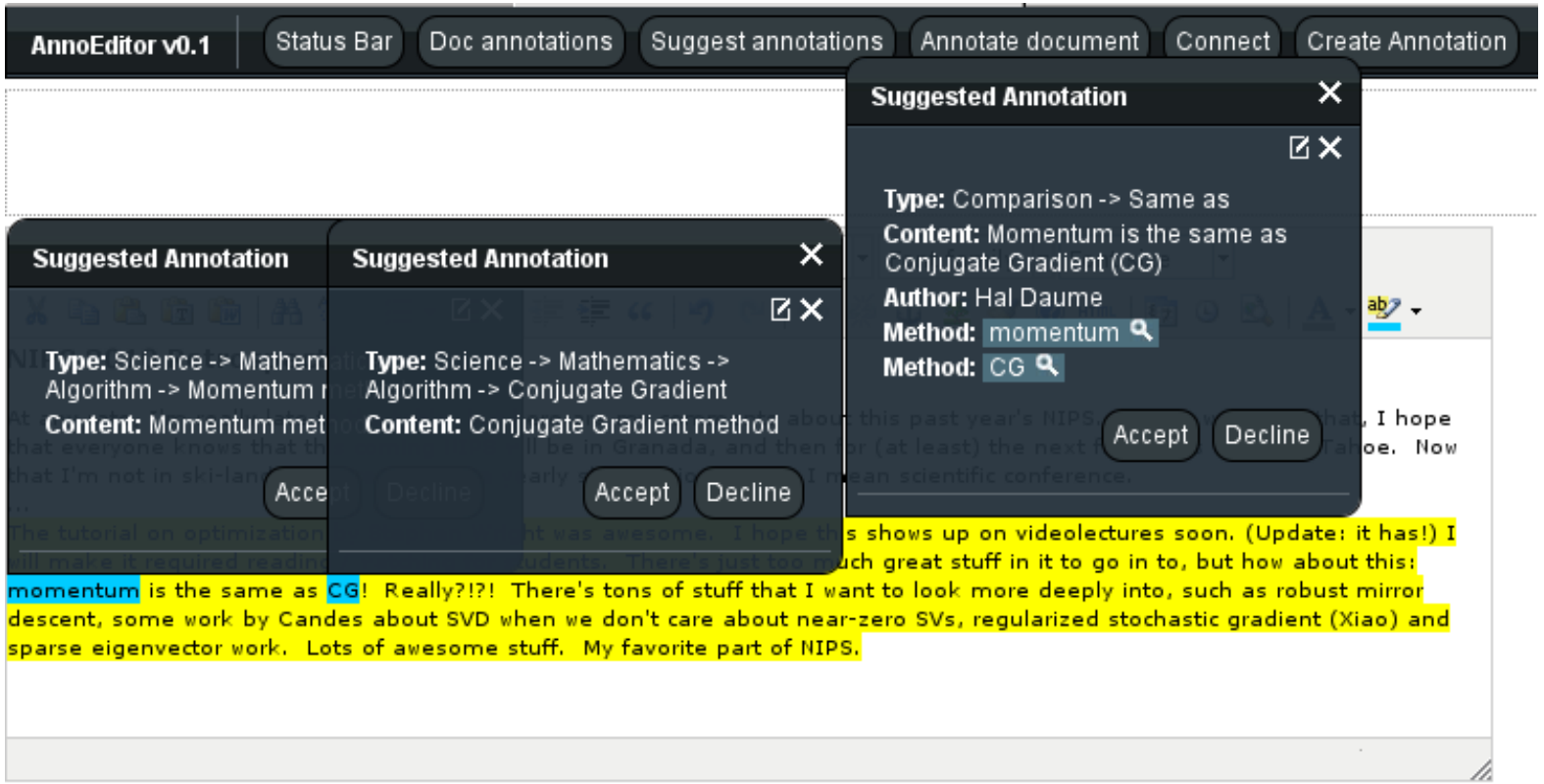
Advanced visualization



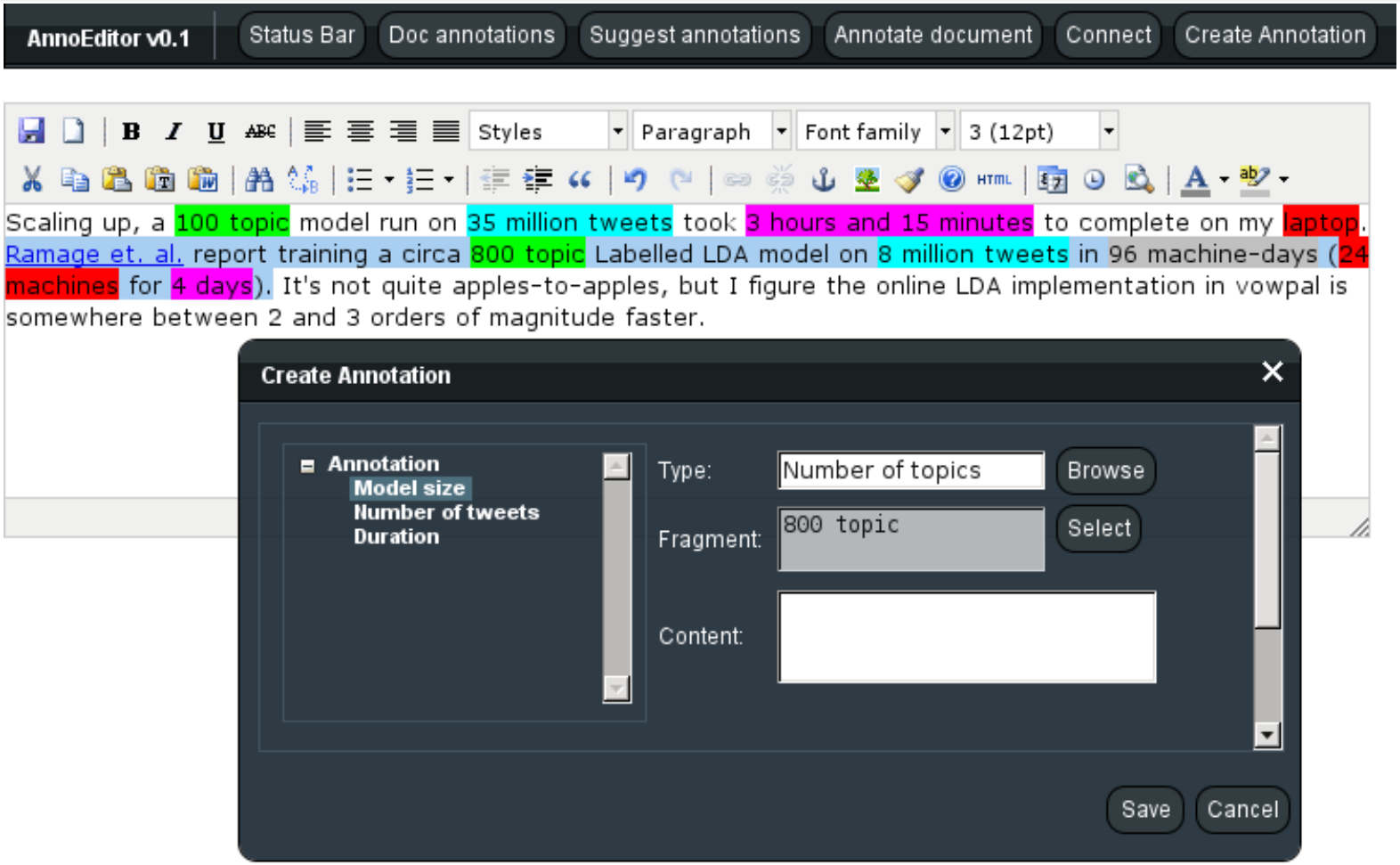
Working solution



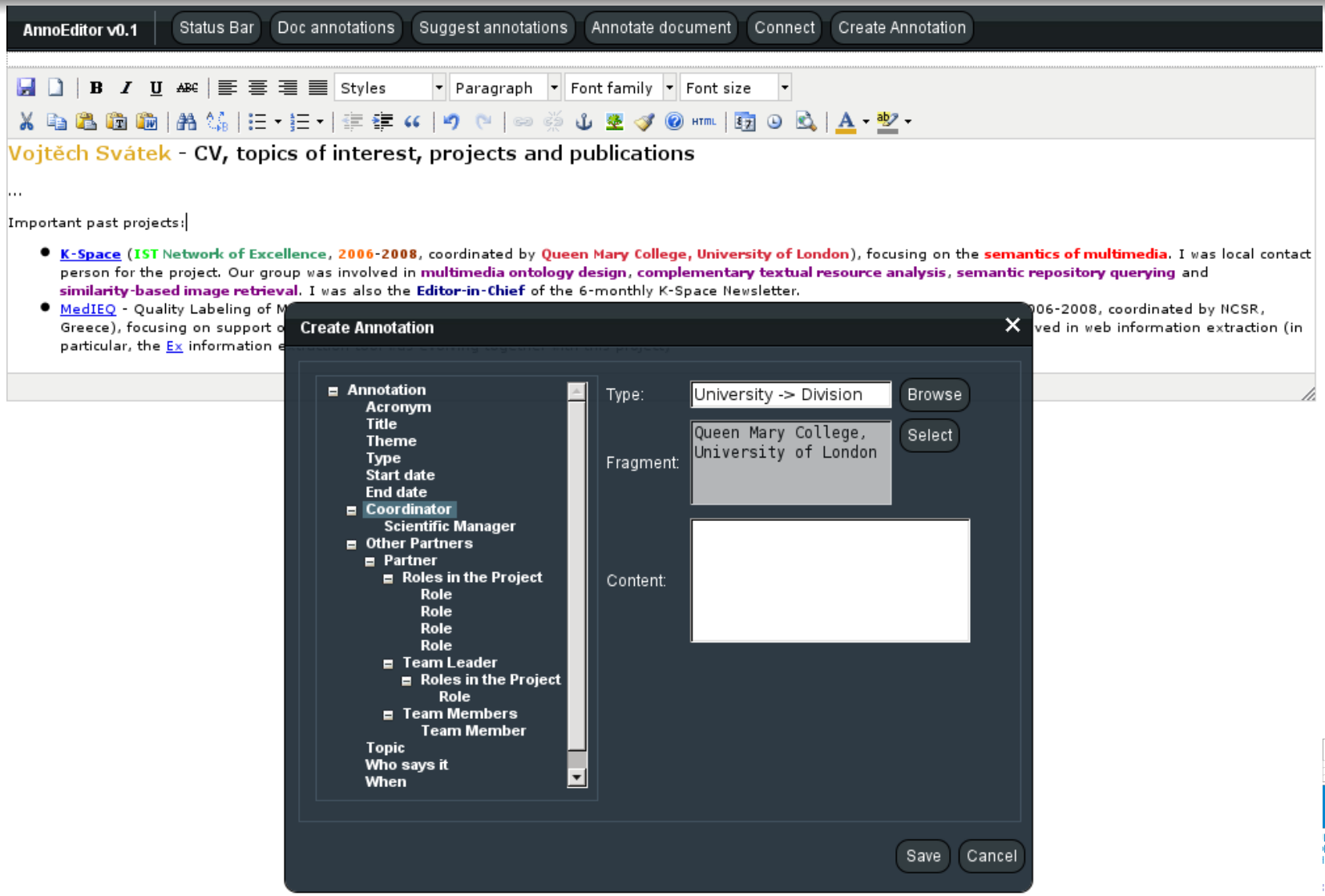
Suggestions



Semiautomatic annotation



Multimedia ontology



Outline

Future development and validation

- Finish clients for web browsers
- A client for annotating PDFs
- Server enhancements
  - better annotation suggestions
  - many instances of the server, distributed environment
  - additional modules (e.g., workflow support)
- Validation by means of annotation experiments