Design of the User Interface of a Scientific Digital Library System for Large-Scale Evaluation Campaigns

Marco Dussin and Nicola Ferro
Department of Information Engineering, University of Padua, Italy
{dussinma, ferro}@dei.unipd.it

Abstract

This paper describes the design and development of an effective user interface for a digital library system which manages the different types of information resource produced during the course of a large-scale evaluation campaign. The paper discusses a conceptual model for the interface, presents its architecture, and shows some examples of the actual interface implemented in the prototype system we have developed.

Categories and Subject Descriptors

H.3 [Information Storage and Retrieval]: H.3.4 Systems and Software; H.3.7 Digital Libraries;

General Terms

Measurement, Performance, Experimentation

Keywords

Experimental Evaluation, Scientific Data, Digital Library System, User Interface, Evaluation

1 Introduction

Large-scale evaluation campaigns for Information Retrieval (IR) promote and stimulate the research and development of Information Retrieval Systems (IRS) by:

- creating an evaluation infrastructure and organising regular evaluation campaigns for system testing where ideas can be exchanged and different approaches can be discussed;
- building a strong multidisciplinary research community where problems are faced from different points of view – e.g. information retrieval, question answering, natural language processing – and multiple techniques are merged and harmonized together;
- constructing publicly available test-suites which can also be used outside the evaluation campaigns for system benchmarking.

Furthermore, large-scale evaluation campaigns aim at anticipating the emerging needs of the research community by designing tasks which stimulate research in appropriate directions.

Finally, large-scale evaluation campaigns impact not only the IR field but also other research fields which adopt and apply results from it, such as the Digital Library (DL) field. Indeed, information access and extraction components of a Digital Library System (DLS), which deal with the indexing, search and retrieval of documents in response to a user's query, rely on methods and techniques taken from the IR field. In this context, large-scale evaluation campaigns provide qualitative and quantitative evidence along the years as to which methods give the best results in certain key areas, such as indexing techniques, relevance feedback, multilingual querying, results merging, and contribute to the overall problem of evaluating a DLS (Fuhr, et al., 2007).

(Agosti, et al., 2007) point out that "the experimental data produced during an evaluation campaign are valuable scientific data, and as a consequence, should be archived, enriched, and curated in order to ensure their future accessibility and re-use". In this perspective, (Agosti, Di Nunzio, & Ferro, 2007a, b) have revised the current methodology for conducting IR evaluation campaigns and have highlighted that it has to extended to support the curation and enrichment of the produced scientific data from their creation up to their re-use by researchers outside the evaluation campaign. To this end, (Agosti, Di Nunzio, & Ferro, 2007c) highlight as a DLS is "the

natural choice for managing, making accessible, citing, curating, enriching, and preserving all the information resources produced during an evaluation campaign" since it provides a more mature way of dealing with the scientific data produced during the IR experimental evaluation.

This paper discusses the design and development of an effective user interface for a scientific DLS able to support the course of an IR evaluation campaign and to assist the different actors involved in an evaluation campaign to perform their own tasks. In this context, we have developed our prototype system, called Distributed Information Retrieval Evaluation Campaign Tool (DIRECT), which has been adopted and tested in the Cross-Language Evaluation Forum (CLEF)[1] campaigns from 2005 to 2007 (Di Nunzio & Ferro, 2005, 2006). The paper presents how the user interface has been designed in the light of the proposed extension to the IR evaluation methodology and describes how it has evolved to satisfy the needs of the users and to offer them new features and an improved usability by treasuring their feedback.

The paper is organized as follows: Section 2 introduces the conceptual model that is at the basis of the design of the user interface; Section 3 discusses the architecture of the user interface; Section 4 presents some examples of interface as implemented in the DIRECT running prototype; finally, Section 5 draws some conclusions.

2 Exploiting the DIKW Hierarchy to Design the User Interface

The current approach used to evaluate information access systems relies on the Cranfield methodology, which makes use of *experimental collections* that consist of a set of *documents*, a set of *topics* expressing user's information needs and from which the actual queries are derived, and a set of *relevance judgments* that determine the relevant documents for each topic. An experimental collection allows the comparison of two information access systems according to some measurements which quantify their performances, and provides a common test-bed in order to guarantee the possibility of replicating the experiments (Cleverdon, 1997).

Nevertheless, the Cranfield methodology is generally mainly focused on creating comparable experiments and evaluating the performances of an IRS rather than modeling and managing the data produced during the evaluation campaign. On the other hand, the curation, enrichment, interpretation, and citation of scientific data are essential components of scientific research: researchers greatly benefit from an integrated vision of the produced scientific data, their analyses and their interpretations, and from the possibility of keeping, re-using, and enriching them with further analyses and interpretations (Agosti, Di Nunzio, & Ferro, 2007b).

(Agosti, Di Nunzio, & Ferro, 2007a, c) have framed the problem of extending the current evaluation methodology in the context of the Data, Information, Knowledge, Wisdom (DIKW) hierarchy (Ackoff, 1989; Zeleny, 1987):

- *data*: the *test collections* and the *experiments* correspond to the "data level" in the hierarchy, since they are the raw, basic elements needed for any further investigation and have little meaning by themselves;
- *information*: the *performance measurements* correspond to the "information level" in the hierarchy, since they are the result of computations and processing on the data;
- *knowledge*: the *descriptive statistics* and the *hypothesis tests* correspond to the "knowledge level" in the hierarchy, since they represent further processing of the information provided by the performance measurements and provide us with some insights about the experiments;
- *wisdom*: *theories*, *models*, *algorithms*, *techniques*, and *observations*, which are usually communicated by means of papers, talks, and seminars, correspond to the "wisdom level" in the hierarchy, since they provide interpretation, explanation, and formalization of the content of the previous levels.

Furthermore, as recognized by (Lord & Macdonald, 2003), the possibility of citing scientific data and their further elaboration is an effective way of making scientists and researchers an active part of the digital curation process. Indeed, this opportunity would strengthen the passing from data to wisdom because experimental collections and experiments would become citable and accessible

[1] http://www.clef-campaign.org/

just like any other item in the reference list of a paper.

Different types of actors are involved in an evaluation campaign:

- the *participant* takes part in the evaluation campaign in order to have a forum to test his new algorithms and techniques, to compare their effectiveness, and to discuss and share his proposals. He needs support for the submission of his experiments and their validation; he then expects to receive measurements about the performances of his experiments and overall indicators that allows for the comparison of his experiments with the ones submitted by other participants. Moreover, he should have the possibility of properly citing his experiments and other information resources and to get a citation correctly resolved to the corresponding information resources;
- the *assessor* contributes to the creation of the experimental collections by both proposing the topics and assessing the relevance of the documents with respect to those topics. He needs support in both these tasks which are labour-intensive and require the inspection of great amounts of data;
- the *visitor* needs to consult, browse and access all the information resources produced during the course of an evaluation campaign in a meaningful fashion which provides him insights about the conducted experiments. Moreover, he should have the possibility of properly citing the accessed information resources and to get a citation correctly resolved to the corresponding information resources;
- the *organizer* manages the different aspects of an evaluation forum: he contributes to the creation of the experimental collections by preparing the documents and overseeing the creation of the topics and the relevance assessments; he provides the framework for the participants to conduct their experiments and for the assessors to create the topics and perform the relevance assessments; he computes the different measures for assessing the performances of the submitted experiments as well as descriptive statistics and statistical tests to characterize the overall features of the submitted experiments; finally, he provides the visitors with the means for accessing all the information resources they are looking for.

These actors interact together in various ways during the course of an evaluation campaign and contribute differently to the DIKW hierarchy discussed above.

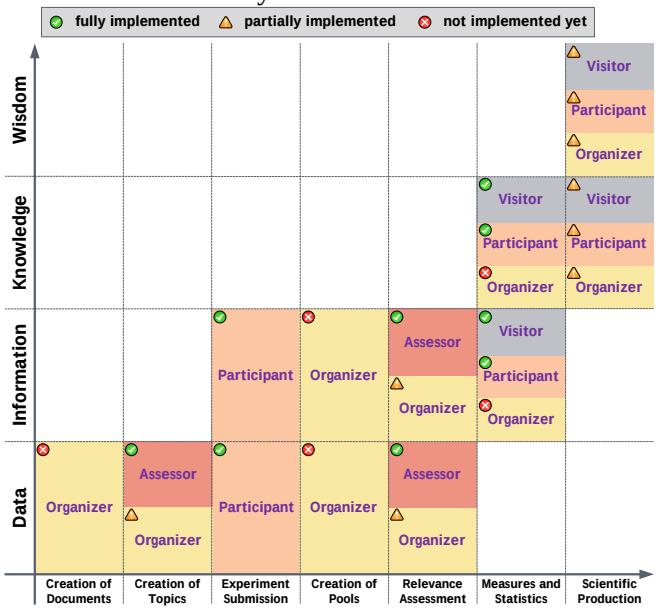

Figure 1: Relationships between the DIKW hierarchy, the different types of actors and the main steps of an evaluation campaign.

Figure 1 summarizes the relationships between the main steps of an evaluation campaign, shown in chronological order on the horizontal axis, the elements of the DIKW hierarchy, shown on the vertical axis, and the main actors involved in an evaluation campaign. In addition, Figure 1 shows the degree to which each step has been implemented in the user interface of the current prototype.

From a glance at Figure 1, it can be noted that the early stages of an evaluation campaign are mainly devoted to the preparation of the data and require limited interaction between the different actors. As time goes on and we enter in the heart of the campaign, there is a progressive movement from data to wisdom, and the number of actors involved and their interaction grows.

We will now examine the elements shown in Figure 1 in more detail:

- *creation of documents*: the organizers are responsible for acquiring and preparing the set of documents which will be released to the participants afterwards. These documents are part of the *data* on which the experiments are built.
  Organizers need an interface that allows them to upload the collections of documents, that can in general be multimedia documents, into the DLS in order to make them available to participants and assessors later on;
- *creation of topics*: the organizers and the assessors cooperate to create the topics for the test collection. For each topic, this step usually requires the preparation of a first draft of the topic and a search for the set of document to verify that there are relevant documents for that topic; then the topic is refined by discussing its content and facets until a final version is reached. These topics are part of the *data* on which the experiments are built.
  Organizers need an interface that allows them to set up the topics to be created, to monitor the creation process, and to commit the topics once they are in the final form.
  Assessors need an interface that allows them to insert and modify the content of a topic, to search the collections of documents to verify that there are relevant documents for the topic, and to discuss together the contents of the topic.
  Note that topics are created by inspecting the documents which are, in a sense, a kind of more basic than the topics. This fact is also reflected in the user interface which needs to support more complex tasks which reflect the relationships between these two kinds of data;
- *experiment submission*: the participants submit their experiments which are built using the documents and the topics created in the previous steps. Each experiment consists of a list of retrieved documents in relevance order for each topic and represents the output of the execution of the IRS developed by the participant. The experiments are part of the *data* which are produced during an evaluation campaign.
  Participants need an interface that allows them to upload their experiments into the DLS, to validate them, e.g. to check that the correct document identifiers have been used or that no topic has been skipped, and to provide all the necessary information for describing their experiments.
  Note that experiments are created by starting from documents and topics and represent, in a sense, a kind of more complex data with respect to them. This fact is also reflected in the user interface which provides support for checking the correctness of the experiments with respect to topics and documents;
- *creation of pools*: the organizers collect all the experiments submitted by the participants and, by using some appropriate sampling technique, select a subset of the retrieved documents to be manually assessed in the next step to determine their actual relevance. The pools are midway between *data* and *information*, since they are still quite raw elements but represent a first form of processing of the experiments.
  Organizers need an interface that allows them to select and sample the documents to be inserted in the pool and to dynamically see how the pools change when the selection criteria are modified in order to determine the best strategy for creating the pools.
  This hybrid nature of the pools between data and information is reflected also in the user interface which explicitly has to show as a pool – i.e. something that relates documents, topics, and experiments – changes when the selection and sampling criteria are modified;
- *relevance assessment*: the organizers and the assessors cooperate to assess each document in the pool with respect to the topic, i.e. to determine whether the document is relevant or not for the given topic. As in the case of the pools, the relevance judgements are midway between *data* and *information*, since they are raw elements which constitute the experimental collection but represent human-added information about the relationship between topics and documents of an experiment.
  Organizers need an interface that allows them to monitor the relevance assessment process and

to commit the relevance judgements once they are in the final form.
Assessors need an interface that allows them to assess the relevance of a document with respect to a topic, to have some basic search functionalities for the documents and topics to assess, and to discuss together in the event of topics which are difficult or ambiguous to assess.
This dual nature of the relevance assessment between data and information is reflected also in the user interface which explicitly request the assessors to enter a human judgement (relevant or not relevant) about the relationship between a document and a topic;

- *measures and statistics*: the organizers exploit the relevance assessments in order to compute the performance measures and plots about each experiment submitted by a participant; then, these measurements are used for computing descriptive statistics about the overall behaviour of both an experiment and all the experiments in a given task; furthermore, these measurements are also employed for conducting statistical analyses and tests on the submitted experiments. As discussed above, performance measures are *information*, since they are the results of a processing on the data; descriptive statistics and hypothesis tests are *knowledge*, since they provide us with some more insights about the meaning of the obtained performances.
  Organizers need an interface that allow them to perform all the computations and statistical analyses that are needed.
  Participants and visitors need an interface which gives access and presents performance measurements, plots, descriptive statistics and statistical analyses in a meaningful way in order to facilitate their comprehension and interpretation.
- *scientific production*: both organizers and participants prepare reports where the former describe the overall trends and provide an overview for the evaluation campaign and the latter explain their experiments, the techniques that have been adopted, and the achieved findings. This work usually continues also after the conclusion of the campaign, since the investigation and understanding of the experimental results require deep analysis and reasoning, which usually takes the form of conference papers, journal articles, talks, and discussion among researchers. Furthermore, not only the organizers and the participants but also external visitors may exploit the information resources produced during the evaluation campaign in order to carry out their research activity. As introduced above, the outcomes of this process are *wisdom*.
  Organizers, participants, and visitors need a user interface which provides easy access and meaningful interaction with the information resources, allows them to cite and reference the information resources relevant for their work, and supports the enrichment of the offered information resources.

This detailed discussion shows how multifaceted the needs of the users involved in a large-scale evaluation campaign are and how different and complex the tasks are that the DLS used to manage the evaluation campaign has to support. This complexity is also reflected in the user interface, which needs to offer different types of interaction with the system according to the task and user at hand.

It is therefore necessary to design a user interface which is functional and responsive in order to make it easy to accomplish the desired task by studying solutions based on the user needs, the analysis of the interaction among users, and the user feedback. Furthermore, a large-scale evaluation campaign involves people who come from different countries, have different languages and different cultures; this factor has to be taken into account by providing a proper internazionalization and localization of the interface in order to lower language and cultural barriers.

The results of this study and analysis have been applied in designing and developing the user interface of DIRECT, a prototype DLS which has been used in the CLEF evaluation campaigns since 2005.

3 Architecture of the DIRECT User Interface

According to the investigation of user requirements and needs, the DIRECT user interface is designed to meet the following goals:

- to be cross-platform and easily deployable to end users;

- to be as modular as possible, clearly separating the application logic from the interface logic;
- to be intuitive and capable of providing support for the various user tasks described in the previous section, such as experiment submission, consultation of metrics and plots about experiment performances, relevance assessment, and so on;
- to support different types of users, i.e. participants, assessors, organizers, and visitors, who need to have access to different kinds of features and capabilities;
- to support internationalization and localization: the application needs to be able to adapt to the language of the user and his country or culturally-dependent data, such as dates and currencies.

The modularity of the components has enormous benefits when building interactive applications, since it helps the designer to better understand and develop each component and modify it without affecting the others. Therefore, we used the Model-View-Controller (MVC) (Krasner & Pope, 1988) approach as provided by Apache STRUTS[2] framework to clearly separate the following three layers:

- *model layer*: contains the underlying data structures of the application and keeps the state of the application;
- *view layer*: the way the model is presented to the user;
- *controller layer*: manages the interaction between the view and the input devices, such as the keyboard or the mouse, and updates the model accordingly.

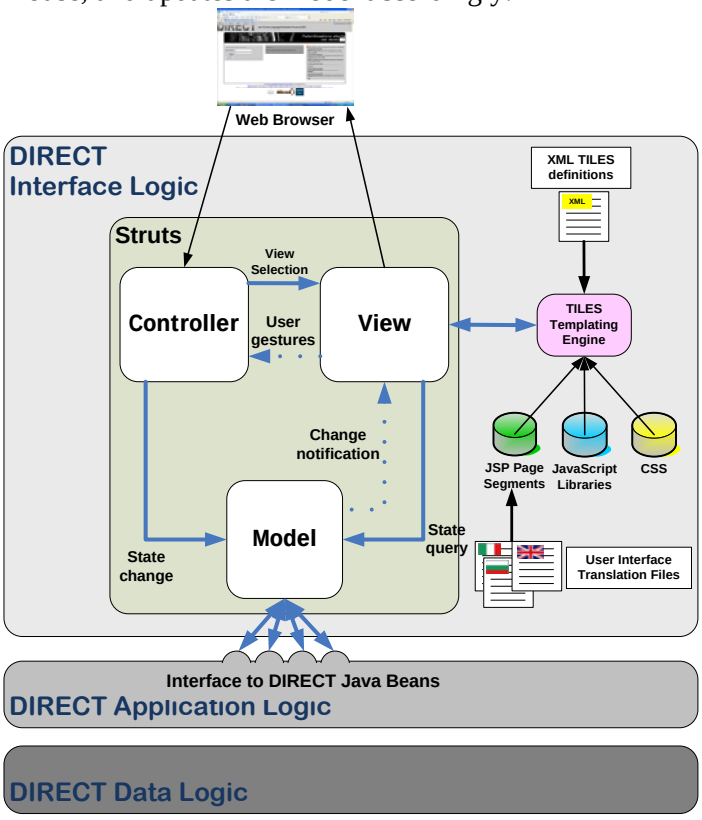

Figure 2: Architecture of the DIRECT User Interface.

Figure 2 shows the architecture of the DIRECT user interface which is Web-based in order to be cross-platform and easily deployable and accessible without the need of installing any software on the end-user machines. The user interface is based on JavaServer Pages (JSP)[3]; in addition, we developed JavaScript[4] library which provides event listeners, Document Object Model (DOM)[5] manipulation, and Asynchronous JavaScript and XML (AJAX)[6] support in order to make the user

[2] http://struts.apache.org/
[3] http://java.sun.com/products/jsp/
[4] http://www.ecma-international.org/publications/standards/Ecma-262.htm
[5] http://www.w3.org/DOM/
[6] http://www.w3.org/TR/XMLHttpRequest/

interaction more successful and responsive. In particular, AJAX allows us to make asynchronous calls to the server and to speed up the user interaction by loading only the requested portion of the data without requiring to download huge amounts of data in one time or to completely refresh a page when only a part of it is changed.

Moreover, the user interface is made more modular by using the STRUTS TILES[7] templating framework, which allows for a rapid development and reuse of components. As shown in Figure 2, when the browser requests a page, the STRUTS controller asks the TILES engine to put together the page components, according to instructions provided by an eXtensible Markup Language (XML)[8] configuration file. Then, TILES loads the JSP reusable code segments to create the page skeleton, adds the JavaScript libraries needed for enhancing the user interaction, fills the page with the contents provided by the STRUTS controller, applies the necessary Cascading Style Sheets (CSS)[9] for formatting the page, and return the dynamically created page to the View layer of STRUTS, which, in turn, sends it to the browser.

Finally, we also support the internationalization and localization of the user interface by adapting it to the language and country of the user. As shown in Figure 2, this is implemented by using translation files according to the Java internationalization capabilities[10]. The correct language and country are initially loaded according to the browser settings and, in the case of not supported locales, it falls back to a default configuration. After the authentication of the user his specific preferences can then be loaded. The user interface has been translated in the following languages: Bulgarian, Czech, English, French, German, Indonesian, Italian, Portuguese, and Spanish.

The user interface has evolved a lot from 2005 to 2007 in order to better model the interaction between the different kinds of users and their tasks and in response to user comments and suggestions. The evolution of the user interface is described in (Dussin & Ferro, 2007).

In the following section, due to space reasons, we will present only some of the various interfaces which have been developed to support the tasks discussed in the previous section; an idea of the full list of implemented interfaces can be had from Figure 1. We will show screenshots of the interface only in the English language, to improve readability.

3 User Interfaces of the DIRECT Prototype

3.1 Login Page

The login page, shown in Figure 3, adopts a three-column layout and, besides the login form, provides additional features, like the password recovery form, the automatic detection of user type, and a news ticker. The news, made available through a Really Simple Syndication (RSS) 2.0[11] feed, allows the user to be informed in real time about the latest events in the course of the evaluation campaign; moreover, users can subscribe to the news feed to get notifications and alerts from DIRECT; finally, news are also available in a browser-friendly format, created client-side using eXtensible Stylesheet Language Transformations (XSLT)[12].

After authentication, the user is automatically redirected to the main page of his user type where a navigation bar allows him to select available tasks, to personalize his account, change email, password, language, and country.

Error and information messages from the system are shown at the top of the page with a suitable iconography: green, yellow, or red according to the gravity of the message.

---

[7] http://struts.apache.org/1.x/struts-tiles/
[8] http://www.w3.org/XML/
[9] http://www.w3.org/Style/CSS/
[10] http://java.sun.com/javase/technologies/core/basic/intl/
[11] http://www.rssboard.org/rss-2-0/
[12] http://www.w3.org/TR/xslt20/

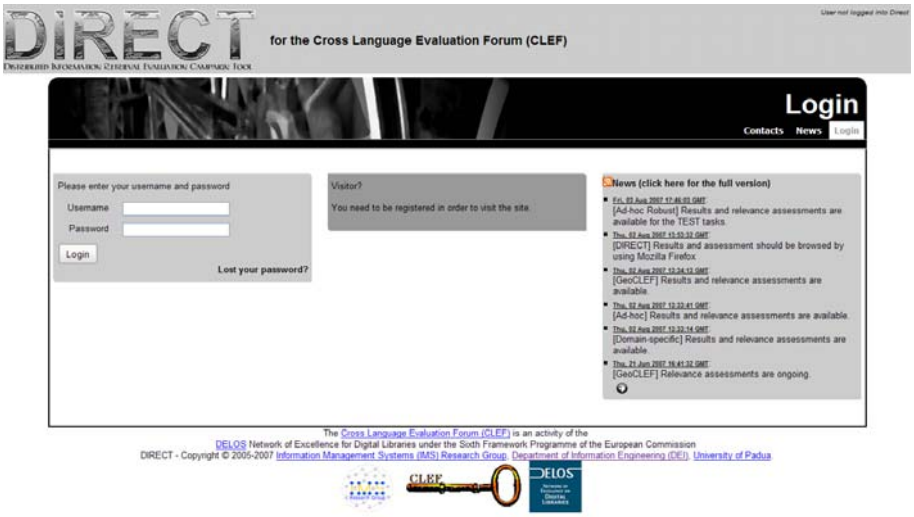

Figure 3: Login page.

3.2 Topic Creation Interface

Figure 4 shows the main page which the assessor uses for topic creation. The list of topics to be created is presented in a tabular way in order to display the information in a compact and coherent way; for each topic, the topic identifier, a summary, and the number of estimated relevant and not relevant documents are shown. Moreover, the table can be folded and expanded so that the user can concentrate on the contents of his own interest without having to read all the data or scroll the whole page. In addition, the system remembers the state of the table by using browser cookies[13] so that the user can find the table opened as he left it, if he reloads the page or logs in again.

Finally, the folding table also helps to increase the loading speed and responsiveness of the interface, as the needed data are loaded only when an assessor asks for a given portion of the table; indeed, asynchronous AJAX calls to the server allow us to retrieve only the portion of information of interest. Since a large amount of data is needed for filling in this table and it has to be downloaded to the client and rendered by the browser, this choice reduces the amount of data exchanged and, consequently, improves the responsiveness of the interface.

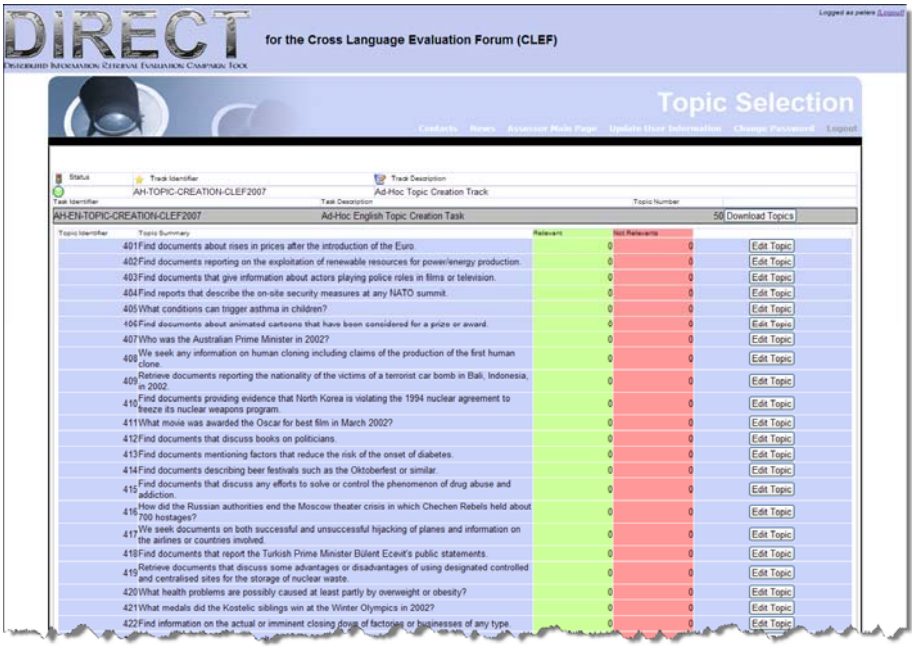

Figure 4: Main page for the topic creation.

[13] `http://www.ietf.org/rfc/rfc2965.txt`

Once the user clicks the “Edit Topic” button on the main page, the interface for the creation of the topics appears, as shown in Figure 5. On the left side, the assessor can modify the fields of the topic – i.e. title, description, and narrative. On the top of the interface, there is a search box which allows the assessor to search the document collections for the entered query and retrieve a list of documents; this is a fundamental operation, since the assessor has to check that there actually are relevant documents in the collection for the topic under creation.

As it can be noted from Figure 5, the list of retrieved documents is shown in the center of the page; for each document, the rank, the document identifier, a relevance assessment and a document summary are shown. The assessor can view the full content of the document by pressing the “View Document” button and can also express a relevance judgment about the document; this judgment is stored in the system and shown whenever the document is retrieved in response to a query. This feature allows assessors to explicitly indicate what document they think might be relevant or not for a given topic; this information is then shared with all the assessors who work on the same topics and contribute to their discussion during the topic creation process. Note that these judgements are different from the ones that are carried out on the pools in the relevance assessment step and which are used for computing the actual performance measures; we could say that the judgements made at topic creation time are a kind of estimate of the relevance of a document but not a final decision.

Finally, as shown in the lower part of the window, the assessor has the possibility to add notes and comments about a topic as well as to respond to other assessors’ notes. This feature is useful for supporting all the discussion that is usually carried out about a topic during its creation.

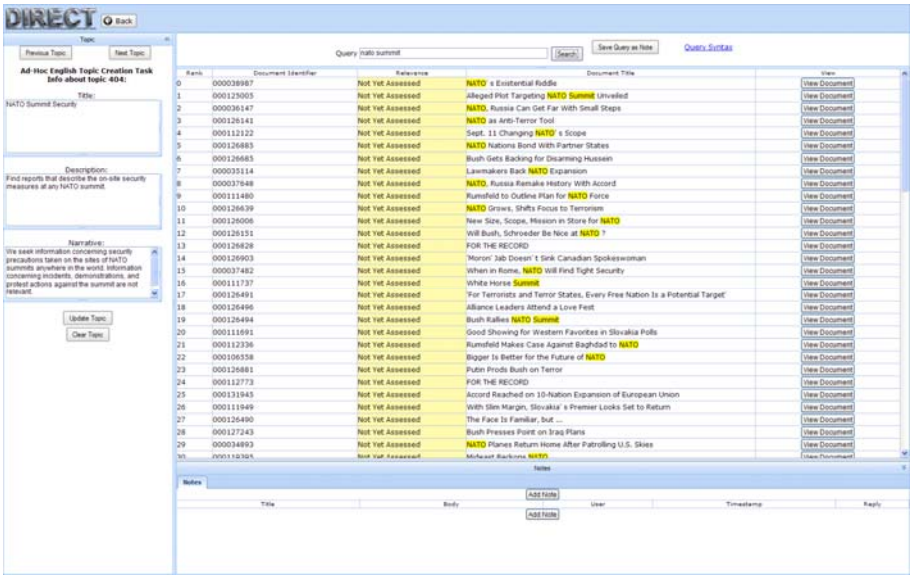

Figure 5: Topic creation interface.

As in the previous case, much care has been paid to make the interface as responsive as possible by loading page fragments only when needed through AJAX calls.

3.3 Experiment Management Interface

Figure 6 presents the main page for the experiment management which uses folding tables, like the topic creation main page, and allows the participant to access all the relevant information about an experiment. Buttons and drop-down lists are enabled or disabled according the current status of the campaign; for example, when the experiment submission is ongoing it is not possible to view the performance measures about an experiment, since these will be computed later on.

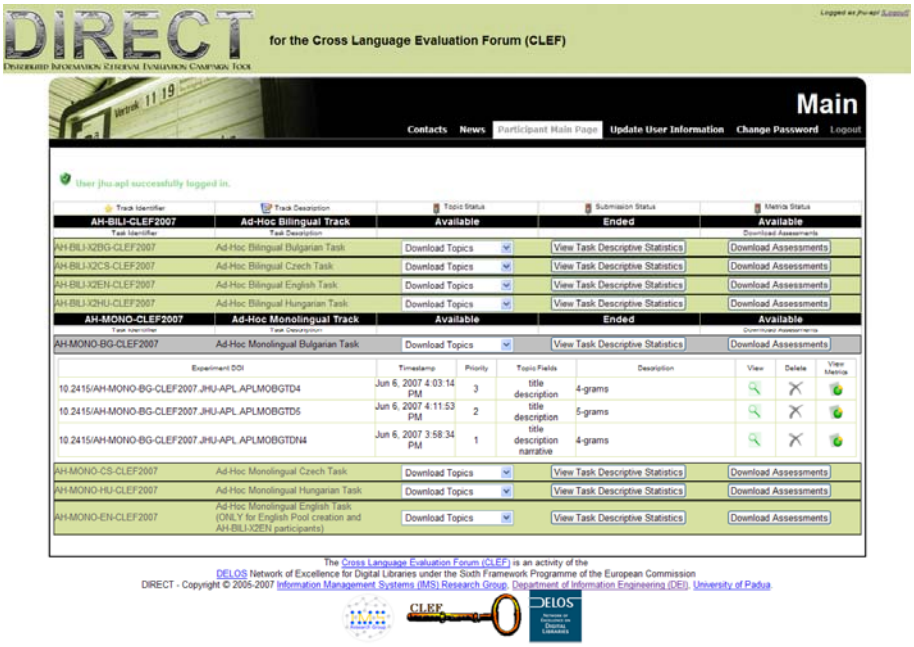

Figure 6: Main page for the experiment management.

This interface allows participants to submit their experiments, to download the topics which have to be used for a given task in a XML format, to download the relevance assessments, once they are available, and to access all the performance measures and statistics about the experiment or the tasks. Finally, for each experiment, summary information is reported: the experiment identifier, the submission timestamp of the experiment, the priority of the experiment, the topic fields used in the experiment, and a short description of the experiment.

When the participant clicks on the "View Metrics" button, the interface for accessing all the performance measures, plots, and statistics about an experiment appears. Figure 7 shows the information about the performance measures of the experiment computed topic-by-topic; for each topic, a list of metrics is reported and the participant can download the computed performances in an XML format.

Similarly, when the participant clicks on the "View Task Descriptive Statistics" button, the interface for accessing all the descriptive statistics and plot about a task appears. Figure 8 shows the information about the descriptive statistics computed for each topic over all the experiments submitted for a given task; for each topic, a list of statistics is reported – minimum, maximum, the different quartiles, the mean, the lower and upper outliers threshold, and so on – and the participant can download the computed descriptive statistics in a convenient XML format.

5 Conclusions

We have discussed the development of a DLS for managing the scientific data produced during the course of an evaluation campaign. In particular, we focused our attention on the design of the user interface of such a system and we faced this problem by framing the different tasks the interface has to support and their relationship with the different types of users of the system in the context of the DIKW hierarchy.

Furthermore, we have presented the architecture of the interface of the prototype DLS, called DIRECT, which we have implemented and tested in the CLEF campaigns since 2005.

Finally, we have shown some of the interfaces of DIRECT in order to give the reader a taste of the approach adopted and the results achieved.

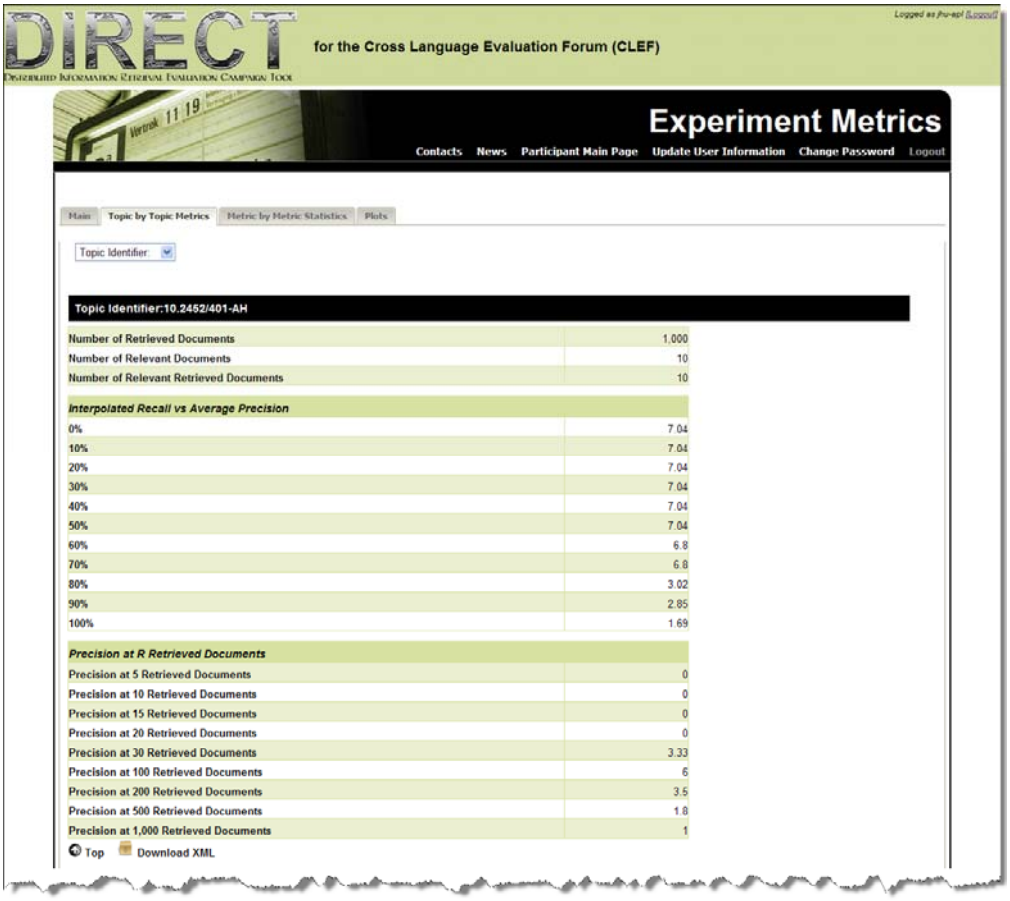

Figure 7: Interface for accessing the performance measures.

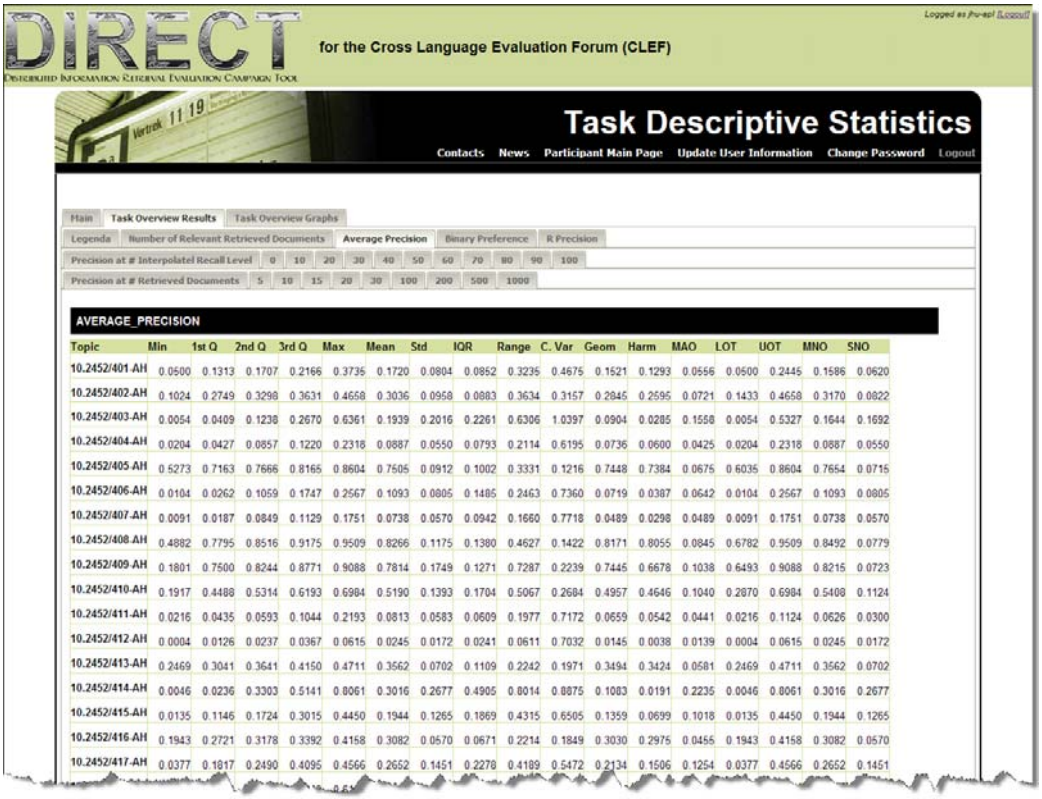

Figure 8: Interface for accessing the descriptive statistics about a task.

Acknowledgements

We would like to warmly thank all the people who contributed to the translation of the user interface: Petya Osenova and Kiril Simov for Bulgarian; Pavel Pecina for Czech; Jacques Savoy for French; Thomas Mandl for German; Mirna Adriani for Indonesian; Diana Santos and Paulo Rocha for Portuguese; and Julio Villena Román for Spanish.

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

Ackoff, R. L. (1989). From Data to Wisdom. *Journal of Applied Systems Analysis , 16*:3–9.

Agosti, M., Di Nunzio, G. M., & Ferro, N. (2007a). A Proposal to Extend and Enrich the Scientific Data Curation of Evaluation Campaigns. In T. Sakay, & M. Sanderson (Eds.), *Proc. 1st International Workshop on Evaluating Information Access (EVIA 2007)*. National Institute of Informatics, Tokyo, Japan. 62–73.

Agosti, M., Di Nunzio, G. M., & Ferro, N. (2007b). Scientific Data of an Evaluation Campaign: Do We Properly Deal With Them? In C. Peters, P. Clough, F. C. Gey, J. Karlgren, B. Magnini, D. W. Oard, et al. (Eds.), *Evaluation of Multilingual and Multi-modal Information Retrieval : Seventh Workshop of the Cross--Language Evaluation Forum (CLEF 2006). Revised Selected Papers*. Lecture Notes in Computer Science (LNCS) 4730, Springer, Heidelberg, Germany. 11–20.

Agosti, M., Di Nunzio, G. M., & Ferro, N. (2007c). The Importance of Scientific Data Curation for Evaluation Campaigns. In C. Thanos, & F. Borri (Eds.), *Digital Libraries: Research and Development Selected and Revised Papers from the First DELOS Conference.* Lecture Notes in Computer Science (LNCS), Springer, Heidelberg, Germany (in print).

Agosti, M., Di Nunzio, G. M., Ferro, N., Harman, D., & Peters, C. (2007). The Future of Large-scale Evaluation Campaigns for Information Retrieval in Europe. In N. Fuhr, L. Kovacs, & C. Meghini (Eds.), *Proc. 11th European Conference on Research and Advanced Technology for Digital Libraries (ECDL 2007)*. Lecture Notes in Computer Science (LNCS) 4675, Springer, Heidelberg, Germany. 509–512.

Cleverdon, C. W. (1997). The Cranfield Tests on Index Languages Devices. In K. Spärck Jones, & P. Willett (Eds.), *Readings in Information Retrieval*. Morgan Kaufmann Publisher, Inc., San Francisco, California, USA. 47–60.

Di Nunzio, G. M., & Ferro, N. (2005). DIRECT: a System for Evaluating Information Access Components of Digital Libraries. In A. Rauber, S. Christodoulakis, & A. Min Tjoa (Eds.), *Proc. 9th European Conference on Research and Advanced Technology for Digital Libraries (ECDL 2005)*. Lecture Notes in Computer Science (LNCS) 3652, Springer, Heidelberg, Germany. 483–484.

Di Nunzio, G. M., & Ferro, N. (2006). Scientific Evaluation of a DLMS: a service for evaluating information access components. In J. Gonzalo, C. Thanos, M. F. Verdejo, & R. C. Carrasco (Eds.), *Proc. 10th European Conference on Research and Advanced Technology for Digital Libraries (ECDL 2006).* Lecture Notes in Computer Science (LNCS) 4172, Springer, Heidelberg, Germany. 536–539.

Dussin, M., & Ferro, N. (2007). An Effective User Interface for a Scientific DLS to Support Information Retrieval Evaluation. In M. Agosti, & C. Thanos (Eds.), *Proc. 3rd Italian Research Conference on Digital Library Management Systems (IRCDL 2007).* ISTI-CNR at Gruppo ALI, Pisa, Italy (in print).

Fuhr, N., Tsakonas, G., Aalberg, T., Agosti, M., Hansen, P., Kapidakis, S., Klas, C.-P., Kovács, L., Landoni, M., Micsik, A., Papatheodorou, C., Peters, C., & Solvberg, I. (2007). Evaluation of Digital Libraries. *International Journal on Digital Libraries , 8* (1):21–38.

Krasner, G. E., & Pope, S. T. (1988). A Cookbook for Using the Model-View-Controller User Interface Paradigm in Smalltalk-80. *Journal of Object-Oriented Programming , 1* (3):26–49.

Lord, P., & Macdonald, A. (2003). *e-Science Curation Report. Data curation for e-Science in the UK: an audit to establish requirements for future curation and provision.* The JISC Committee for the Support of Research (JCSR), `http://www.jisc.ac.uk/uploaded_documents/e-ScienceReportFinal.pdf`.

Zeleny, M. (1987). Management Support Systems: Towards Integrated Knowledge Management. *Human Systems Management , 7* (1):59–70.