EVAL: A Web-based Design Review System

BRIAN JOHNSON
University of Washington

BRANKO KOLAREVIC
University of Hong Kong

Abstract

The Association for Computer Aided Design in Architecture (ACADIA) recently conducted a design competition in which design projects were presented using web sites. As a consequence it was not necessary to co-locate reviewers in order to view submissions. Since the proposals took the form of web sites, it seemed appropriate to use the web as the medium for conducting the review. The review thus became an opportunity to explore online design studio review strategies as well as competition issues. As there were over 600 entries in the competition, each of which was to be reviewed by at least three reviewers, the review process presented certain logistical challenges that might not pertain to a "normal" design studio. Using a globally-distributed review panel and jury meant that synchronous review of projects would not be possible, and that face-to-face interaction between jurors would be lost.

This paper describes the review system which was developed to address this need. It also profiles the conduct of the review itself, and offers some observations about performance, ergonomics and related design issues for future efforts.

Overview

Beginning with a number of ambitious experiments conducted in the early 1990's and described in (Wojtowicz, 1994) a growing number of design studios have been conducted or facilitated by use of Internet and web technology. In the simpler cases these projects consist of student web pages with "mailto:" links for feedback by reviewers. However, other projects, by (Chen *et al.*, 1994), (Kolarevic *et al.*, 1998) and others have attempted to foster and observe the growth of collaborative student communities on the web, using common design programs, distributed design teams, shared storage space, video-conferencing and email exchange. While much has been accomplished, interaction is often cumbersome and dependent on specialized equipment and/or specialized computing skills. There is much room for improvement in terms of making the technology "transparent" and enabling use of such systems to function as design communities.

This particular project, which focuses on design review and evaluation, grew out of the opportunity presented by the First ACADIA International Design Competition (ACADIA, 1998), a design project in which proposals for a "cybrid" (Anders, 1997) library design were developed as web sites. Given that the medium of presentation was the web, and since ACADIA members who would be participating in any review might be located thousands of miles apart, it was desirable to use the web as the review medium as well. Unlike the previous experiments with design review that had been conducted at the University of Washington (Novitski, 1996) this review needed to be private, needed to accommodate sites located away from the review system, needed to handle a sizable number of projects, and ultimately needed to produce a sorted ranking of the projects so that prizes could be awarded.

Traditional competition design reviews have been described this way by Martini (1999): "Entries are typically submitted as a series of standard-sized boards, which are placed around a big room, often stacked so that only the top board is visible. As a juror, it is possible to walk around the room, scanning the front boards, stopping to spend more time and flip through behind-boards of interesting looking projects." Clearly, while semi-independent, the jurors are able to observe each other and interact socially. They are able to quickly overview of the competition responses, as well as invest time in more detailed examination where warranted. An on-line review system should replicate or provide substitutes for various components of this process.

In order to more easily generate a ranking, it was decided to have the critics supply a quantitative score for each web site that they visited in addition to making qualitative comments. A preliminary review panel was assembled and assigned the task of identifying the top projects. Some 600 web sites needed to be visited by three separate reviewers, each of which would then record an assessment. These 1800 "review events" would be generated by a review panel of some 45 individual reviewers located all over the globe, and would take place over a period of about two weeks, at the end of which time we needed to be able to announce the 50 finalists. These projects would then be separated out, and the final jury would use the same system to perform final winnowing and selection of the prize winners.

System Design Considerations

There were four major areas of concern in the design of the evaluation system: The character of the evaluation interaction itself, clear record keeping and data management, system security, and standardized evaluation. A system of direct static links and emailed evaluations was considered, but discarded as requiring

[Originally printed in Proceedings of ACADIA'99: "Media and Design Process", Snowbird, Utah, October 29-31, 1999, pp 30-39]

too much transcription of data by the central record-keepers. A series of links and bulletin board comments was also considered, but discarded because individual jurors would not retain control of their statements, and it would not provide for the necessary scoring. In the end it was decided to implement the system using a combination of a custom CGI "middleware" application and a commercial back-end database. This combination would allow us to address the first three issues, which are explored in more detail below. The issue of standardized scoring will be addressed later.

The Interaction Issues

The review panel and final jury would not have significant amounts of time to practice with the system. It needed to be straight-forward and easy to use.

The reviewers needed a central dispatch point from which to access sites, and they needed to be able to file their conclusions at the end of a review. Obviously, they needed to be able to navigate freely from one site to another, in order to get a general sense of the projects as a group, and they might well wish to review their evaluations as a group, or check the comments posted by other reviewers from time to time.

Given the number of sites, and the sometimes slow connection speeds, it was desirable to minimize the time spent waiting for the dispatch page to reload. It was decided to implement the system as a console, keeping the page on screen. This constant presence would also serve as a visual anchor for reviewers, presenting one consistent element as they jumped from entry to entry.

Implementing the scoring console had another advantage as well. There was no way to anticipate the frame or window structures which the entry sites might utilize (via JavaScript, etc.). By opening each site in it's own window (see the right side of Figure 1), which could then be under complete control of the subject site, problems were avoided.

The reviewers logged the results of their evaluations using as an HTML form. To keep the number of discrete windows on screen to a minimum, this form was presented as a frame within the console window (see the left side of Figure 1).

Finally, since site names were purposely simplistic and generic, reviewsers needed to be able to tell which sites they had reviewed and which not. In Figure 1, only site "E310" has not been visited by this reviewer, as shown by the plain (green) square in the "X" column.

Record Keeping and Data Management Issues

The scoring system would be online for several weeks. During this time it was possible that the server would suffer a hardware failure or that some unanticipated bug in the sofware would corrupt the database. We needed to be confident that the data was accurate.

The list of competition entries and reviewers was fixed at the start of the process (and were entered in the database directly), so the main task of the system was to record evaluations of entries submitted by reviewers, make those evaluations available to the reviewers to modify, and keep summary statistics up to date. However, given the numbers of participants in the process, it was desirable to have a single channel through which to make the inevitable last-minute modifications.

Security Issues

All http interactions are fundamentally public and stateless (independent of previous interactions). Obviously, we needed to know that the scores were being recorded by our reviewers. Since any entrant with a server log would be able to tell the URL of the referring page, it was important that the system provide for privacy and that it validate information before it was entered in the database. However, we did not feel that we needed to carry this to the level of encrypted transactions, nor was it desirable to force the user to include a password with every evaluation posted to the database.

Any state information, such as validation of a user, must be maintained separately from the http interaction itself. In this case, a "login & trust" model was used, in which the reviewer was asked to authenticate themselves with a personal ID and PW when they first connected, at which point the CGI recorded their host ip-number. Subsequent traffic from this host was assumed to come from that individual. An "inactivity timer" removed them from the list of trusted users when 30 minutes had passed without a transaction.

The "Eval" System

The working name for the system which was developed is "Eval". EVAL consists of an AppleScript CGI, a few related HTML pages and three FileMaker Pro databases, one each for Entrant data, Evaluator (reviewer) data, and Score data.

Database Schema

A simplified schema of the database information is shown in Figure 1. As shown, there are three major databases: ENTRANTS, for competition entries needing review; EVALUATORS, for reviewers giving review-group assignments, logon ID and PW, etc.; and SCORES, for the scores or evaluations of the entries.

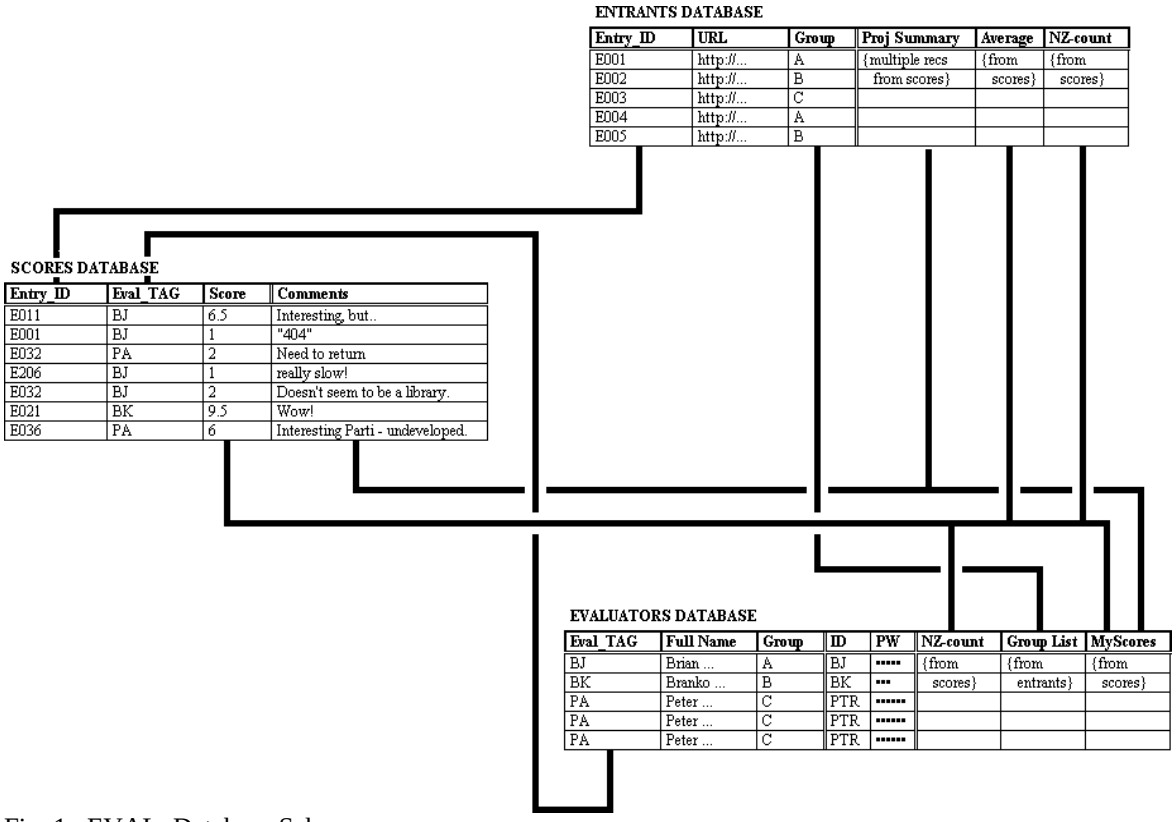


Fig. 1. EVAL: Database Schema

Linking relationships occur between each pair of databases (Group assignments are shared by Entrants and Evaluators, Entry_IDs bind scores to Entries, and Evaluator Tags link scores to evaluators). As a consequence, automated pre-calculation of lists can be achieved in FileMaker through "portals". Thus, each reviewer has an list of Entrants (the Group List) to evaluate, bound to him/her by the Group name. The evaluator also has a list of the scores they have recorded (the MyScores list). Each Entrant has a number of scores, an average score, and a Project Summary listing of all scores and comments, all copied from the Scores database using the Entry_ID.

User Interface

As discussed above, EVAL displays an omnipresent console to facilitate speedy navigation and aid in maintaining reviewer orientation. A simplified illustration of the distribution of screen area is shown in Figure 2, where areas 1 and 2 appear in one window, called the Eval Console, and area 3 is the competition entry window. This console is divided into two parts, an upper area (#1 in Figure 2) where different lists of data are displayed, and a lower area (#2 in Figure 2) where evaluation sheets are displayed. Whenever a score is recorded through use of the evaluation sheet, the upper window is refreshed so that it stays up-to-date.

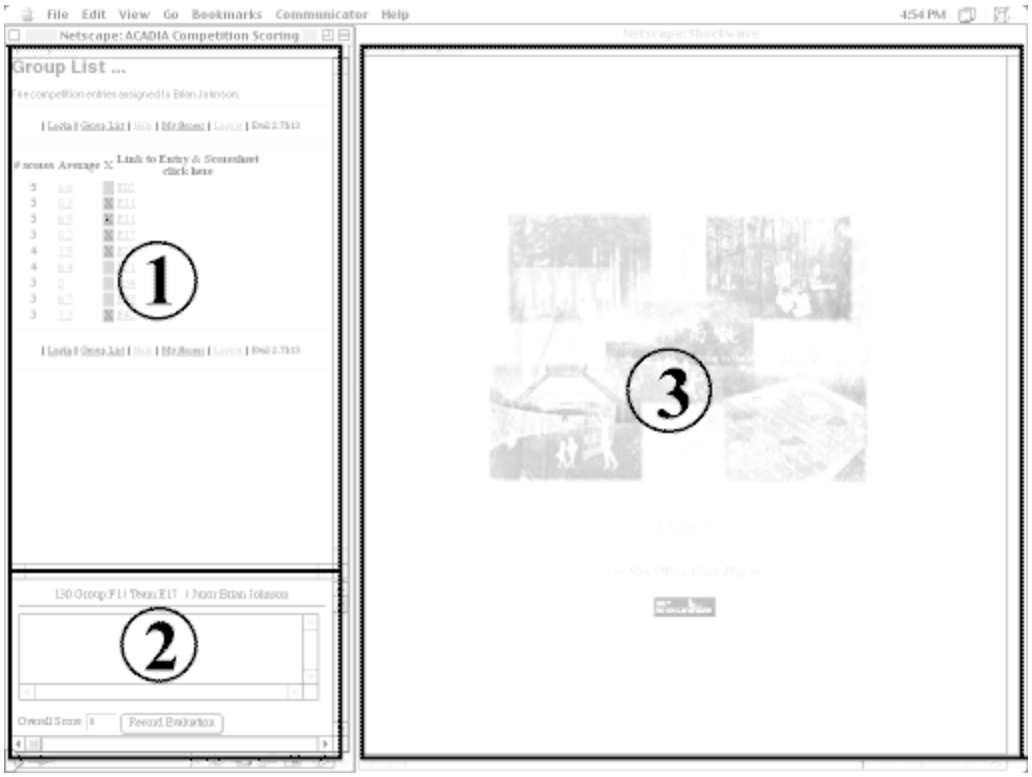


Fig. 2. EVAL: Review Console and one Entry Site Window during a review session

Initial contact occurs though an HTML login form (Figure 3). This page includes a width gauge to aid in setting the window width of the scoring console.

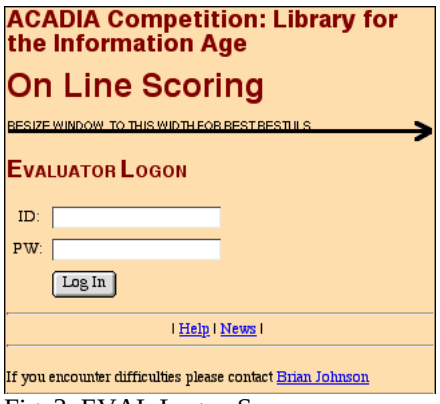


Fig. 3. EVAL Logon Screen

Once a correct ID and PW have been entered, the CGI stores the reviewers ip-number for subsequent trust testing and the full scoring console is displayed, as shown to the left of Figure 2. Initially the "Group List" is displayed, showing only the reviewer's assigned projects (for detail, see Figure 4).

The figure shows the console along the left edge of the screen, with room to the right for "site" windows. Most users probably overlap the two windows, but the relative positions are appropriate. Initially, the score sheet frame in the lower third of the console window (#2 in the figure) is empty. A score sheet is displayed when the reviewer connects to a site.

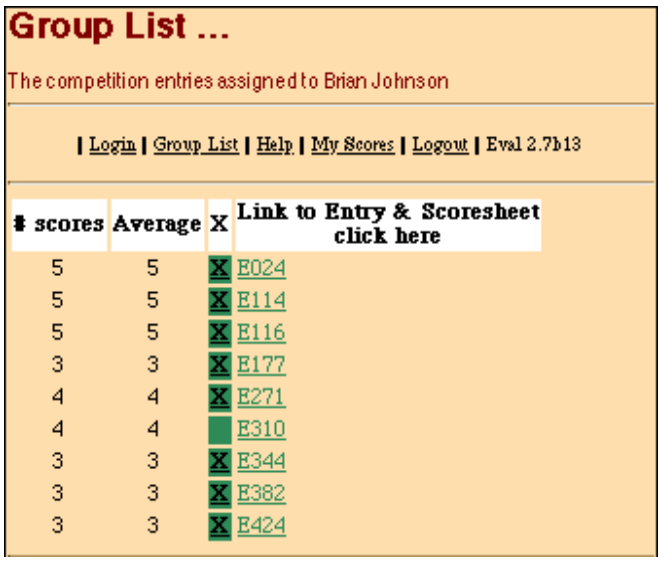


Fig. 4. EVAL: Group List Detail

The "Group List"/Scoring Mechanics

As illustrated in Figure 4, the "Group List" includes the following information for each project: the number of scores already entered, the average of those scores (here altered to obscure the actual scoring data), a "visited" check-box, and a link to the project web site.

To initiate evaluation of a project the reviewer clicks on the project ID in their Group List. Using a JavaScript function, two links are made: a score sheet is loaded in the bottom frame of the console (#2 in Fig. 2), and a new window is opened to display the target site (#3 in Fig. 2). This provides the site with a separate "browsing universe' in which to be viewed and limits the collateral damage that bad HTML or JavaScript can cause.

The score sheet (Figure 5) is generated by the CGI based on data recorded in the SCORES database. Hidden on the form is the ID of the reviewer and of the site being reviewed. The reviewer's current evaluations (comment and score) for this project are provided as the default values for the score sheet. If there is no evaluation on record, a zero-value-score is created and stored in the database as a first step (these scores are not used in the computation of averages or when counting scores logged, but they provide a means of telling whether a site has actually been viewed by the reviewer).

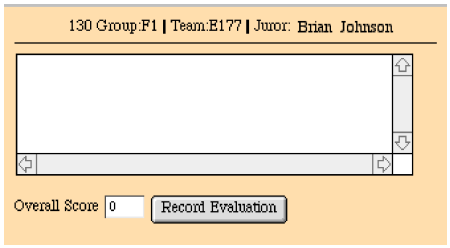


Fig. 5. EVAL Evaluation Sheet Detail

When the reviewer has finished viewing the site they edit or enter the score sheet comment and score fields and then click "Record Evaluation" to update the database. A JavaScript pre-checks the score to insure that it is within the acceptable range of values (1..10), incidentally insuring that the reviewer has actually entered a score.

The reviewer may now close the site window or keep it available for reference as they look at the next site. If another site window is called up by selecting a new entry from the Group List, a new score sheet is displayed in the scoring frame. No change is made to the database. This enables the reviewer to re-visit sites, review their comments and score, and move on without re-affirming or redundantly entering data.

When a score sheet is submitted, the CGI checks that the source ip-number is currently trusted, then updates the database, logs the score change to the log file, emits an archival email message, and returns a page

which acknowledges the score and forces an update of the listing in the upper part of the console window, refreshing that display so it is accurate.

The Visited checkbox-Differentiating Evaluators.

Using FileMaker to pre-compute the HTML for the Group List is much faster than doing it through AppleScript. Unfortunately, this means that the Group List HTML for each reviewer of a given group is identical. The system cannot specially "tag" those projects that this particular reviewer has visited. This would not normally be a problem because the color of a visited link is different from that of one not yet visited. However, in this case the link is executed by a JavaScript function and does not change color.

To address this, the "X" column was added (see Figure 4) to help reviewers keep track of the projects that have been scored. The X is simply text, defined as a direct link to the target site. The page's "not visited' (LINK) color matches the background color of the table cell, so it is not visible initially. Once a site has been visited through the JavaScript link, however, the X's color changes to the "visited link" (VLINK) color (black) and the X becomes visible. Because the application of LINK and VLINK colors occurs in the browser and depends on that browser's history file, the database/CGI isn't sending different HTML, but the different reviewers get different progress feedback.

Unfortunately, this means that the "X" is machine-specific, and that it only indicates sites visited, not sites scored. While this is a limitation, it was felt that the overall benefit of having (even a crude) "you are here" marker was worth it.

Standardizing Evaluation Scores

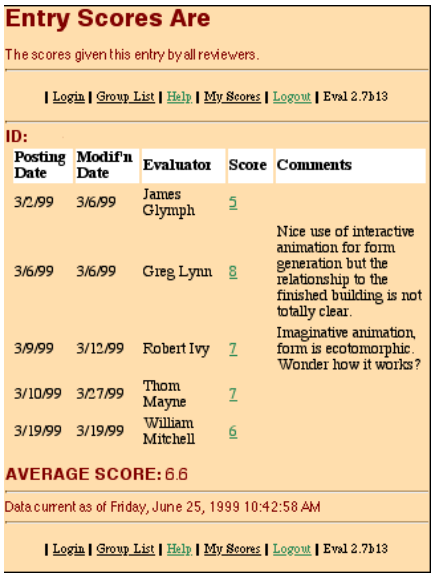
Each score record consisted of a numerical evaluation and text comment. While the Comments provided for a low-key "conversation" of sorts amongst jurors, as well as a means of recording site access problems and so on, the numerical "score" issued by each reviewer was obviously very important. In an effort to standardize this numerical evaluation, the following keywords were established.

Score	Descriptor	Elaboration
1	UNACCEPTABLE	(not there, doesn't load, isn't related to the competition)
2	INCOMPREHENSIBLE	(Lacks navigation tools, unreadable, etc.)
3	MINIMAL	(On topic, but either the design or the web site isn't sufficiently developed.)
4	SO-SO	(addresses the project requirements but isn't interesting; lacks "pizzazz")
5	COMPLETE	(when that's the best you can say about it)
6	GOOD	(Offers some interesting insights into web/library. Competitive)
7	PRETTY GOOD	(begins to display mastery of design/web issues)
8	VERY GOOD	(displays high level of skill and understanding of design/web issues)
9	OUTSTANDING	(you want to E-mail somebody to tell them about it)
10	SUPERB	(Your personal favorite (a possible winning entry?))

Fig. 6. Scoring vocabulary and values

Alternate Views of the Scoring Data

In addition to the Group List, two other listings exist for viewing and accessing scores: the ***Project Summary*** listing, and the ***My scores*** listing. If the reviewer clicks on the average score shown in the Group List, they will be taken to the detailed score summary for that one project (Figure 7). In this view they can see the scores and comments offered by other reviewers, as well as the dates of posting and modification. By clicking on one of the scores they can call up a score sheet and use it to change their own score and comments for the project (but not those of others).



Entry Scores Are

The scores given this entry by all reviewers.

| Login | Group List | Help | My Scores | Logout | Eval 2.7b13

ID:

Posting Date	Modif'n Date	Evaluator	Score	Comments
3/2/99	3/6/99	James Glymph	5	
3/6/99	3/6/99	Greg Lynn	8	Nice use of interactive animation for form generation but the relationship to the finished building is not totally clear.
3/9/99	3/12/99	Robert Ivy	7	Imaginative animation, form is ecotomorphic. Wonder how it works?
3/10/99	3/27/99	Thom Mayne	7	
3/19/99	3/19/99	William Mitchell	6	

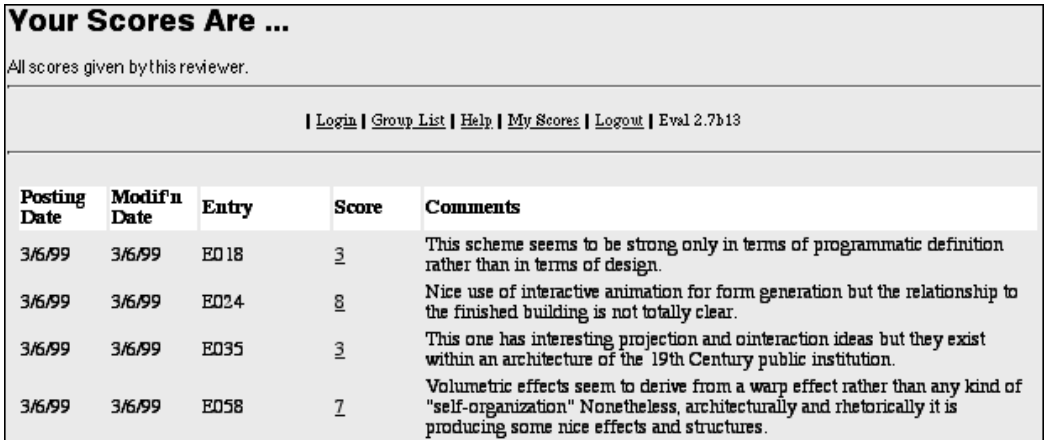
AVERAGE SCORE: 6.6

Data current as of Friday, June 25, 1999 10:42:58 AM

| Login | Group List | Help | My Scores | Logout | Eval 2.7b13

Fig. 7. EVAL Project Scoring Summary

The **My Scores** link takes the reviewer to a summary of the scores and comments which they have given so far (Figure 8). As with the **Project Summary** listing, the reviewer may click on any score in order to bring up a score sheet in the bottom frame for editing.



Your Scores Are ...

All scores given by this reviewer.

| Login | Group List | Help | My Scores | Logout | Eval 2.7b13

Posting Date	Modif'n Date	Entry	Score	Comments
3/6/99	3/6/99	E018	3	This scheme seems to be strong only in terms of programmatic definition rather than in terms of design.
3/6/99	3/6/99	E024	8	Nice use of interactive animation for form generation but the relationship to the finished building is not totally clear.
3/6/99	3/6/99	E035	3	This one has interesting projection and ointeraction ideas but they exist within an architecture of the 19th Century public institution.
3/6/99	3/6/99	E058	7	Volumetric effects seem to derive from a warp effect rather than any kind of "self-organization" Nonetheless, architecturally and rhetorically it is producing some nice effects and structures.

Fig. 8. EVAL Reviewer score summary showing some of a reviewer's scoring.

Logging Out

When a reviewer is done scoring they "Logout" to leave the system. Logging out cancels the trust relationship between their ip-number and the scoring CGI, returning them to the general category of generic web user.

Timing Out

In the event that the reviewer doesn't log out, but simply stops communicating with the CGI, their trusted status is retained for 30 minutes and then the CGI performs a log-out for them.

Administrative Links

In addition to the main scoring interface, the CGI responds to two other commands: ***Summary*** and ***Users***. Summary produces a list of all reviewers, with the number of scores each has recorded. This was quite useful for the competition webmaster, who was able to redistribute review assignments to cover situations in which a reviewer did not complete their assigned group. The *users* command the CGI to display a list of currently logged-on reviewers and the amount of time left on their inactivity timer.

Results

Implemented in late 1998, the system was refined and tested in early 1999. It was then used, as planned, to manage evaluation of the First ACADIA Design Competition. Early on in the scoring, a problem was uncovered when multiple evaluation hits occurred simultaneously, but this was resolved, and the database corrected by hand. There were no other bugs uncovered.

In terms of usability, the 45 members of the review panel were able to complete their review in about two weeks, with very few reported problems (there were two reports of access difficulties, never completely resolved).

During the review process, the CGI handled over 4000 service requests. By the end of the review nearly 1800 scores had been recorded, some going through several revisions. As an added bonus, the competition webmaster found that he was able to easily track progress, easily assign jurors to additional or different review panels, and easily correct such problems as project URL's that changed at the last minute.

The system was implemented initially on a Macintosh WGS 7250. However, with the number of CGI invocations involved in active scoring, and the time required to process each transaction, it became apparent that system performance was not adequate. The evaluation system was then moved to a faster machine (a PowerMac 8600/300) which was dedicated to hosting the scoring site for the remainder of the judging. This system provided response times consistent with browsing static HTML from most servers.

Conclusions

The first ACADIA design competition was conceived as an event that would establish a marker against which future web-based competitions might be compared. Response to the competition indicates that it has. Most of the credit goes to good organization and promotion, an interesting design program, strong support from vendors, a highly-regarded jury, and an unexplored medium. The international pool of entrants, using the web as both topic and medium, broke new ground in web-based presentation. The number of entries, the distributed review panel, and a narrow window of time for review all mandated development of an on-line system for managing the competition review process. Such a system, EVAL, was successfully developed and deployed using a custom AppleScript CGI application and the FileMaker Pro DBMS on a Macintosh web server.

While there is not as much direct feedback as we would like, we believe that this system demonstrated the value of an omnipresent evaluation console when navigating a complex set of web pages and interactions. Almost all reviewers were able to use the system without assistance from competition organizers. Only one juror, who was attempting to use WebTV to score projects, reported insurmountable problems.

On the whole, we believe that the use of an evaluation vocabulary mapping a "verbal" evaluation to a score was successful in providing some standardization amongst reviewers. Certainly some reviewers had higher average scores than others, and there were discrepancies between reviewers, as might be expected. However, the very highest scores were not overused, and there was often strong similarity in scoring by different reviewers.

The potential exists for review processes to ignite conversations amongst reviewers. This was almost entirely missing from this review, even though the evaluation system provided for both text and numeric evaluation. Reviewers had knowledge of each other, including email addresses, but did not (as far as we know) use this extensively. One juror went so far as to indicate that the lack of this more informal "conversational" interaction made it quite difficult to conduct the judging, and another was observed to have *removed* a comment from the scoring database. Whether the lack of interaction was due to the asynchronous character of the evaluation process, the "formalized" recording of comments into a database, or the sheer number of projects that needed to be looked at remains an open question.

Future development

EVAL represents a first approximation. Its strength lies in management of a formalized review process such as occurred in this competition. Further enhancements are being considered, including more complete "cradle to grave" support for the competition organizers, as well as enhancements to further simplify and streamline the evaluation process. In addition, an alternative, more democratic, "browser's choice" paradigm is being considered for future competitions. Finally, related systems are being developed which do more to directly foster informal designer interaction *during* design, rather than at the end of the process.

Availability

The distributable components of the EVAL system are available to others on the web at <http://www.caup.washington.edu/software/>. In addition to the EVAL files, you will need a Macintosh web server and the FileMaker Pro database management system to implement a complete setup.

References

ACADIA Competition web site, (1998). B. Johnson, (webmaster): <http://www.acadia.org/competition/>.

Anders, P. (1997). Cybrids: Integrating Cognitive and Physical Space in Architecture. Jordan, J.P., B. Mehnert and A. Harfmann (Eds.). ACADIA '97 (pp. 17-34), Cincinnati, Ohio: Berman Printing.

Chen, N., T. Kvan, J. Wojtowicz, D. van Bakergem, T. Casaus, J. Davidson, J. Fargas, K. Hubbell, W. Mitchell, T. Nagakura, P. Papazian (1994). Place, Time and The Virtual Design Studio. Harfmann, A. and M. Fraser (Eds.). ACADIA '94 (pp. 115-131), St. Louis, MO.

Kolarevic, B., G. Schmitt, U. Hirshberg, D. Kurmann, B. Johnson. (1998). An Experiment in Design Collaboration. T. Seebohm and S.van Wyk (Eds.), ACADIA'98 (pp. 90-99), Québec City, Canada: Universal Printing.

Martini, Kirk. in a private email exchange with the author.

Novitski, B.J. (1996). The Web Takes Off. *Architectural Record*, July 1996, (p 49). See also <http://www.caup.washington.edu/courses/arch499/>

Wenz F. and U. Hirschberg (1997). Phase(x) Memetic Engineering for Architecture. B. Martens, H. Linzer and A. Voigt (Eds.), ECAADE '97, Vienna, Austria.

Wojtowicz, Jerzy (Ed.) (1994). *Virtual Design Studio*. Hong Kong: Hong Kong University Press.