TECHNISCHE UNIVERSITÄT DRESDEN

Faculty of Computer Science Institute of Software- and Multimedia Technology, Chair of Multimedia Technology

A semantic extension of a hierarchical storage management system for small and medium-sized enterprises.

Axel Schröder, Ronny Fritzsche, Sandro Schmidt, Annett Mitschick and Klaus Meißner

Berlin, 09/29/2011

Outline

- Problem
- Motivation
- Existing approaches
- Concept
- Prototype
- Conclusion and future work

Problem

- Amount of data stored in companies increases every year [1]
- Fast memory is very expensive
- Current storage management systems treat every file individually without regarding relations between stored files
- Most data is unstructured and content wise unorganized [2]
- Every person has its own understanding of schemas to organize data

Motivation

- Using a storage management system (SMS) can help to save costs
- Even a SMS has to handle a tradeoff between costs, capacity and access time

➔ Using inherent and additional semantic information is the key to a more efficient and more intelligent way to store information

➔ Introduction of a semantic-aware component to enable a SMS to use semantic information (semantic storage extension - SSE)

Existing approaches

- Wide range of projects dealing with semantics in text and multimedia documents [such as 3,4]
- Common used approach: separate meta data from files to increase system performance [e. g. 5]
- Distinction between
  - Semantic Information in File Systems
  - Semantic Information in Additional Systems
    - Such as information from metadata

Existing approaches

Semantic Information in File Systems

- TagFS → annotation of files [6]
  - Automatically and user-controlled annotation of files with keywords (containing metadata, folder names, paths, …)
- Semantic Vectors [7]
  - Meta data converted in vectors spanning a feature space
- Capture events to link related documents [8]

Existing approaches

Semantic Information in Additional Systems

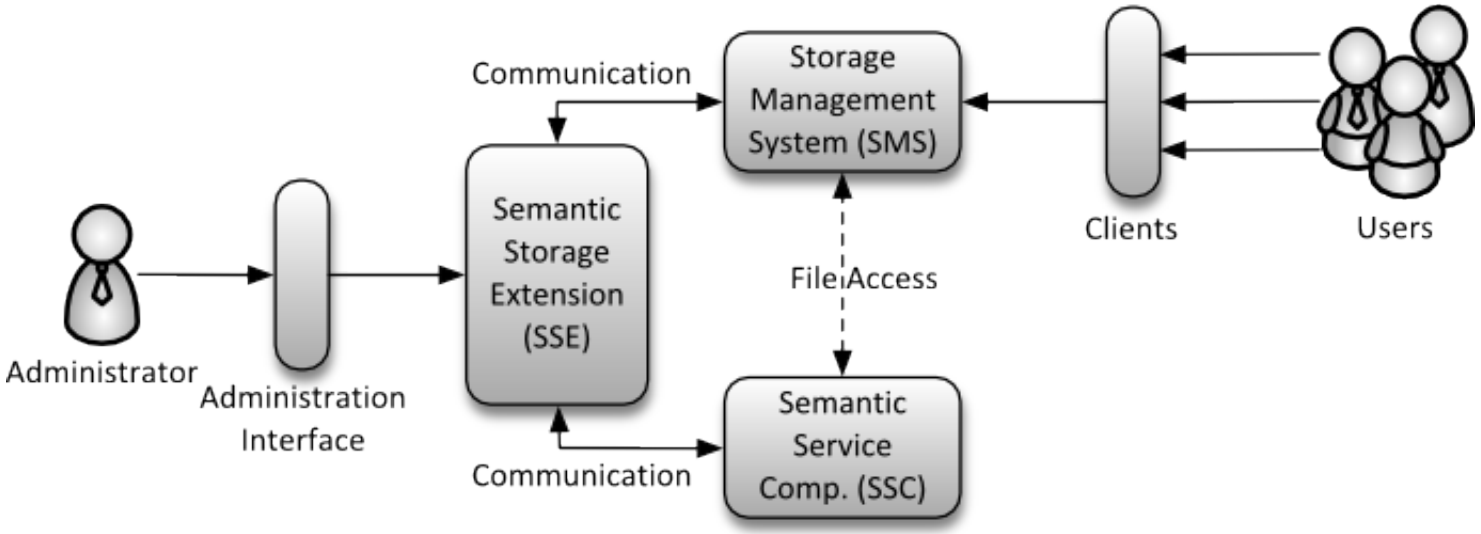
- Information from metadata
- Implicit semantic knowledge (e. g. CBIR, face recognition, …)

➔ consequences for our approach - using of:

- Metadata / attributes from file system
- Explicit knowledge (e. g. from a domain expert)
- Implicit semantic knowledge

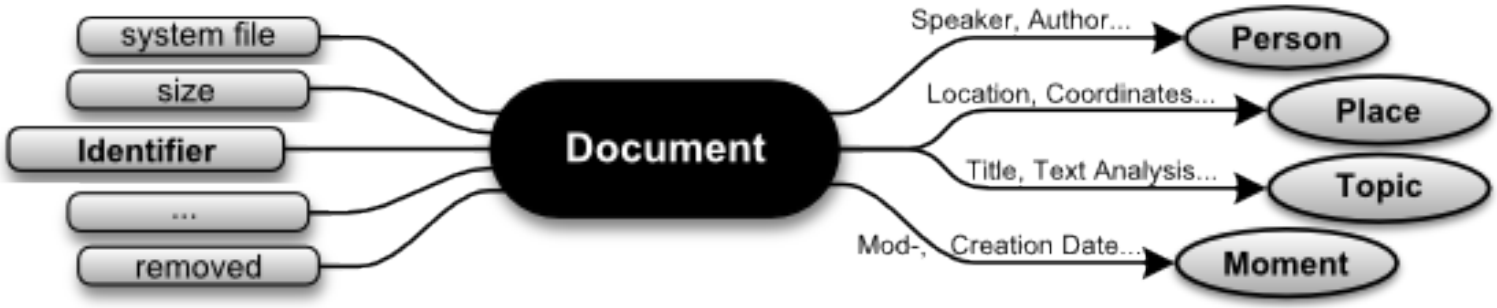
Concept

Architecture



Concept

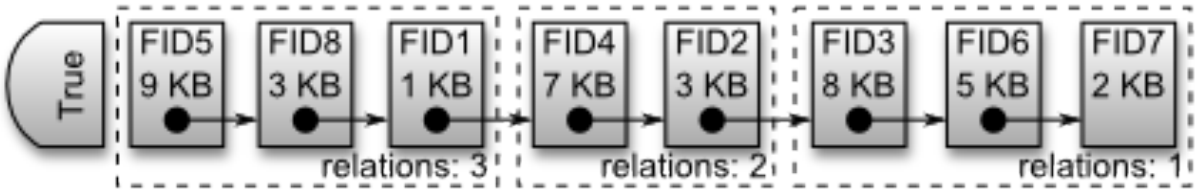
Semantic Service Component



Concept

Modifications in the SMS

- Requesting the SSE
  - Simple request (*file id only*)
  - Directed request (*file id + planned action*)
- Response from SSE
  - Just an advice / recommendation



- Complete analysis and partial analysis of stored data
  - Semantic Service Component need access to stored data

Concept

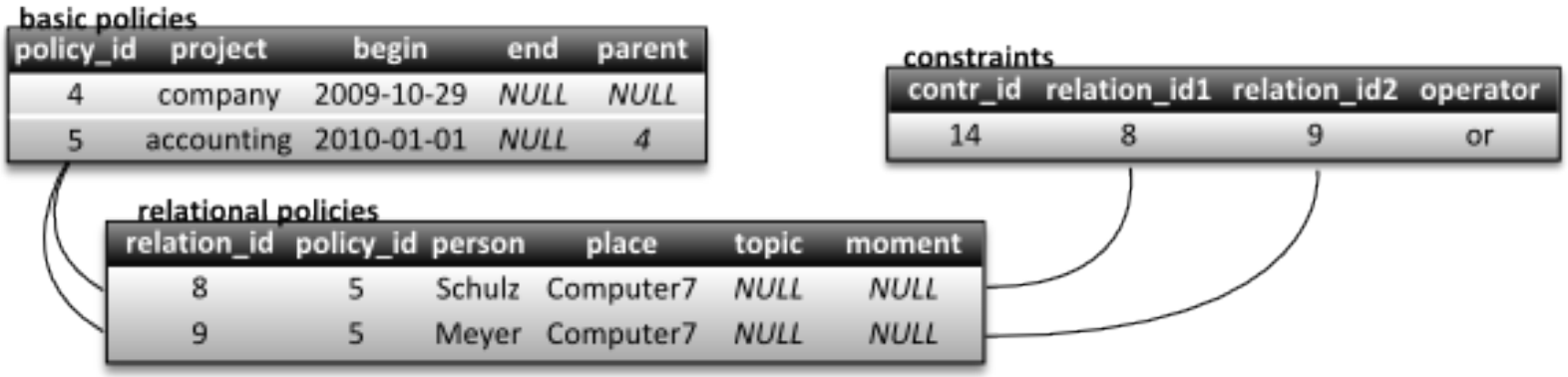
Request the SSE

- Directed request for migration
  - SSE returns *true* if no requested file has relations to used files
- Directed request for retrieval
  - SSE returns true if there are relations to active files
- Directed request for deletion
  - SSE returns false if there is a high relevance to active files

Concept

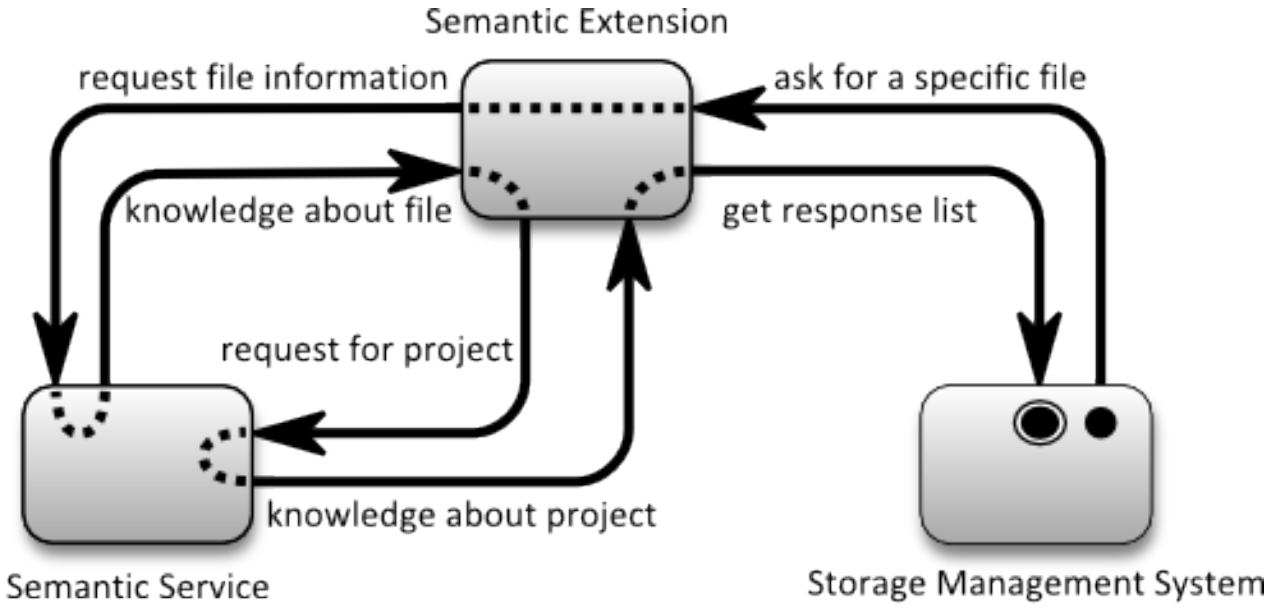
Design of the SSE

- Policies to model explicit semantic knowledge
- Basic policies to model processes / project and temporal boundaries
- Relational and constraint policies to associate projects with persons, places, …



Concept

Design of the SSE



Prototype

- SSE
  - Implemented in Java
  - Using AXIS2
- SSC
  - Implemented in Java
  - Using AXIS2
- Admin interface
  - Implemented in WPF
- HSM (*SMS – not yet connected to the prototype*)
  - Implemented in C# and WPF

Conclusion

- Great lack of using semantics in SMS
- Introduction of the SSE
- Domain-independent architecture through using a SSC and policies to model explicit knowledge
- Future work to:
  - Improve prototype
  - Evaluate concept (especially polices and ontology)

References

(1) Troopens, U., Erkens, R., M uller-Friedt, W., Wolafka, R., Haustein, N.: Storage Networks Explained. John Wiley & Sons Ltd., Mainz, 2 edn. (2009)

(2) Gnasa, M.: Fraunhofer IAIS: Text Mining und Information Retrieval (2009), http://www.iais.fraunhofer.de/4862.html

(3) Orio, N.: Music Retrieval: A Tutorial and Review. Foundations and Trends R in Information Retrieval 1(1), 1{96 (2006)

(4) Sebastiani, F.: Machine learning in automated text categorization. ACM Computing Surveys 34(1), 1-47 (2002)

(5) Hackl, G., Pausch, W., Schönherr, S., Specht, G., Thiel, G.: Synchronous Metadata Management of Large Storage Systems. Proceedings of the Fourteenth International Database Engineering & Applications Symposium pp. 2-7 (2010)

(6) Bloehdorn, S., G orlitz, O., Schenk, S.: TagFS - Tag Semantics for Hierarchical File Systems. Proceedings of the 6th International Conference on Knowledge Management I-KNOW'06 (2006)

References

(7) Mahalingam, M., Tang, C., Xu, Z.: Towards a semantic, deep archival file system. The Ninth IEEE Workshop on Future Trends of Distributed Computing Systems pp. 115-121 (2003)

(8) Weippl, E.R., Klemen, M., Linnert, M., Fenz, S., Goluch, G., Tjoa, A.M.: Semantic Storage : A Report on Performance and Flexibility. Lecture Notes in Computer Science 3588, 586-595 (2005)

Thank you for your attention.

»Wissen schafft Brücken.«