Discussion of Bernstein Integration Strategies

Clare Llewellyn
University of Liverpool Library

Dr. Robert Sanderson
Dept. of Computer Science
University of Liverpool

clare.llewellyn@liv.ac.uk, azaroth@liv.ac.uk

Overview

- Use Cases
- Overall Integration Strategy
- Architectural Options
  - Topic 1: Metadata Enhancement
  - Topic 2: Extent of Description
  - Topic 3: Administration of Databases
- Summary

Use Cases

- **Integration**
  Methods and data sources for integration
- **Search**
  Types of search to be supported by the system
- **Display**
  Ways of displaying the results of Search
- **Statistics**
  Options for post processing the results of Search
- **Other**
  Non use cases or other very specific scenarios

Use Cases: Integration

- DbSearchSpeed
- IncunabularLinkUseCase
- ExternalInterconnectivityUseCase
- DatabasesUniformImageScaling
- DbErrorTreatment
- BriquetSearch

Use Cases: Search

- MultilingualSupportUseCase
- DbSearchReliability
- DbSearchAmbiguity
- DbSearchMissingData
- DbSearchSynonmy
- DbSearchShapeAmbiguity
- DbSearchByElements
- DbSearchIncompleteWatermarks
- DbSearchMulticriteriaAndLogicalOperators
- DbSearchHeterogeneousContent
- WmTerminologyUseCase
- CartographyUseCase
- DatingManuscriptByWatermark
- MappingOfSynonyms
- DbSearchByMeasurement
- SymbolSearch

Use Cases: Display

- SearchAllResource
- MultilingualSupportCase
- DatabaseUniformImageScaling
- BookmarkFunction

Use Cases: Statistics

- DbSearchStatistics
- DbAnswerExport
- PaperDatingUseCase
- PaperAuthenticationUseCase
- CartographyUseCase
- CartographyBibliographyIntegrationUseCase
- BibliographicalUseCase

Use Cases: Other

- DbSearchWaterMarkandCounterMark
- DbSearchWaterMarkTwins
- ImageProcessingUseCase
- DataPullSchema
- DoiUse

Integration Strategy

Primary Consideration:

It is not possible to support all of the use cases provided without additional metadata describing the components that make up each watermark.

Primary Decision:

a) There will be some additional description done

or ***b)*** Not all use cases will be supported

Integration Strategy 2

You are thinking: ?!?!!! >:(!!!

- First step is always to analyse existing work:

  classification in terms of concept and terminology
- Need terminology mapping for interfaces and textual search
- Possibility of automatic construction from existing descriptions
- Mapping to (for example) IconClass easier if already harmonized internally

Suggested Architecture

- Each watermark would have a description in terms of components (cross, bull's head, circle) rather than a hierarchy
- Components have their own descriptions (as appropriate) and relationships to other components (eg above, left-of, within)
- Component fields to be decided by watermark experts
- Type of component (cross vs circle vs shield) to be referenced by a number for language independence
- Descriptions linked to original databases and images, or integrated into them as desired

Search and Display

- By searching for numeric components with fields defined by domain experts, rather than free or structured text, we achieve language and description methodology independence.
- Also for display: 'shield' vs 'escutcheon', 'cross' vs 'croix'

- Allows for incomplete search, boolean searching, ambiguous shape matching and searching for individual elements

- Increases accuracy of matching and preciseness of searches expressed

- Makes manipulation of cross-database results easier and more reliable

Search Interfaces

- **Textual Search:**
  By translation into the numeric representation depending on language and type of user (expert vs lay person), we do not sacrifice precision for multilingualism

- **User Constructed Image:**
  Using recent web technology, it is possible to graphically construct searches rather than relying on language. This would allow easy setting of known search attributes per component

- **Hybrid Text/Image Search:**
  The graphical query could be constructed automatically from textual input, giving the user options to then correct the generated image

Topic 1: Metadata Enhancement

- Status Quo (no additional description)

- Components automatically extracted from existing descriptions
- Components manually described

- High level component description (snake, bull's head)
- Detailed components (horns/no horns, eyes, tongue)
- Very detailed components (lines, hooks, letter shapes)

- No additional attributes
- Additional attributes deemed important
- All additional attributes deemed useful

- Spatial relationships between components or not

Topic 2: Extent of Enhancement

- Status Quo (no additional description)

- Some databases extended
- All databases extended

- Some watermarks in database described
- All watermarks in database described

- Some watermarks described in detail
- All watermarks described in detail

- All descriptions done centrally (by same person/team)
- All descriptions done distributedly (by database provider)

Topic 3: Database Administration

- All descriptions held in one central database
- Descriptions held centrally in multiple databases
- Descriptions held centrally and distributed
- All descriptions held in distributed databases

- Descriptions completely separate from original
- Descriptions linked to original database
- Descriptions integrated into original database

- Proxy server(s) to act as gateway to database
- Direct connection to database

Summary

- Use Cases analysed
- In order to fulfil use cases, additional description needed
- 1: Types of description enhancement
- 2: Extent of description enhancement
- 3: Administration of enhanced data
- Construction Interface "Demo"