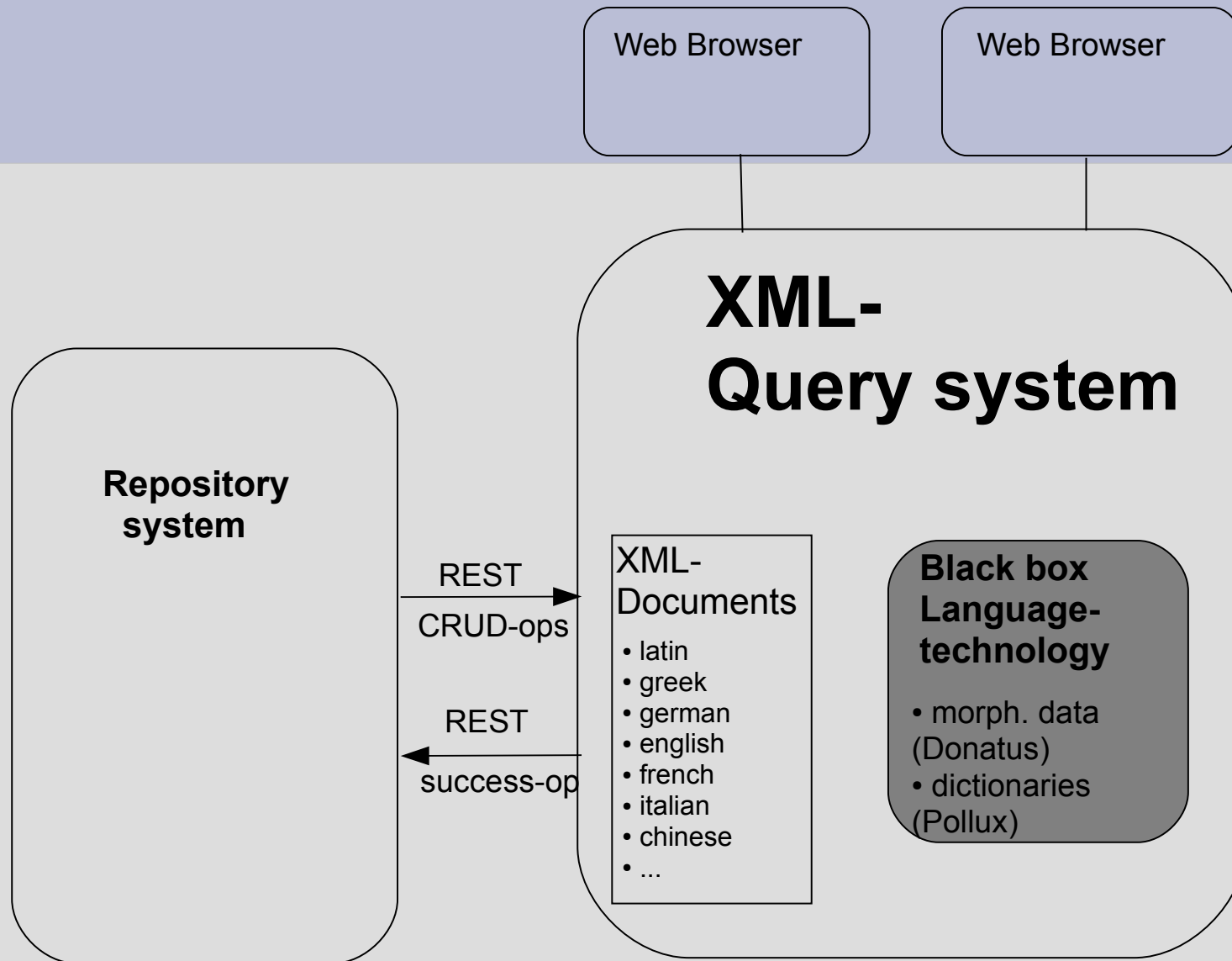


# XML query system/language technology

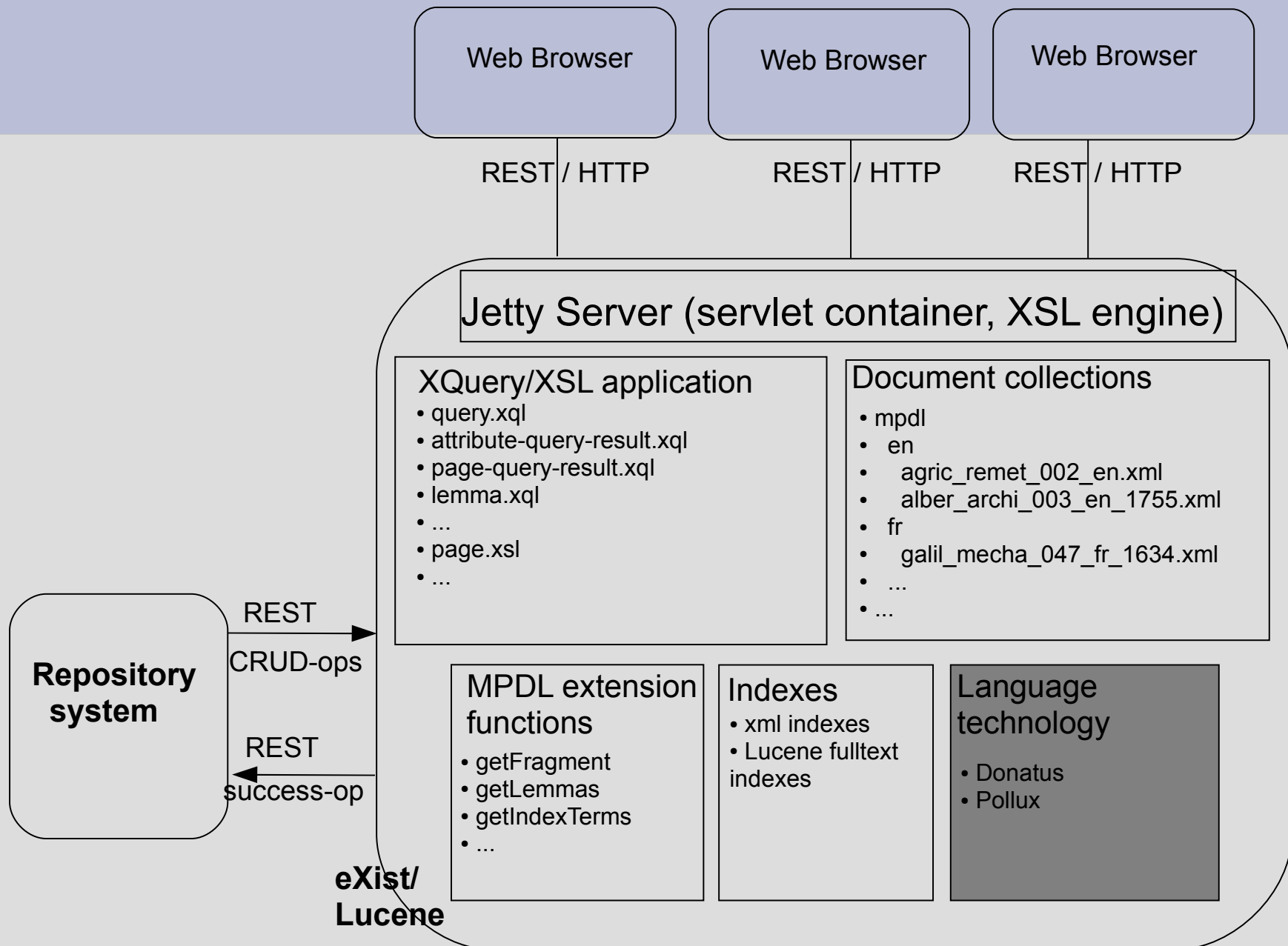
## Content

1. Architecture
2. Functionality
3. Language technology
4. Presentation

# Architecture (1)



# Architecture (2)

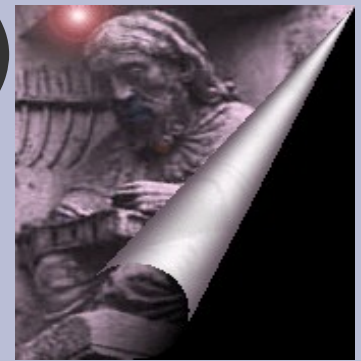


# Functionality

- **Presentation of XML documents**
  - Presentation of pages (XML fragment)
    - modes: Text, Text/Pollux, Image, XML
    - fast: special extension of eXist (getFragment)
  - Navigation in document: page down/up, jump to page
  - Integration of Digilib
  - Enrichment of pages
- **Query system**
  - Document bases: Archimedes, Echo, TEI
  - Search in Document bases and in documents
    - boolean queries in metadata (Lucene query power)
    - (morphological) fulltext queries (also with polyhierarchical lemmas)
    - structural queries: XPath und XQueries within/without documents
    - hit marking / presentation of morph. variants
  - dokument index with links to dictionaries (Pollux) and Wikipedia and to morph. databases
  - Presentation of morph. data (Donatus) and dictionaries (Pollux)
- **CRUD interface**
  - create, read, update, delete documents (web interface)
  - XML-RPC interface to eXist
  - REST interface to repository system eSciDoc

# Language technology (1)

## Donatus



### What is Donatus language technology (see: [archimedes.fas.harvard.edu](http://archimedes.fas.harvard.edu))

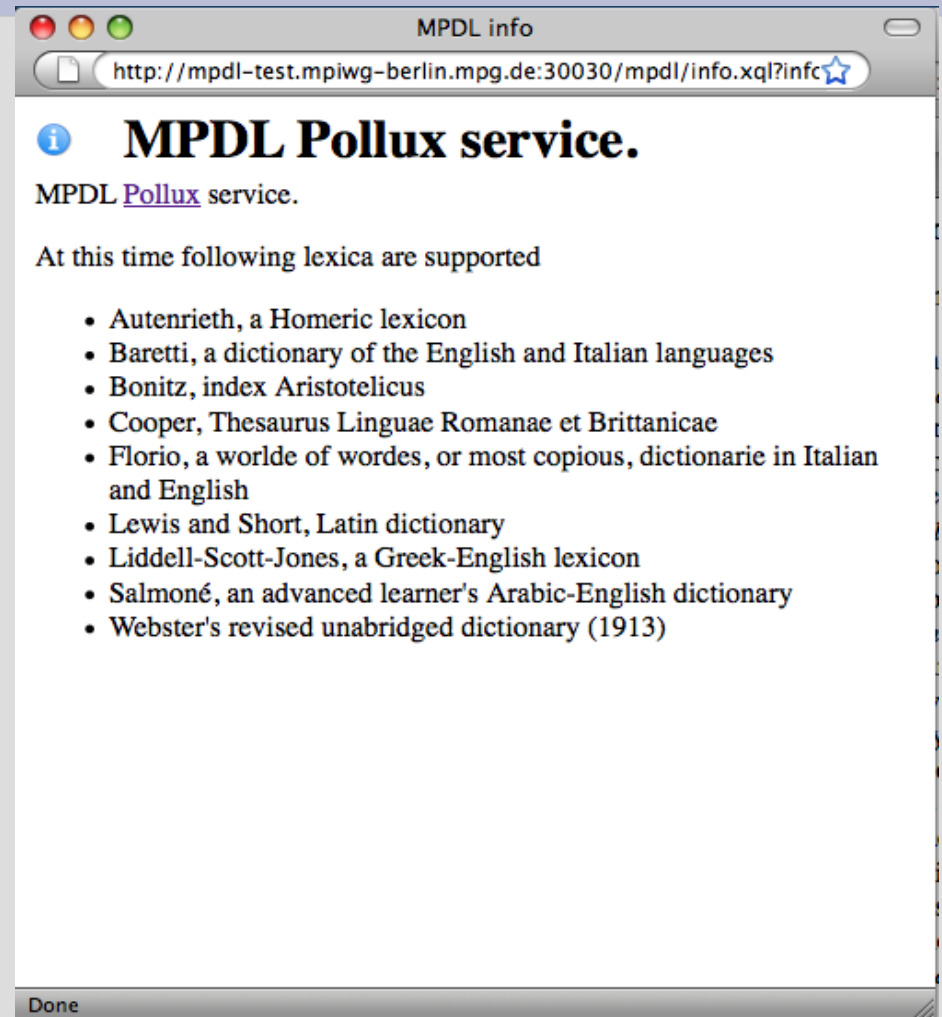
- morphological database in languages arabic, german, english, french, greek, italian, latin, dutch
- overall ca. 3 millions of word forms with its lemmas und grammatical information
- Usage: during morph. indexing of documents and query by morph. entries (pure Java)
- special greek and arabic Unicode conversion
- Presentation of morph. expansion in query result
- Presentation of morph. information in document index
- Web interface:
  - e.g.: <http://mpdl-proto.mpiwg-berlin.mpg.de/mpdl/lt/lemma.xql?language=la&form=accedo>

# Language technology (2)

## Pollux

### What is Pollux language technology (siehe: [archimedes.fas.harvard.edu](http://archimedes.fas.harvard.edu))

- 9 digital dictionaries/thesauri (arabic, english, greek, italian, latin)
- ca. 460.000 entries



# Language technology (3)

## Pollux

### What is Pollux language technology

- datenbase of 9 dictionaries (BerkeleyDB)
- special greek and arabic Unicode conversion
- Presentation of document pages: Pollux mode with links to Pollux entries (over morph. database)
- Document index: link to Pollux entry
- Web interface
  - e.g.: <http://mpdl-proto.mpiwg-berlin.mpg.de/mpdl/lt/lex.xql?language=la&form=accedo>

# Language technology (4)

- Language technology (Pollux, Donatus) implemented modular (pure Java)
- could be used in different usage scenarios (in client or servers)
  
- All project software is free (open source)



# Presentation

- <http://mpdl-proto.mpiwg-berlin.mpg.de/mpdl/query.xql>