

eHumanities and Shifting Asymmetries in Cultural Flows – an Ontology for cultural concepts?

Workshop: Mapping HASS – e-Research for Humanities,
Arts and Social Sciences,
OGF 23, Barcelona, 5. June 2008

Peter Gietz, EXCluster “Asia and Europe”,
University of Heidelberg
peter.gietz@daasi.de



RUPRECHT-KARLS-
UNIVERSITÄT
HEIDELBERG



The Cluster of Excellence

- "Asia and Europe in a Global Context"
- Part of the German Federal Excellence Initiative
- "to establish internationally visible, competitive research and training facilities"
- A cluster of ca. 50 interdisciplinary projects
- Analysing cultural interactions between Asia and Europe
- Includes a digital research infrastructure called Heidelberg Research Architecture (HRA)



RUPRECHT-KARLS-
UNIVERSITÄT
HEIDELBERG



The Cluster of Excellence

- Has a dedicated agenda:
 - Finding and analysing flows of concepts
 - That happen within shifting asymmetries
 - Globalisation is not a new phenomenon, but has happened since the beginning of man kind
 - Every thing is part of this global process
 - People are a medium of such flows



What are the technical requirements?

- Large amount of different databases already exist
 - Language resources
 - Image resources
 - Bibliographical resources
 - Music, Films, etc. etc.
- Most Cluster project will have the need for such or new data bases

What else?

- Martin Doerr:
"The ultimate goal of users seeking information is not to get an "object" but to understand a topic. Understanding is built on associations. Associations are found in database records, digital objects, metadata or indices."



How to store concepts

- How to get all these data into one system?
- How can one-dimensional metadata be enhanced?
- How can flows be described?
- How can phenomena be linked with each other?
- How can the system answer intelligent questions?



First answers

- Different data bases can be integrated in a loosely coupled system
 - In a Service Oriented Architecture
- One central metadata base can be a central retrieval point
- One big full text index on cluster resources can be helpful
- New Semantic Web technologies might be even better for storing and finding conceptual flows



More first answers

- We are trying to use as much infrastructure if the University as possible
- We will need more than that
 - Storage (Grids?, Clouds?)
 - Communication platforms
 - E-Learning systems
- We want to use existing Software where possible
 - Some Interfaces will have to be done by ourself



Ontology

- A multi dimensional system for relating information objects
- A classification system (like Dewie Decimal System) can be called a single dimensional metadata system
 - there is only one relation type: Is subclass of
- A Thesaurus (like Roget's Thesaurus) uses more dimensions:
 - is part of semantic field, is connected with
 - Other thesauri have even more, like synonyms, antonyms, subclass, etc.



Ontology

- An Ontology is all the above and much more
 - We have a hierarchical class model (like in DDC)
 - We have an unrestricted number of relation types (not only the few of a thesaurus)
- So we can store classifications and thesauri in ontology stores
- The best way to formalize an ontology entry is:
 - Subject – predicate – object
 - Where subject and object are classes or class instances
 - And predicate is a relation type

– RDF triples



What do we want to do with ontologies

- find integrated knowledge
- produce new knowledge
- provide evidence for new hypotheses
- verify or challenge old hypotheses



How could a cluster ontology look like?

- There could be an overall ontology which includes classes and relation types that all cluster projects will need, e.g.,
 - classes like person, time, location, event
 - relations like hasType, wasInfluencedBy, isIdentifiedBy, isComposedOf
- Every cluster project using the ontology technologies could built a sub ontology fulfilling their very special needs, e.g.
 - Classes like picture elements, poses, etc.



Are there standards out there?

- CIDOC CRM

- is an international standard (ISO 21127:2006)
- Is a extensible ontology for concepts and information in the field of cultural heritages
- Meant for museums (any kind of), archives and libraries – also for information exchange
- Creates a formal semantic for information integration
- Supports standard technologies like RDF and OWL



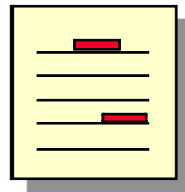
CIDOC CRM

- “it describes in a formal language the overt and implicit concepts and relations relevant to cultural heritage information”
- Consists
 - of 87 classes and subclasses
 - and 184 predicates
- Some of them are very useful for us



The CIDOC CRM

Epistemic Networks: orthogonal to sources



Next steps

- Evaluating ontology tools like protégé
- Creating a test ontology for a single cluster project
- Analysing CIDOC CRM for applicability in the cluster context
- Setting up an RDF database and experiment with RDF triples vs. OWL/RDF vs. OWL
- Retro engineering of a (cluster) typical article with an ontology

