

Scientific Evaluation of a DLMS: A Service for Evaluating Information Access Components

Giorgio Maria Di Nunzio and Nicola Ferro

Department of Information Engineering – University of Padua
Via Gradenigo, 6/b – 35131 Padova – Italy
{dinunzio, ferro}@dei.unipd.it

Abstract. In this paper, we propose an architecture for a service able to manage, enrich, and support the interpretation of the scientific data produced during the evaluation of information access and extraction components of a *Digital Library Management System (DLMS)*. Moreover, we describe a first prototype, which implements the proposed service.

1 Introduction

As observed in [4], “*Digital Library (DL)* development must move *from an art to a science*” in order to design and develop *Digital Library Management Systems (DLMSs)*, based on reliable and extensible services. This shift from DLs to service-based DLMSs and the requirement for improved reliability points out, among other issues, the need of proper evaluation methodologies in order to assess a DLMS along different dimensions.

The evaluation itself of a DLMS turns out to be a scientific activity whose outcomes, such as performance analyses and measurements, constitute a kind of *scientific data* that need to be properly considered and used for the design and development of DLMS components and services. These scientific data should, in turn, be managed by a DLMS which takes care of supporting their enrichment and interpretation. We propose to name this type of DLMS a *scientific reflection DLMS*, since it deals with scientific data, information, and interpretations about the design and development of another DLMS.

There are many aspects to take into consideration when evaluating a DLMS, such as information access capabilities, interaction with users, and so on. As a consequence, the scientific reflection DLMS should be constituted by different and cooperating services, each one focused on supporting the evaluation of one of the aspects mentioned above. In particular, we face the problem of designing and developing a service for evaluating the information access components of a DLMS.

2 Design of an Information Access Evaluation Service for a Scientific Reflection DLMS

Figure 1 shows the architecture of the proposed service. It consists of three layers – data, application and interface logic layers – in order to achieve a better

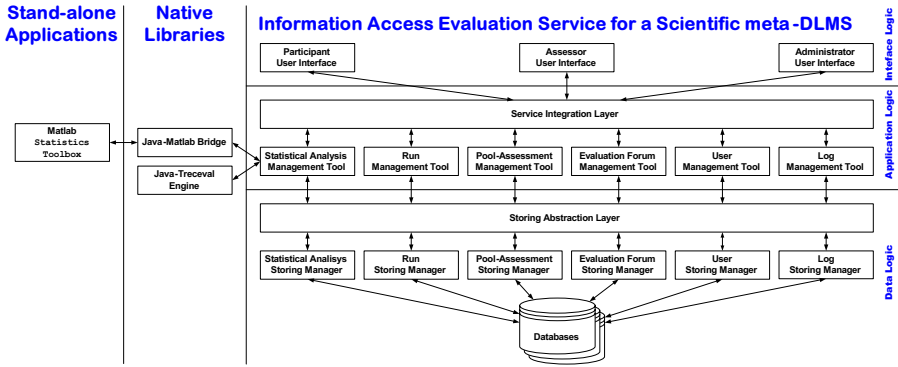


Fig. 1. Architecture of a service for supporting the evaluation of the information access components of a DLMS

modularity and to properly describe the behavior of the service by isolating specific functionalities at the proper layer.

Data Logic. The data logic layer deals with the persistence of the different information objects coming from the upper layers. There is a set of “storing managers” dedicated to storing the submitted experiments, the relevance assessments and so on. We adopt the *Data Access Object (DAO)*¹ and the *Transfer Object (TO)*¹ design patterns. The DAO implements the access mechanism required to work with the underlying data source, acting as an adapter between the upper layers and the data source.

In addition to the other storing managers, there is the *log storing manager* which fine traces both system and user events. It captures information such as the user name, the *Internet Protocol (IP)* address of the connecting host, the action that has been invoked by the user, any error condition, and so on. Thus, besides offering us a log of the system and user activities, the log storing manager allows us to fine trace the provenance of each piece of data from its entrance in the system to every further processing on it.

Finally, on top of the various “storing managers” there is the *Storing Abstraction Layer (SAL)* which hides the details about the storage management to the upper layers. In this way, the addition of a new “storing manager” is totally transparent for the upper layers.

Application Logic. The application logic layer deals with the flow of operations within *Distributed Information Retrieval Evaluation Campaign Tool (DIRECT)*. It provides a set of tools capable of managing high-level tasks, such as experiment submission, pool assessment, statistical analysis of an experiment.

For example, the *Statistical Analysis Management Tool (SAMT)* offers the functionalities needed to conduct a statistical analysis on a set of experiments.

¹ <http://java.sun.com/blueprints/corej2eepatterns/Patterns/>

In order to ensure comparability and reliability, the SAMT makes use of well-known and widely used tools to implement the statistical tests, so that everyone can replicate the same test, even if he has no access to the service. In the architecture, the MATLAB Statistics Toolbox² has been adopted, and an additional library has been implemented to allow our service to access MATLAB in a programmatic way. As an additional example aimed at wide comparability and acceptance of the tools, a further library provides an interface for our service towards the `trec_eval` package³, which represents the standard tool for computing the basic performance figures, such as precision and recall.

Finally, the *Service Integration Layer (SIL)* provides the interface logic layer with a uniform and integrated access to the various tools. As we noticed in the case of the SAL, thanks to the SIL also the addition of new tools is transparent for the interface logic layer.

Interface Logic. It is the highest level of the architecture, and it is the access point for the user to interact with the system. It provides specialised *User Interfaces (UIs)* for different types of users, that are the participants, the assessors, and the administrators. Note that, thanks to the abstraction provided by the application logic layer, different kind of UIs can be provided, either stand-alone applications or Web-based applications.

3 The First Running Prototype

The proposed service has been implemented in a first prototype, called *Distributed Information Retrieval Evaluation Campaign Tool (DIRECT)* [2], and it has been tested in the context of the *Cross-Language Evaluation Forum (CLEF)* 2005 evaluation campaign. The initial prototype moves a first step in the direction of an information access evaluation service for scientific reflection DLMSs, by providing support for: the management of an evaluation forum: the track set-up, the harvesting of documents, the management of the subscription of participants to tracks; the management of submission of experiments, the collection of metadata about experiments, and their validation; the creation of document pools and the management of relevance assessment; common statistical analysis tools for allowing the comparison of the experiments; common tools for summarizing, producing reports and graphs on the measured performances and conducted analyses.

The prototype was successfully adopted during the CLEF 2005 campaign. It was used by nearly 30 participants spread over 15 different nations, who submitted more than 530 experiments; then 15 assessors assessed more than 160,000 documents in seven different languages, including Russian and Bulgarian which do not have a latin alphabet. It was then used for producing reports and overview graphs about the submitted experiments [1,3].

² <http://www.mathworks.com/products/statistics/>

³ <ftp://ftp.cs.cornell.edu/pub/smart/>

DIRECT has been developed by using the Java⁴ programming language. We used the PostgreSQL⁵ *DataBase Management System (DBMS)* for performing the actual storage of the data. Finally, a Web-based interface, which make the service easily accessible to end-users without the need of installing any kind of software, has been developed by using the Apache STRUTS⁶ framework.

4 Conclusions

We proposed the architecture of a service which evaluates the information access components of a DLMS. This service should be part of a wider *scientific reflection DLMS*, which allows for enriching and interpreting the scientific data produced during the evaluation of a DLMS. A first prototype of the proposed service has been implemented and widely tested during the CLEF 2005 evaluation campaign.

Acknowledgements

The authors would like to thank Carol Peters, coordinator of CLEF, for fruitful discussions in the context of the CLEF effort. The work reported in this paper has been 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. G. M. Di Nunzio and N. Ferro. Appendix A. Results of the Core Tracks and Domain-Specific Tracks. In *Working Notes for the CLEF 2005 Workshop*. http://www.clef-campaign.org/2005/working_notes/workingnotes2005/appendix_a.pdf, 2005.
2. G. M. Di Nunzio and N. Ferro. DIRECT: a System for Evaluating Information Access Components of Digital Libraries. In *Proc. 9th European Conference on Research and Advanced Technology for Digital Libraries (ECDL 2005)*, pages 483–484. LNCS 3652, Springer, Heidelberg, Germany, 2005.
3. G. M. Di Nunzio, N. Ferro, G. J. F. Jones, and C. Peters. CLEF 2005: Ad Hoc Track Overview. In *Accessing Multilingual Information Repositories: Sixth Workshop of the Cross-Language Evaluation Forum (CLEF 2005). Revised Selected Papers*. LNCS 4022, Springer, Heidelberg, Germany (in print), 2006.
4. 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*, 5(4):266–274, 2005.

⁴ <http://java.sun.com/>

⁵ <http://www.postgresql.org/>

⁶ <http://struts.apache.org/>