Extending the DelosDLMS by the FAST Annotation Service

Maristella Agosti[*], Gert Brettlecker[§], Nicola Ferro[*], Paola Ranaldi[§‡], and Heiko Schuldt[§]

[*]Department of Information Engineering – University of Padua – Italy
{agosti, ferro}@dei.unipd.it

[§]Department of Computer Science – University of Basel – Switzerland
{gert.brettlecker, heiko.schuldt}@unibas.ch

[‡] Department of Computer and Information Science – University of Konstanz – Germany
paola.ranaldi@unibas.ch

I. Introduction

DelosDLMS [9] is a prototype of a next-generation *Digital Library Management System (DLMS)*, jointly developed by partners of the EU-funded project DELOS[1] (a Network of Excellence on Digital Libraries). The goal of DelosDLMS is to combine text and audio-visual searching, to offer personalized browsing using new information visualization and relevance feedback tools, to allow retrieved information to be annotated and processed, to integrate and process sensor data streams, and finally, from a systems engineering point of view, to be easily configured and adapted while being reliable and scalable.

The DelosDLMS prototype is currently being built by integrating digital library functionality provided by DELOS and non-DELOS partners into the OSIRIS/ISIS platform, a middleware environment developed by ETH Zürich and now being extended at the University of Basel. *OSIRIS (Open Service Infrastructure for Reliable and Integrated process Support)* has been chosen as basis for integration since it follows a service-oriented architecture and thus allows to seamlessly add more functionality which is provided behind a (Web) service interface. *ISIS (Interactive SImilarity Search)* consists of a set of DL services that are built on top of OSIRIS. The ISIS services provide content-based retrieval of images, audio and video content, and the combination of any of these media types with sophisticated text retrieval.

From a content point of view, DelosDLMS has been extended in first integration activities by support for content-based retrieval of 3D objects[2] and advanced audio features[3]. Moreover, a new visualization by a *Self-Organizing Map (SOM)* representation of high-dimensional feature spaces[4] has been added and both Daffodil[5] and an interactive paper interface[6] allow for advanced interaction with DelosDLMS. In this paper, we report on recent integration activities which extend DelosDLMS by the *Flexible Annotation Service Tool (FAST)* annotation service.

Annotations are not only a way of explaining and enriching an information resource with personal observations, but also a means of transmitting and sharing ideas to improve collaborative work practices. Furthermore, annotations allow users to naturally merge and link personal contents with the information resources provided by a DLMS. For example, instead of simply downloading a paper and then working on a printed version, users could work directly with the paper by means of the tools provided by the DLMS and share their work with colleagues. As a consequence, the DLMS is no longer perceived as something external to the intellectual production process, nor as a mere consulting tool, but instead as an intrinsic and active part of the intellectual production process.

A relevant aspect of annotations is that they can take a part of a hypertext [1] since they allow the creation of new relationships among existing objects, by means of links that connect annotations together with existing objects. In addition, the hypertext between annotations and annotated objects can be exploited not only for providing alternative navigation and browsing capabilities, but also for offering advanced search functionalities, able to retrieve more and better ranked objects in response to a user query by also exploiting the annotations linked to them [2], [3]. Moreover, DLMSs usually offer some basic hypertext and browsing capabilities based on the available structured data, such as authors or references. On the other hand, DLMSs do not normally provide users with advanced hypertext functionalities, where the information resources are linked on the basis of the semantics of their content and hypertext information retrieval functionalities are available. Therefore, annotations can turn out to be an effective way of associating this kind of hypertext to a DLMS to enable the active and dynamic use of information resources. In addition, this hypertext can span and cross the boundaries of the single DLMS, if users need to interact with the information resources managed by diverse DLMSs [4]. This latter possibility is quite innovative, because it offers the means for interconnecting various DLMSs in a personalized and meaningful way for the

[1] http://www.delos.info
[2] http://viplab.dsi.unifi.it/research/
[3] http://www.ifs.tuwien.ac.at/mir/
[4] http://infovis.uni-konstanz.de/
[5] http://www.daffodil.de
[6] http://www.globis.ethz.ch/research/paper

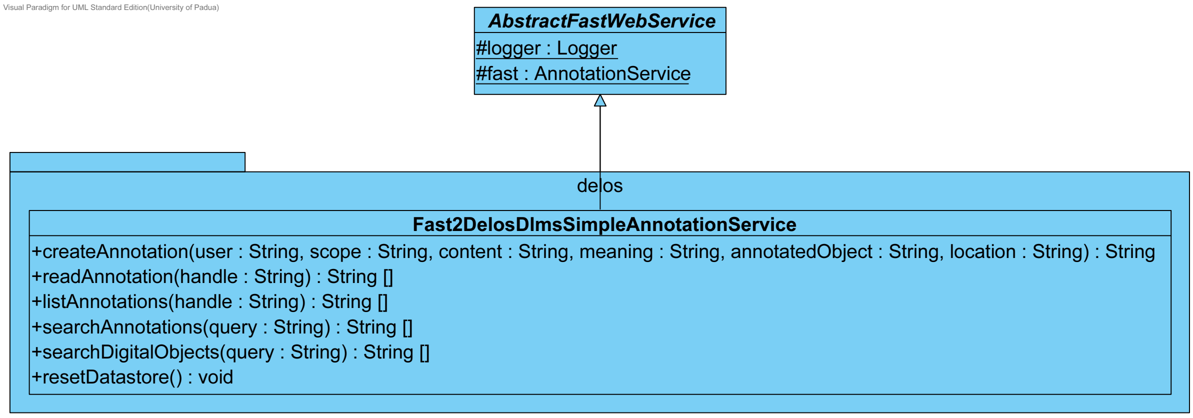

Visual Paradigm for UML Standard Edition(University of Padua)
AbstractFastWebService
#logger : Logger
#fast : AnnotationService
delos
Fast2DelosDlmsSimpleAnnotationService
+createAnnotation(user : String, scope : String, content : String, meaning : String, annotatedObject : String, location : String) : String
+readAnnotation(handle : String) : String []
+listAnnotations(handle : String) : String []
+searchAnnotations(query : String) : String []
+searchDigitalObjects(query : String) : String []
+resetDatastore() : void

Fig. 1. UML class diagram of the designed Web service and its prototype implementation.

end-user, and, as also recognized by [7], this is a big challenge for the next generation DLMSs.

We envisage the following scenario, as demonstrator of the first outcomes of the integration of FAST into the DelosDLMS. A user can annotate an object shown in the DelosDLMS user interface, and he can retrieve the annotations on a given object. Then, the user can perform a keyword search based on annotations. In this case, the annotations that match the query are selected and, starting from those annotations, a list of annotated documents is identified. These annotated documents can be used to filter out the result list produced with other kinds of search, such as an image search.

The paper is organized as follows: Section II provides an overview of the OSIRIS/ISIS architecture; Section III introduces the FAST annotation service; Section IV describes the activities conducted to integrate FAST into the DelosDLMS; finally, Section V draws some conclusions and provides an outlook for future work.

II. OSIRIS/ISIS Architecture

OSIRIS [11], [12] is a middleware that supports the combination of different distributed services into processes (this is also known as ’programming-in-the-large’) and the reliable execution of these processes following the model of transactional processes [10]. When different specialized digital library application services are made available to OSIRIS, users can define and run powerful digital library processes.

In addition to transactional guarantees and reliability, OSIRIS features a high degree of scalability. The decentralized peer-to-peer approach for process execution in OSIRIS, which is realized by sophisticated replication mechanisms for control flow dependencies, avoids any single point of failure during process execution. Peer-to-peer process execution also incorporates sophisticated load balancing in order to distribute process load among available, suitable peers.

OSIRIS distinguishes between *loosely-coupled* and *tightly-coupled* services. Usually, a part of the distributed OSIRIS middleware (called OSIRIS layer) runs on each host providing application services. These application services are considered as tightly coupled since, in case information on their transactional properties like compensation or re-invocation (retriability) is available, dedicated transactional guarantees for processes can be provided. Loosely-coupled application services are those that have to be called remotely, without a local OSIRIS layer available. The integration/invocation is done via the *Web Services Description Language (WSDL)* for service description and the *Simple Object Access Protocol (SOAP)* for invocation.

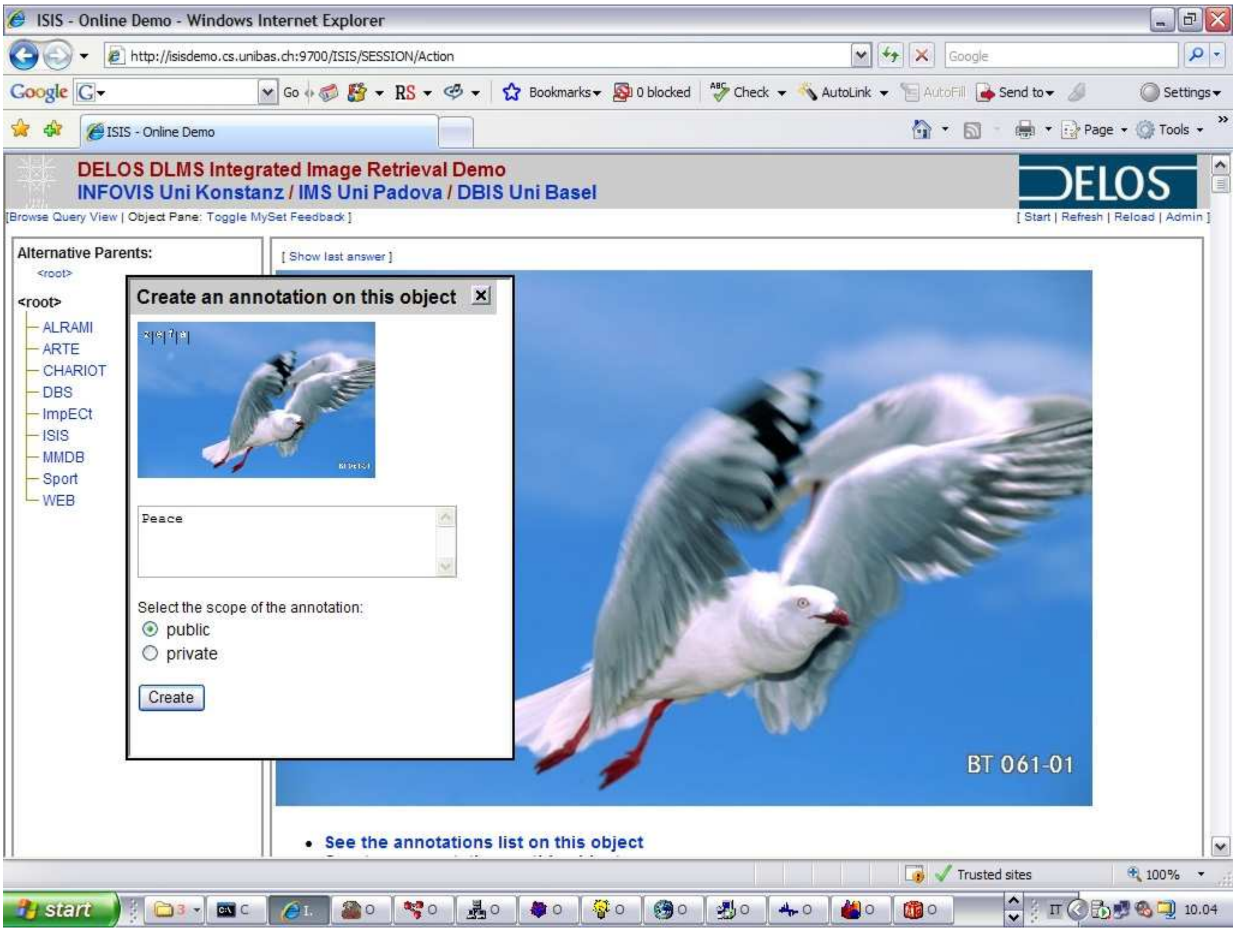
ISIS [8] consists of a set of specialized application services for similarity search which are combined by the OSIRIS middleware. The ISIS services can be easily distributed among several nodes in a network [13]. The implementation of ISIS consequently follows the idea of service-orientation. In particular, the digital library functionality provided by ISIS includes *storage services* for arbitrary types of media objects, *metadata services* for derived and manually generated metadata, *feature extraction services* for given media types (image, audio, and video features), and finally *indexing and search services* that are able to deal with high dimensional feature descriptors [14].

III. FAST Architecture

FAST [4] is a flexible system designed to support both various architectural paradigms and a wide range of different DLMSs. In order to achieve the desired flexibly:

1) FAST is a stand-alone system, i.e., it is not part of any particular DLMS;
2) the core functionalities of the annotation service are separated from the functionalities needed to integrate it into different DLMSs.

The flexibility of FAST and its independence from any particular DLMS is a key feature to provide users with a uniform way of interaction with annotation functionalities, without the need of changing their annotative practices only because a user works with different DLMSs. From an architectural point of view, FAST adopts a three-layers architecture – the data, application and interface logic layers – and is designed at a high level of abstraction in terms of abstract *Application Program Interfaces (APIs)* using an object-oriented approach [4]. In this way, we can model the behaviour and the functioning of FAST without worrying about the actual implementation of each component. Different alternative implementations of



ISIS - Online Demo - Windows Internet Explorer

http://isisdemo.cs.unibas.ch:9700/ISIS/SESSION/Action

DELOS DLMS Integrated Image Retrieval Demo
INFOVIS Uni Konstanz / IMS Uni Padova / DBIS Uni Basel

[Browse Query View | Object Pane: Toggle MySet Feedback]

[Start | Refresh | Reload | Admin]

Alternative Parents:
<root>

<root>
- ALRAMI
- ARTE
- CHARIOT
- DBS
- ImpECt
- ISIS
- MMDB
- Sport
- WEB

[Show last answer]

Create an annotation on this object

Peace

Select the scope of the annotation:
● public
○ private

Create

- See the annotations list on this object

Fig. 2. Pop-up Window for the Creation of Annotations

each component could be provided, still keeping a coherent view of the whole architecture of the FAST system.

Within FAST, annotations are composite multimedia objects, where each part of the annotation, called *sign of annotation*, have a well-defined and explicit semantics, called *meaning of annotation*. Annotations can annotate multiple part of a given digital object and can relate this annotated digital object to various other digital objects, if needed. Furthermore, once it has been created, an annotation is considered as a first class digital object, so that it can be annotated too. In this way, the model supports users in creating not only sets of annotations concerning a digital object, but also threads of annotations, i.e., annotations which reply to one another. These threads of annotations are the basis for actively involving users with the system and for enabling collaboration [5], [1], [6].

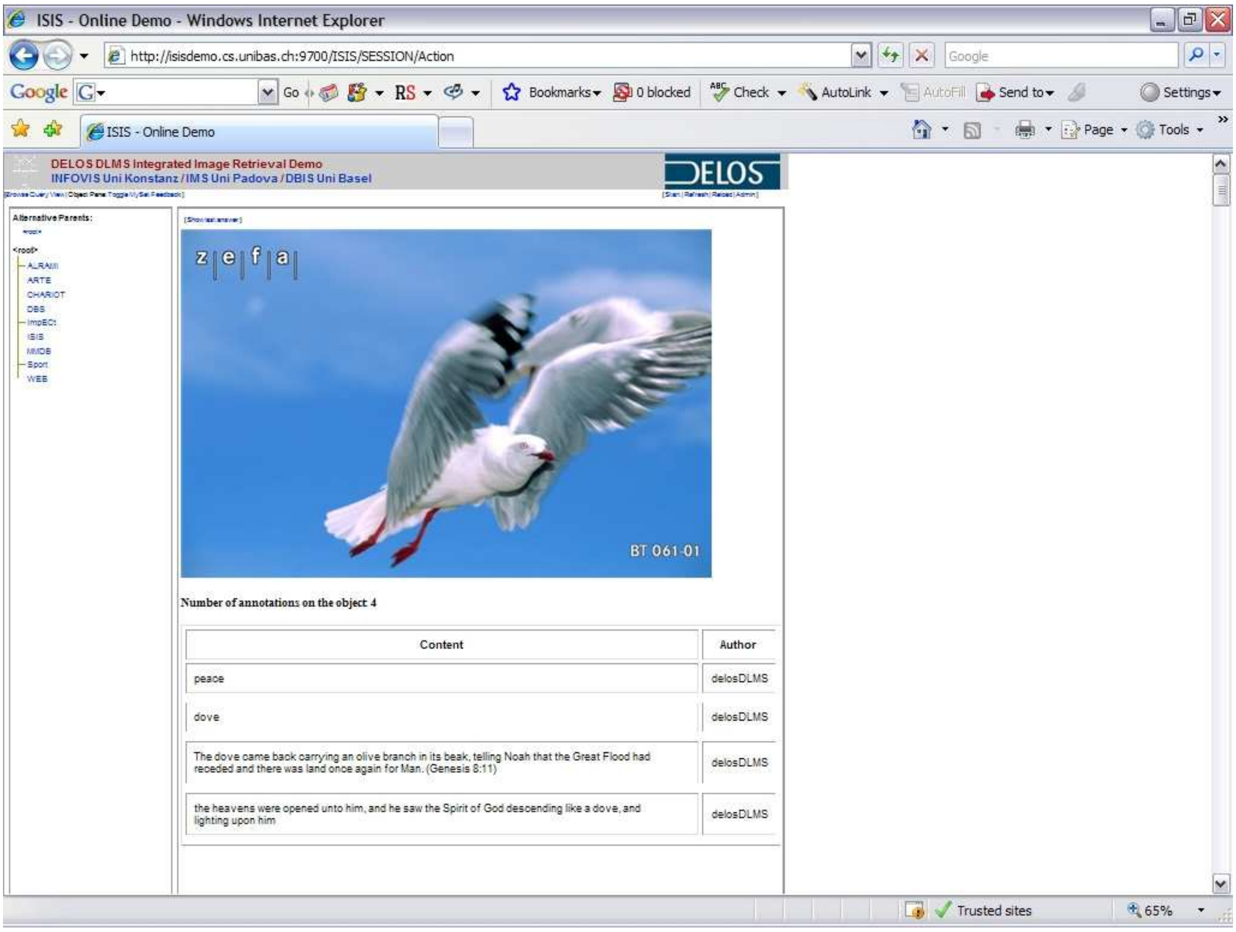
From a functional point of view, FAST provides annotation management functionalities, such as creation, access, and so on. Furthermore it supports collaboration among user by introducing scopes of annotation and groups of users: annotations can be private, shared or public; if an annotation is shared, different groups of users can share it with different permissions, e.g., one group can only read the annotation while another can also modify it. Note that the annotation management engine ensure that some validity constraints are complied with: for example, a private annotation cannot be annotated by a public annotation. In such cases there is a *scope conflict* – in the example, the author of the private annotation could see both the public and the private annotation, but another user could see only the public annotation which would be annotating something hidden to this user. Finally, FAST offers advanced search functionalities based on annotations by exploiting annotations as a useful context in order to search and retrieve relevant documents for a user query [2]. The aim is to retrieve more documents that are relevant and to have them ranked in a way which is better than a system that does not makes use of annotations.

IV. Integration Activities

A. FAST Web Service

In order to integrate the FAST annotation service into the DelosDLMS we follow a *loosely-coupled* integration strategy, making FAST accessible to the DelosDLMS as a Web service.

In particular, we selected a subset of the functionalities exposed by FAST at the application layer and wrapped them into a Web service, which has been called `Fast2DelosDlmsSimpleAnnotationService` since it provides the basic functionalities needed to carry out the scenario described in Section I.



DELOS DLMS Integrated Image Retrieval Demo
INFOVIS Uni Konstanz / IMS Uni Padova / DBIS Uni Basel

Number of annotations on the object 4

Content	Author
peace	delosDLMS
dove	delosDLMS
The dove came back carrying an olive branch in its beak, telling Noah that the Great Flood had receded and there was land once again for Man. (Genesis 8:11)	delosDLMS
the heavens were opened unto him, and he saw the Spirit of God descending like a dove, and lighting upon him	delosDLMS

Fig. 3. Object Annotation List Window

The functionalities made available by the developed Web service are:

- `createAnnotation`: creates a new annotation, assuming that the annotation is constituted by only one textual sign and can be either public or private. In addition, a specific location of the annotated digital object can be specified, e.g. the upper left corner of an image;
- `readAnnotation`: reads an existing annotation with all the related information;
- `listAnnotations`: returns a list of the annotation identifiers on a given digital object.;
- `searchAnnotations`: performs a keyword-based search on the textual content of the annotations;
- `searchDigitalObjects`: performs a keyword-based search for digital objects on the basis of the content of their annotations by exploiting also hypertext between digital objects and annotations;
- `resetDatastore`: completely resets the FAST datastore and its use is limited to the testing phase of the integration.

Figure 1 depicts the *Unified Modeling Language (UML)* class diagram of the designed Web service where the functionalities made available are shown together with the input and output parameters. Note that `Fast2DelosDlmsSimpleAnnotationService` is a concrete subclass of an abstract class, called `AbstractFastWebService`, which actually wraps the FAST application logic layer in order to make it available to all its implementing subclasses. Furthermore, `AbstractFastWebService` provides its subclasses the means for accessing the logging infrastructure built in into FAST in order to provide detailed information about their functioning. The `Fast2DelosDlmsSimpleAnnotationService` has been implemented in Java[7] by using the Apache Axis[8] implementation of the SOAP protocol with the RPC/encoded binding style.

B. Extensions to ISIS

ISIS has been extended to make use of the FAST annotation services. In order to call a Web service, the WSDL description has to be registered at the DelosDLMS SOAP Gateway. After this registration, the loosely coupled SOAP services are available for use by all nodes of the DelosDLMS infrastructure. The ISIS Session component is offering a Web based user interface,

[7] `http://java.sun.com/`

[8] `http://ws.apache.org/axis/`

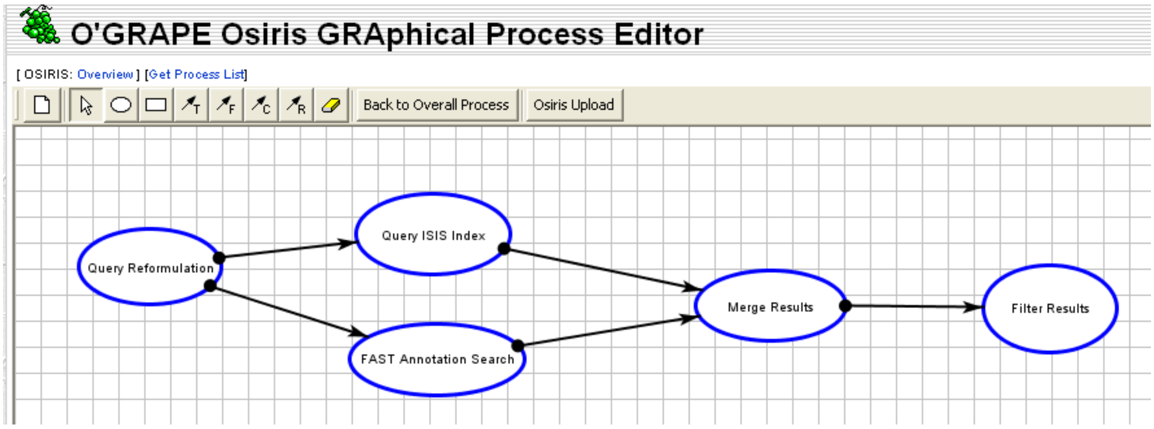


Fig. 4. DelosDLMS Extended Annotation-Aware Search Process used by ISIS

which has been extended to allow for creating, searching, and viewing of annotations on digital library objects.

The Session component is receiving user requests coming from the Web browser via the *HyperText Transfer Protocol (HTTP)* and converting them into requests for the DelosDLMS infrastructure.

In order to create a new annotation for some object, the user has to click on a link which is added to the ISIS page that displays the object. From the pop-up window the user can now edit her own annotation and select whether it should be a public or a private annotation (see Figure 2). Once the new annotation is created all the annotations on the object will be displayed (Figure 3) when the object is selected. The list of the annotations on the object can be displayed also clicking the *see the annotation list on the object* link, above the *create annotation* link.

If a user issues a search, the Session component is in charge of starting a DelosDLMS process executing the query tasks. This query process has been adapted in order to allow also for searching within annotations. If the user has specified keywords for the search, it is possible to query also the FAST annotation service. This query option has been integrated into the user interface. In particular, when opening the content options, the user can decide to ask either for a content-based search, or for the annotation search, or for both. Fig. 4 illustrates this extended query process, which offers search in annotations as a parallel execution branch. Once the *Query ISIS Index* and the *FAST annotation search* activities are completed, the results have to be merged by an additional *Merge Result* step. The *Query Reformulation* and *Filter Results* steps will allow for relevance feedback over the overall query process.

Finally, the results are presented to the user via the session component. Internally, this is done by using *XSL Transformations (XSLT)* stylesheets for the conversion of the DelosDLMS result message into an *HyperText Markup Language (HTML)* page for the browser. For the combined ISIS and annotation search the Session component will display the result objects differently depending on whether they are coming from content based ISIS search or from annotation search (Fig. 5).

V. Conclusions

Digital libraries are no longer systems that just provide access to static content. Rather, users demand support for enriching and personalizing content in a collaborative environment. Therefore, annotation management plays a vital role in future Digital Libraries.

In this paper, we have presented the DelosDLMS system which has been extended by the FAST annotation service. Currently, this integration is realized in a loosely-coupled way by using Web services. Among others, future activities will aim at making user management hosted in the annotation service available to DelosDLMS and at closely linking annotation management and personalization services.

Acknowledgments

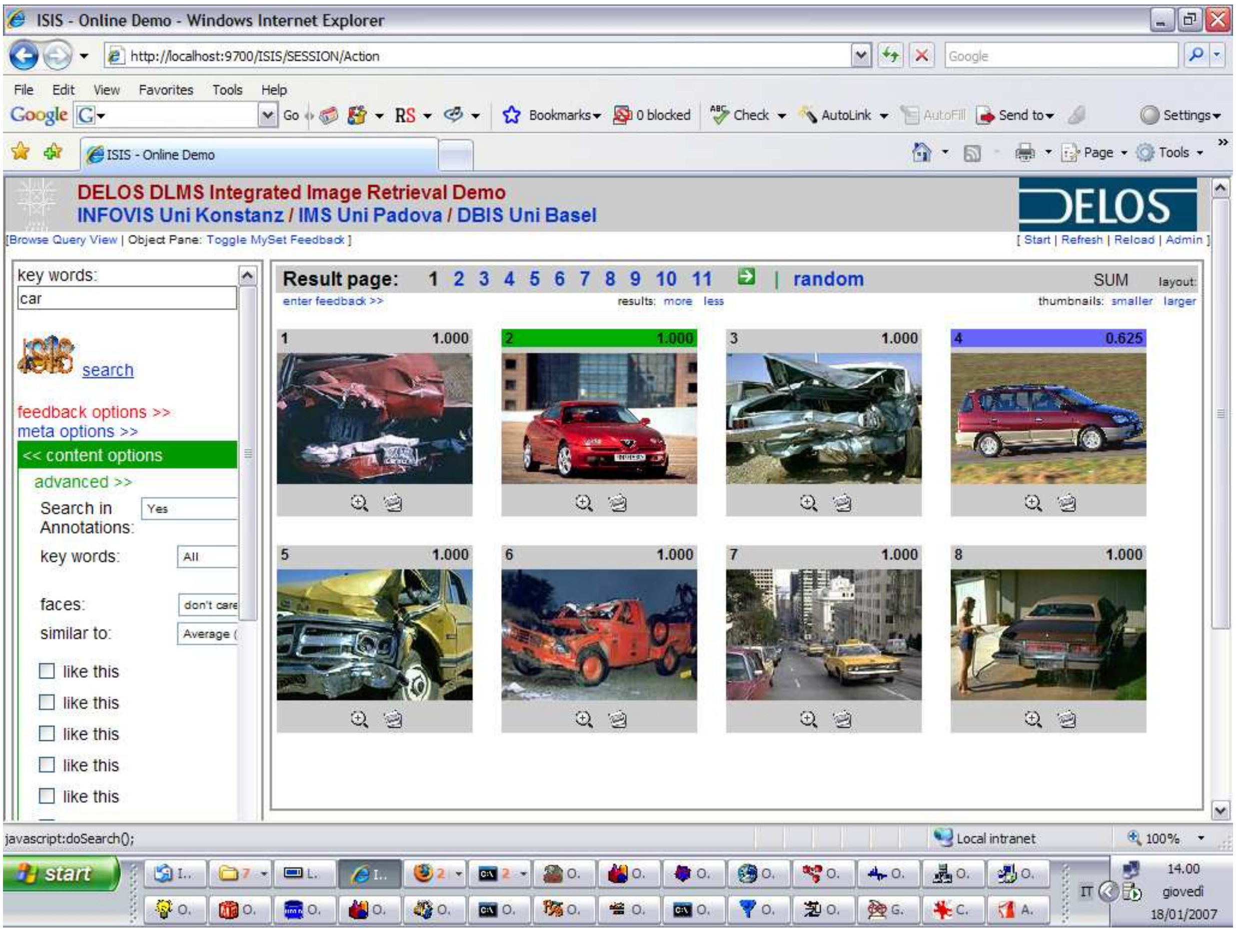
The work was partially supported by the DELOS Network of Excellence on Digital Libraries, as part of the Information Society Technologies (IST) Program of the European Commission (Contract G038-507618).

References

[1] M. Agosti, N. Ferro, I. Frommholz, and U. Thiel, “Annotations in Digital Libraries and Collaboratories – Facets, Models and Usage,” in *Proc. 8th European Conference on Research and Advanced Technology for Digital Libraries (ECDL 2004)*, R. Heery and L. Lyon, Eds. Lecture Notes in Computer Science (LNCS) 3232, Springer, Heidelberg, Germany, 2004, pp. 244–255.

[2] M. Agosti and N. Ferro, “Annotations as Context for Searching Documents,” in *Proc. 5th International Conference on Conceptions of Library and Information Science – Context: nature, impact and role*, F. Crestani and I. Ruthven, Eds. Lecture Notes in Computer Science (LNCS) 3507, Springer, Heidelberg, Germany, 2005, pp. 155–170.

[3] M. Agosti and N. Ferro, “Search Strategies for Finding Annotations and Annotated Documents: the FAST Service,” in *Proc. 7th International Conference on Flexible Query Answering Systems (FQAS 2006)*, H. Legind Larsen, G. Pasi, D. Ortiz-Arroyo, T. Andreasen, and H. Christiansen, Eds. Lecture Notes in Artificial Intelligence (LNAI) 4027, Springer, Heidelberg, Germany, 2006, pp. 270–281.



ISIS - Online Demo - Windows Internet Explorer

http://localhost:9700/ISIS/SESSION/Action

DELOS DLMS Integrated Image Retrieval Demo

INFOVIS Uni Konstanz / IMS Uni Padova / DBIS Uni Basel

key words: car

Search in Annotations: Yes

key words: All

faces: don't care

similar to: Average (

Result page: 1 2 3 4 5 6 7 8 9 10 11 | random

1	1.000
2	1.000
3	1.000
4	0.625
5	1.000
6	1.000
7	1.000
8	1.000

javascript:doSearch();

Fig. 5. Objects Resulting from the Combined Annotation and Content-based Searches: the objects marked in green are the ones retrieved from the annotation search; the blue ones are retrieved from both the content-based search and the annotation-based one.

[4] M. Agosti and N. Ferro, "A System Architecture as a Support to a Flexible Annotation Service," in *Peer-to-Peer, Grid, and Service-Orientation in Digital Library Architectures: 6th Thematic Workshop of the EU Network of Excellence DELOS. Revised Selected Papers*, C. Türker, M. Agosti, and H.-J. Schek, Eds. Lecture Notes in Computer Science (LNCS) 3664, Springer, Heidelberg, Germany, 2005, pp. 147–166.

[5] M. Agosti and N. Ferro, "Annotations: Enriching a Digital Library," in *Proc. 7th European Conference on Research and Advanced Technology for Digital Libraries (ECDL 2003)*, T. Koch and I. T. Sølvberg, Eds. Lecture Notes in Computer Science (LNCS) 2769, Springer, Heidelberg, Germany, 2003, pp. 88–100.

[6] M. Agosti, N. Ferro, and N. Orio, "Annotating Illuminated Manuscripts: an Effective Tool for Research and Education," in *Proc. 5th ACM/IEEE-CS Joint Conference on Digital Libraries (JCDL 2005)*, M. Marlino, T. Sumner, and F. Shipman, Eds. ACM Press, New York, USA, 2005, pp. 121–130.

[7] Y. Ioannidis, D. Maier, S. Abiteboul, P. Buneman, S. Davidson, E. A. Fox, A. Halevy, C. Knoblock, F. Rabitti, H.-J. Schek, and G. Weikum, "Digital library information-technology infrastructures," *International Journal on Digital Libraries*, vol. 5, no. 4, pp. 266–274, 2005.

[8] M. Mlivoncic, C. Schuler, and C. Türker. Hyperdatabase Infrastructure for Management and Search of Multimedia Collections. In *DELOS Workshop: Digital Library Architectures*, pages 25–36, S. Margherita di Pula, Cagliari, Italy, 2004.

[9] H.-J. Schek and H. Schuldt. DelosDLMS - Infrastructure for the Next Generation of Digital Library Management Systems. *ERCIM News, Special Issue on the European Digital Library*, No. 66:22–24, July 2006.

[10] H. Schuldt, G. Alonso, C. Beeri, and H.-J. Schek. Atomicity and Isolation in Transactional Processes. *ACM Transactions on Database Systems*, 27(1):63–116, March 2002.

[11] C. Schuler, H. Schuldt, C. Türker, R. Weber and H.-J. Schek, Peer-to-Peer Execution of (Transactional) Processes. *International Journal of Cooperative Information Systems (IJCIS)*, 14 (2005) 377–405

[12] C. Schuler, C. Türker, H.-J. Schek, R. Weber and H. Schuldt, Scalable Peer-to-Peer Process Management. *International Journal of Business Process Integration and Management (IJBPIM)*, vol. 1, no. 2, pp. 129–142, 2006.

[13] R. Weber, J. Bolliger, T.R. Gross, and H.-J. Schek. Architecture of a Networked Image Search and Retrieval System. In *Proc. of the CIKM Conference*, pages 430–441, Kansas City, Missouri, USA, 1999.

[14] R. Weber, H.-J. Schek, and S. Blott. A Quantitative Analysis and Performance Study for Similarity-Search Methods in High-Dimensional Spaces. In *Proc. of the 24th Int. Conf. on Very Large Data Bases (VLDB98)*, pages 194-205, New York City, USA, 1998.