The ImageCLEF Management System

Ivan Eggel[a], Henning Müller[ab]

[a] Business Information Systems, University of Applied Sciences Western Switzerland
(HES–SO), Sierre, Switzerland
[b] Medical Informatics, University and Hospitals of Geneva, Switzerland
ivan.eggel@hevs.ch

Abstract. The ImageCLEF image retrieval track has been part of
CLEF (Cross Language Evaluation Forum) since 2003. Organizing
ImageCLEF and its large participation of research groups involves
a considerable amount of work and data to manage. Goal of the
management system described in this paper was to create a system
for the organization of ImageCLEF to reduce manual work and
professionalize the structures. All ImageCLEF sub tracks having a
page in a single run submission system reduces work of organizers
and makes submissions easier for participants. The system was de-
veloped as a web application using Java and JavaServer Faces (JSF)
on Glassfish with a Postgres 8.3 database. The main functionality
consists of user, collection and subtrack management as well as run
submissions. The system has two main user groups, participants
and administrators. The main task for participants is to register
for subtasks and then submit runs. Administrators create collec-
tions for the sub tasks and can define the data and constraints for
submissions. The described system was used for ImageCLEF 2009
with 86 subscribed users and more than 300 submitted runs in 7
subtracks. The system has proved to significantly reduce manual
work and will be used for upcoming ImageCLEF events and other
evaluation campaigns.

1 Introduction

ImageCLEF is the cross–language image retrieval track, which is run as part of
the Cross Language Evaluation Forum (CLEF). ImageCLEF[1] has seen participa-
tion from both academic and commercial research groups worldwide from com-
munities including: cross–language information retrieval (CLIR), content–based
image retrieval (CBIR) and human computer interaction. The main objective
of ImageCLEF is to advance the field of image retrieval and offer evaluation
in various fields of image information retrieval. The mixed use of text and vi-
sual features has been identified as important because little knowledge exists on
such combinations and most research groups work either on text or on images

---

[1] http://www.imageclef.org/

but only few work on the two. By making available visual and textual baseline results ImageCLEF gives participants data and task to obtain the information that they do not have themselves [1, 2]. ImageCLEF 2009 was divided into 7 subtracks (tasks) each of which provides an image collection:

- ImageCLEFmed: medical retrieval;
- ImageCLEFmed–annotation–IRMA: automatic medical image annotation task for the IRMA (Image Retrieval in Medical Applications) data set;
- ImageCLEFmed–annotation–nodules: automatic medical image annotation for lung nodules;
- ImageCLEFphoto: photographic retrieval;
- ImageCLEFphoto–annotation: annotation of images using a simple ontology;
- ImageCLEFwiki: image retrieval from a collection of Wikipedia images;
- ImageCLEFrobot: robotic image analysis.

ImageCLEF has been part of CLEF since 2003, with the number of registered research groups having grown from 4 in 2003 to 86 in 2009.. Taking the ever growing number of participants, it has become increasingly difficult to manage the registration, communication with participants and run submission manually. The data includes a copyright agreement for CLEF, submitted runs, task a user registered for, contact details for each participant. Registered groups became passwords for data download of each of the sub tasks that were send upon signature of the copyright agreement manually. The many manual steps created misunderstandings, data inconsistencies, and a large amount of email requests.

After several years of experience with much manual work, a computer–based solution was created in 2009. In this paper we present the developed system based on Java and JSF (Java Server Faces) to manage ImageCLEF events without replacing other already existing tools such as Easychair[2] for review management or DIRECT to evaluate results in several other CLEF tasks [3]. The new system was developed to integrate into the ImageCLEF structure and to facilitate organizational issues. This includes a run submission interface to avoid every task developing own solutions.

2   Methods

For the implementation of the system we relied on Java and JSF running on Glassfish v2.1. For data integration a Postgres 8.2 database was employed. The bridge between Java and Postgres was established with a Postgres JDBC 3 driver. Other Technologies used for client side interaction were pure Javascript and AJAX. The server used an Intel Xeon Dual Core 1.6 GHz processor with 2 GB of RAM and total disk space of 244 GB running on SuSe Linux.

---

[2] http://www.easychair.org/

3 Results

The ImageCLEF management system[3] mainly handles 4 functions: management of users, collections, sub tracks and runs. The possibility of dynamic sub track creation makes the system usable for other events and data of participants can be transferred from one event to another. Participating in a new event mainly includes setting up a new database making the application flexible.

3.1 User Management

Account Types Generally, there are two user groups in the management system: participants and administrators. Participants are users with the goal to participate in one or more ImageCLEF tasks and submit runs. After the registration and the validation of the copyright agreement by the organizers, a user is allowed to submit runs. Administrators are users that enjoy rights to set up and modify the system with essential data, e.g. creating subtracks or delete users. They can also act as participants for run submissions. Usually, all ImageCLEF organizers have their own administrator accounts. To become an administrator the user needs to be registered as a participant. An existing administrator can then convert an existing participant account into an administrator account.

User Registration Each participating group can register easily and quickly. A link for the registration on the initial login page will guide the user to the registration process. For security reasons it is not possible to register as an administrator, so it is necessary to register as a participant first. To complete the registration, the following information needs to be provided:

– group name (e.g. name of association, university, etc.);
– group e–mail address (representative for the group);
– group address;
– group country;
– first name of contact persony;
– last name of contact person;
– phone number of contact person (not mandatory);
– selection of sub tracks the participant wishes to participate in.

After submitting the registration form the system validates all input fields and (in case of validity) stores the participant's registration information to the database, which at the same sends the login password to the participant by e–mail.

General Resources/Tasks of User Management There are several resources and tasks for user management, which include viewing a list of all users, users' details, updating and deleting a user as well as validating pending participant signatures. In Figure 1 the list of all users shows a table with users row

Fig. 1. List of all the users, allowing to sort by various criteria and with different views.

by row. Every row represents a user with the possibility to navigate to the detail and update pages by clicking the according links in the table. There is also a delete button in the row, which will remove the user from the database. Only administrators are allowed to delete participants, however it is not permitted to remove another administrator account. It is possible for every user, regardless of being administrator or participant to view a user detail page, however with the restriction of participants not being able t o see the list of submitted runs within another user's page (see Figure 2). The system also provides an update function. While participants can only update their own accounts, administrators are allowed to update all participants they wish to. Only administrators possess the authorization to validate a participant's signature for the copyright agreement.

3.2 Collection Management

A collection describes a dataset of images used for the retrieval. Since all sub-tracks are associated with a collection the creation of a collection has to be performed before adding a sub track. Theoretically, the same collection can be part of several sub tracks. Any administrator can create new collections. For a new collection the user needs to provide information like the name of the collection, the number of images in the collection and the address to its location on the web. Additionally, the user has to provide an imagenames–file, which represents a file containing the names of all images in the collection with one imagename per line. Providing this file is essential to perform checks for run submissions, i.e.

User detail of User No. 47

Group name	LIRIS
E-mail	loic.maisonnasse@insa-lyon.fr
Address	7 avenue Jean Capelle B.501.315 69100 Villeurbanne
Country	France
Firstname	Loïc
Lastname	Maisonnasse
Phone Number	
Registration date	Apr 16, 2009
Account type	**Participant**
Signature OK	✓

Subscribed to the following ImageCLEF subtrack(s)

ImageCLEFmed

User has submitted the following runs

Subtrack	Retrieval Type	Submission Date	Primary Run	Validated	Detail
ImageCLEFmed	Textual	Jun 10, 2009	✗	✗	Detail
ImageCLEFmed	Textual	Jun 10, 2009	✗	✗	Detail
ImageCLEFmed	Textual	Jun 10, 2009	✓	✗	Detail
ImageCLEFmed	Textual	Jun 10, 2009	✗	✗	Detail
ImageCLEFmed	Textual	Jun 10, 2009	✗	✗	Detail
ImageCLEFmed	Textual	Jun 10, 2009	✗	✗	Detail
ImageCLEFmed	Textual	Jun 10, 2009	✗	✗	Detail
ImageCLEFmed	Textual	Jun 10, 2009	✗	✗	Detail
ImageCLEFmed	Textual	Jun 10, 2009	✗	✗	Detail

Fig. 2. The view of the details of one user.

if the images specified in the submitted run file are contained in the collection. It is also only possible for administrators to perform updates on existing collections if necessary. The update page provides the possibility to change ordinary collection information as well as the exchange of the imagenames–file.

3.3 Subtrack Management

Each subtrack determines a beginning and an end date preventing participants from submitting runs for this subtrack when the time period for submission is over. Every subtrack allows only a limited number of submitted runs per participant. Like all organizational tasks, creating a new subtrack is only possible for administrators. The interface for the creation of new subtracks asks to provide information like the name of the collection, the maximal number of runs allowed as well as start and end dates of the task. Providing these dates will prevent a participant from submitting runs for this task before the task starts or after the task has finished. It is equally important to select the collection associated with the subtrack, which demands prior creation of at least one collection. In a task view, all submitted runs for the task are listed in a table (only accessible to administrators). Administrators also enjoy the privilege to download all submitted runs for the task in one zip file. All participants in the subtrack are listed.

Fig. 3. Example for a run submission.

3.4 Runs

Run submission is one of the central functions of the presented system. Each participant has the opportunity to submit runs. Administrators can act as participants and thus submit runs. Figure 3 shows an example of run submission.

The main item of a run submission is the runfile itself, which can be uploaded on the same page. After the file upload and before storage of the metadata to the database, the system executes a runfile validation. Due to varying file formats among the tasks there are specific validators created for each task. In case of invalid files the transaction will be discarded, i.e. the data will not be stored to the system and an error message will notify the user avoiding the submission of runs in incorrect format. Likewise, the validator assures that each image specified in the run file has to be part of the collection. All this avoids the submission of incorrect run files and thus manual work of the organizers.

Administrators have the possibility to see all submitted runs in a table, whereas ordinary participants are only allowed to see their own runs. The simplest way for a admin to view his or another user's submitted runs is to inspect the user's detail page. For administrators, a table with all submitted runs of all users appears also on the initial sub track page. A useful feature for administrators is the opportunity to download all runs of a subtrack in one zip file. The system generates (at runtime) a zip file including all runs of a particular task. The same page equally provides the facility to download a zipped file of run

meta data xml files with each file corresponding to a run. After submission it is still permitted to modify own runs by replacing the runfile or by altering meta information on the run.

4 System Use in 2009

The registration interface of the system provided an easy way for users to register themselves to ImageCLEF 2009. The system counted 86 registered users from 30 countries. 10 of these users were also system administrators, the rest normal ImageCLEF participants. ImageCLEF 2009 consisted of 7 sub tracks (see Table 1). With 37 the ImageCLEFphoto–annotation task had the largest number of participants whereas the RobotVision task with its 16 participants recorded the smallest number. As shown in Table 1, participants of the ImageCLEFmed task submitted 124 runs in total, which was the highest number of submitted runs by subtrack, although the task did not have the largest number of participants. The high number of submitted runs was partly due to ImageCLEFmed being devided into image–based and case–based topics, allowing groups to submit twice as many runs. Both ImageCLEFmed–annotation tasks as well as ImageCLEFphoto did not use the system's run submission interface and used other tools. However, it is foreseen that all tasks will provide their run submission in the future. There were a total of 39 participants that did not submit any run on the system. Some of these participants only participated in tasks that did not use the described interface and others finally did not submit any runs.

Sometimes groups registered with more than one email address and in these cases we ask groups to remove the additional identifiers and have a unique submission point per group.

Table 1. ImageCLEF tasks with number of users and submitted runs.

Task	# users	# runs
ImageCLEFmed	34	124
ImageCLEFmed-annotation-IRMA	23	19
ImageCLEFmed-annotation-nodules	20	0
ImageCLEFphoto	34	0
ImageCLEFphoto-annotation	37	74
ImageCLEFwiki	30	57
RobotVision	16	32
TOTAL	86	306

5 Conclusion

This paper briefly presents a solution to reduce manual and redundant work for benchmarking events such as ImageCLEF. Goal was to complement already

existing systems such as DIRECT or Easychair and supply the missing functionality. All seven ImageCLEF tasks were integrated and almost all participants who registered for ImageCLEF on the paper–based registration also registered electronically. Not all tasks used the provided run submission interface but this is foreseen in the future. With 86 registered users and more than 300 submitted runs the prototype system showed to work in a stable and reliable manner.

Several small changes were performed to the system based on comments from the users, particularly in the early registration phase. Reminder emails for forgotten passwords were added as well as several views and restrictions of views on the data. In the first version, run file updates were not possible once the run was submitted. This was changed. Confusion caused the renaming of the original run file names by the system after submission, which was meant to unify the submitted names based on the identifiers given inside the files. Some participants were then unable to properly identify their runs without a certain effort. To avoid this, the system will keep original names of runfiles in the future. There is also more flexibility in the meta data for each of the runs before submission but the goal is to harmonize this across tasks as much as possible.

The management system could enormously reduce manual interaction between participants and organizers of ImageCLEF. As the standard CLEF registration was still on paper with a signed copyright agreement, the electronic system gave the possibility to have one contact with participants and then make all information available at a single point of entry, the ImageCLEF web pages and with it the registration system. Passwords did not need to be sent to participants manually but access was organized through the system. Having a single submission interface also lowered the entry burden for participants of several sub tasks. Having only fully validated runs avoided a large amount of manual work for cleaning the data and contact with participants.

6 Acknowledgements

References

1. Clough, P., Müller, H., Deselaers, T., Grubinger, M., Lehmann, T.M., Jensen, J., Hersh, W.: The CLEF 2005 cross-language image retrieval track. In: Cross Language Evaluation Forum (CLEF 2005). Springer Lecture Notes in Computer Science (2006) 535–557
2. Peters, C., Jijkoun, V., Mandl, T., Müller, H., Oard, D.W., Peñas, A., Petras, V., Santos, D.: Editorial. In: CLEF 2007 Proceedings. Volume 5152 of Lecture Notes in Computer Science (LNCS)., Budapest, Hungary, Springer (2008)
3. Nunzio, G.M.D., Ferro, N.: Direct: A system for evaluating information access components of digital libraries. In Rauber, A., Christodoulakis, S., Tjoa, A.M., eds.: ECDL. Volume 3652 of Lecture Notes in Computer Science., Springer (2005) 483–484