CREAM – CREAting Metadata for the Semantic Web [1]

Paper Reviewed By:

Chris Halaschek

CSCI8380 – Advanced Topics in Information Systems

Large Scale Distributed Information Systems (LSDIS) Lab

The University of Georgia

Paper Review: “CREAM CREAting Metadata for the Semantic Web”, Siegfried Handschuh, Steffen Staab 2003

Motivation

- Currently
  - Web pages with HTML, XML, etc.
  - Very little semantic markup
- For the Semantic Web vision to become a reality, we need semantic annotation of Web pages, both existing and new

Case Studies

- KA2 Initiative and KA2 Portal [2]
  - Add semantic markup to HTML for knowledge acquisition community and Web portal
  - Pages were *crawled* for annotation
  - Created OntoPad
    - Problems
      - Syntactical human errors – typos
      - Semantic errors – violation of constraints
      - Pragmatic errors - wrong assumptions about the existence of object descriptions

Case Studies

- TIME2Research Portal
  - Annotation for portal in telecom, IT, multimedia, e-business markets
    - People watch news tickers, new reports
      - Compose all in one portal
  - Annotate HTML, Word, Excel
  - Provided OntoAnnotate [3]
    - Could highlight and drag and drop to ontological concept
    - Problem
      - Need to create documents and metadata in one step

Case Study

- Annotations of Paper Abstracts at ISWC-2002
  - Authors provided annotation of title pages and abstracts
  - Provided OntoMat [4] tool
    - Same problems as before

Case Studies - Summary

- Common Problems
  - Hand written annotation – too time consuming
  - Syntactic
    - Human syntactical errors
  - Semantic
    - Violation of semantic constraints
      - String instead of object
  - Pragmatic
    - Incorrect assumptions about the existence of object descriptions

System Requirements

- Consistency
  - Adhere to an ontology
- Proper Reference
  - Identifiers should be unique
- Avoid Redundancy
  - In collection of metadata
- Relational Metadata
  - Have meaningful relationships
- Maintenance
  - Data changes
- Ease of use
- Efficiency
- Support Multiple Ontologies

CREAM – Overview

- Framework that allows for creation of metadata
  - Annotation Mode: create metadata for existing Web pages
  - Authoring Mode: create metadata "Almost for free"…make Web page and annotate at same time
- Creates metadata annotated with respect to domain specific ontology
- CREAM framework implemented in OntoMat

CREAM Modules

- Document Editor
  - Viewer: HTML, PDF, XML, etc
  - Content Generation
    - Allows conventional authoring with insertion of concepts in the ontology
- Ontology Guidance / Fact Browser
  - Browse the concepts and relations in the ontology – allows drag'n'drop

CREAM Modules

- Crawler
  - Determine if entities already exist by crawling other sites
  - Ensures proper reference and avoids redundancy
- Annotation Inference Server
  - Reasoner on metadata for querying
    - Used Ontobroker's underlying F-Logic
- Document Manager
  - Manage his/her documents as well as crawled pages…here maintenance of foreign pages is needed

CREAM Modules

- Meta Ontology
  - Discussed Later
- Metadata Re-recognition & Information Extraction
  - Extract knowledge from pages – semi-automatically
- Storage
  - Annotations in document in Document Manager
  - Annotations in Inference Server

Meta – Ontology

- For ease of use in annotator
  - Labels: Person may be uniquely identified by SSN, but also have a name attribute
    - When creating authoring a document in a drag'n'drop fashion, you want his name not SSN

CREAM Architecture

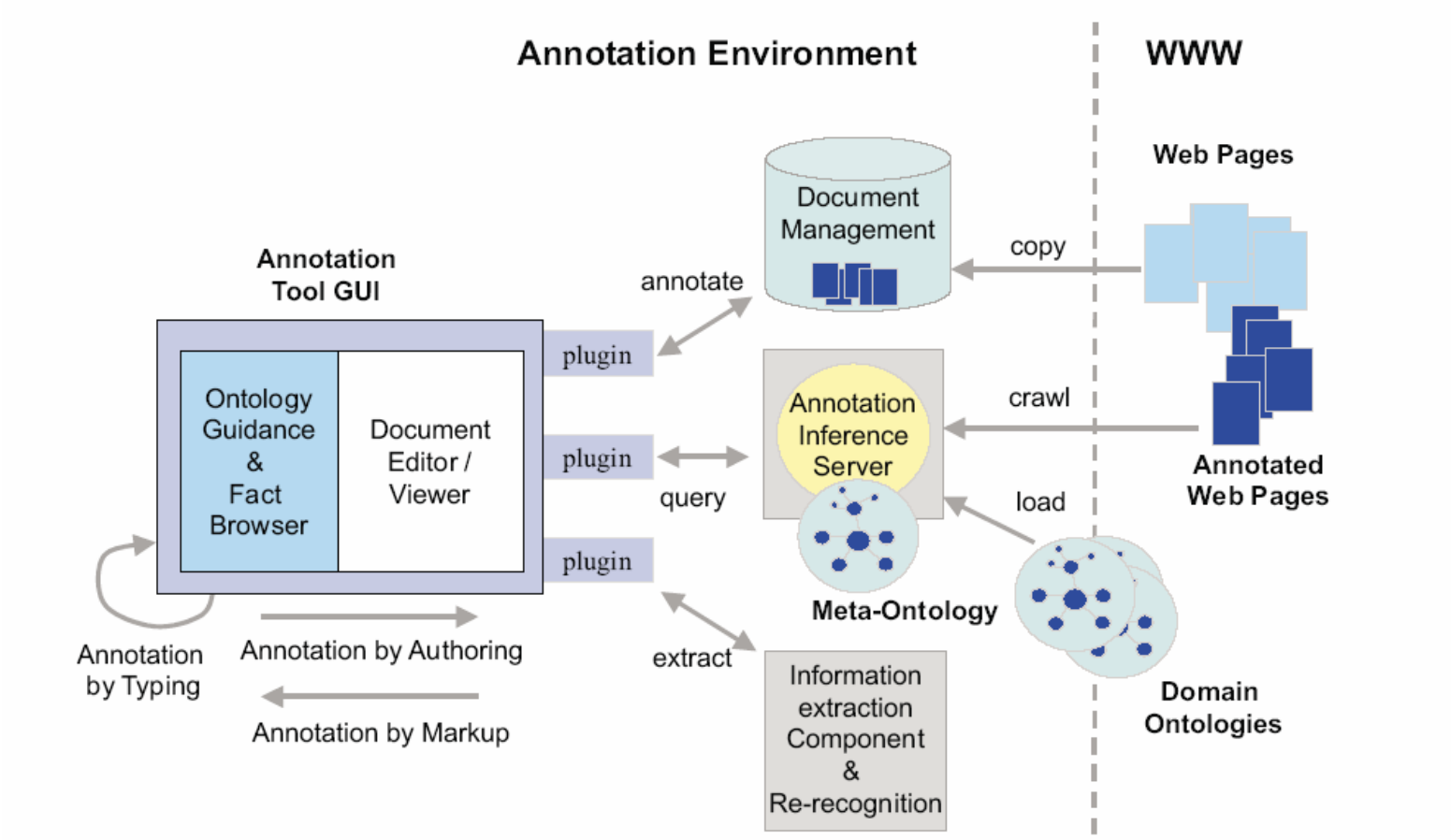

Fig. 2. Architecture of CREAM.

OntoMat

- Implementation of CREAM framework
  - Plug-in structure
- Java based
- Free to download [4]

Interaction with OntoMat

- Types
  - Annotation by typing statements
    - Generation of metadata independent of the document being viewed
  - Annotation by markup
    - Generate internal annotation from already annotated data in the Web page being viewed
    - Drag from Web page to ontology
  - Annotation by authoring pages
    - Creating a Web page in their environment and dragging from ontology to web page – where Meta Ontology helps

Demo

Conclusion

- Framework for creating semantic metadata
- Supports multiple modes of interaction for annotation of web page
- OntoMat is an implementation of the CREAM framework

Questions

- How annotate multimedia
- Will people take the time to annotate already written HTML pages
  - For new pages, think how many people use Word, Dreamweaver, Frontpage, etc.
    - Seems possible

References

[1] S. Handschuh and S. Staab. “CREAM CREAting Metadata for the Semantic Web” 2003 available at

http://www.aifb.uni-karlsruhe.de/WBS/sst/Research/Publications/cream-computer-networks.pdf

[2] http://ka2portal.aifb.uni-karlsruhe.de

[3] OntoAnnotate is now a commercial tool available from Ontoprise GmbH.

[4] http://annotation.semanticweb.org/ontomat

Questions / Comments