Blackboard 6 usage patterns and implications for the Universiteit van Amsterdam

Alan Berg, Victor Maijer, Frank Benneker

Central Computing Services, Universiteit van Amsterdam, The Netherlands.

A.M.Berg@uva.nl, V.N.Maijer@uva.nl, W.F.M.Benneker@uva.nl

Abstract

The Blackboard 6 system at the University of Amsterdam [8] is an E-learning ecosphere. 12,000 students regularly participate actively in studying via this online 24 hour a day system. Usage tracking over the last half year contains more than 28,000,000 hits and the content has a corpus of 112,000 files ranging over 2200 courses and costing a small 28 GB in storage space. This paper will step through the methods of harvesting analyzing and relating gross statistics to university policy. The important conclusions being that: (1) the Microsoft Office series of file formats is dominant. (2) Video usage is suppressed (3) University policy can have an impact on quality of content (4) We can track the impact of policy on content (5) We suggest improvements in our own infrastructure that make the tracking and data mining simpler in the near future.

Keywords: Blackboard, Statistics, Usage

1 Introduction

In the world of E-learning and course design a lot of effort has been put into research and ideas on how to organise and structure on-line course environments. However, most of these studies do not rely on empirical data from the actual use of a web based educational system. This papers intention is in part to help reverse this trend by describing a few relatively simple statistics. These statistics have the benefit that they can be tracked and used as a part of a set up that is a viable description of an improved infrastructure possibility for Blackboard. This improvement in infrastructure can deskill the finding of relevant and optimizing patterns. With so much data it is easy to overload the reader with the microscopic view. It is not our intention to drill down, but rather to zoom out and highlight gross detail and not by accident at the same time show the obvious potential of such manipulations.

The introduction of the Blackboard system as a virtual learning environment for the University of Amsterdam dates back in the prehistoric days of 2000. At this time it was a matter of policy not to impose any paradigms from the community of instructional design on the teachers that were going to use Blackboard. Teachers were supported on how to use the systems but also allowed to find their way in using the system and building courses and course materials in such way that suited their needs. A second point from the perspective of teacher and student is that the system is used in blended situation. The University of Amsterdam does not provide full distance education courses, though it may do at some undefined point in the future. As developers of idea's and infrastructure it is our responsibility to support the current and be prepared for change. Hence the data gathering evolution mentioned in this paper.

2 Methodology current and intended

The Blackboard 6 E-learning environment is a compound of a number of different subsystems. The main course material and metadata resides both physically on hard disk and to a degree ghosted within the database backend. Within the storage reside course content and log files and within the database is also a partial map of the usage information especially accumulated for the "Advanced Reporting System". The Advance Reporting System has a special accumulator table that describes in a wealth of detail user requests. For a policy researcher the important point to note is that databases are easier to query and find relationships in than storage that is not designed for this purpose. Therefore to make statistical analysis between content and usage without over taxing production system databases we took the novel approach of placing the file structure information within an open source, in this case MySQL database and then perform advanced queries. The details of these queries will not be presented here only a few germane examples. The methodology has the major advantages of flexibility and millisecond response times and all from within the context of a personal computer. The creation of such a system comprises a few simple stages: A directory listing of the content area is placed in a text file. This is processed by a small script and the data filtered into a database import file a grand name for what is basically a tab delimitated version of the directory listing. Finally the import is absorbed by a local database. This turns out to be a very effective and rapid process to obtain malleable file system information.

The database contains a table column named type which contains the file extension name of each file and another size that holds the size of the file in bytes. An example SQL query is shown next. It produces the results beautified in table 1. Please note that the constant represents the total storage used in bytes. This information was found by another one line query.

select sum(size)from blackboard; ➔ 298765379.89

select type,sum(size)*100/298765379.89 as ct from blackboard group by type order by ct desc limit 12;

Type	Size (%)
Ppt	**46%**
Pdf	18%
Doc	11%
Zip	6%
Jpg	3%
Pps	2%
Exe	1%
Other	13%

Table 1: Storage space usage per file type.

Please note that the "Other" category represents all types not mentioned (zip, xsl etc).

From just one query we can see that PowerPoint presentations take around half of the storage space up. Since storage is the cheapest component in the system this fact alone is not particularly relevant. The trick is to relate different data sources and gain insight of impact of specific policies.

Figure 1 is a mindmap highlighting the diverse threads that need to be weaved into a story. The first thread is that of an intended web interface. This interface should integrate information from disparate sources so that a functional manager can have an immediate and flexible oversight. The four main data sources processed are as described:

- **File system:** Harvesting simple directory listings into a database. It makes information amenable to manipulation.
- **Blackboard database**: This database has all the raw metadata over each content object. The database also has special accumulation tables that are directly translatable into usage reporting.
- **Log files:** The Apache/Tomcat log files are daily transported to a special log file analyzing server that generates usage reports. The processed accumulations are also stored in a database. The tool used at present is the commercial product Webtrends. However, if this proves to be difficult to communicate with programmatically then it is a practical step to migrate to one of a number of free Open source tools.
- **System monitoring**: At present at the gross level Blackboard 6 is monitored for CPU load, bandwidth utilization, memory usage via a custom in house tool. The data from this tool is transacted into another one of the multiple databases. And this again if required may be exposed to predefined queries.

The main theme of this method is that of placing information into a database. Databases are highly optimized search tools with simple API's allowing programmatic querying without complex customizations. For example the Java language has the JDBC interface [7] that hides implementation details away from the programmer. This implies construction effort for a unified interface is small and within the grasp of our development group.

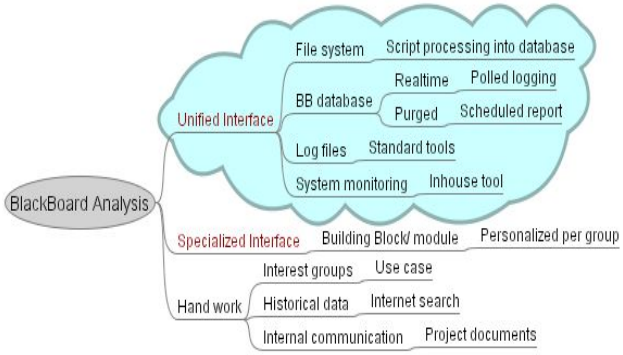

Figure 1: Mindmap of relevant and easy to get at data sources within the Blackboard infrastructure. Much detail is missing for the sake of readability.

Figure 2 is an architectural interpretation of thread one of the mindmap. Basically all databases are exposed to a uniform interface. The interface is intended to be web form based with the ability to perform well defined queries and if need be ad hoc open quieries. The cost in man hours is not the building of the interface. As stated this is trivial once the data sits in databases, but rather one of understanding relationships and building the correct correlating queries, the conceptual and statistical model.

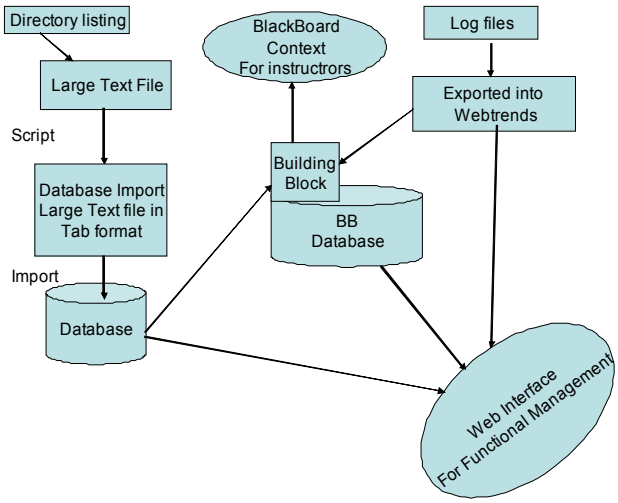

Figure 2: A proposed data harvesting infrastructure.

The second thread within the mindmap is that of a specialized interface. This interface is intended to work within the Blackboard context. Once an instructor logs onto the

Blackboard system there should be a special link to relevant course statistics. The authors have experience using the proprietary building block framework [4] and see it as a fundamentally simple task to enhance value with this method. A simple list of the top five popular files and the least can give the sort of feedback that pushes quality up. Numerous other examples come to mind: Warnings for when students have not visited the course or when a specific homework has not been clicked on as help. These pieces of information are available via simple queries. Zooming further out, the teacher can discover directly which of their courses are popular and which are not. In conclusion a specialized interface gives an almost tactile feel for the instructor over the details of what in specific is happening in their responsibility domain.

3 Example Results

This sections intention is to give you the reader the strong feeling that the design mentioned previously can easily fulfil expectations. This is achieved via example interpretations of specific data.

Table 2 shows that around 60% of all files within Blackboard content areas are of the office format. As previously seen this is also true for storage utilization. Powerpoint takes up the most storage and Word documents the most number of files as shown in figure 3. Video formats account for 7% of total number of files stored.

Native Office	68233 (61%)
Other	43030 (39%)

Table 2: Percentage of files in Office format as measured by number of files.

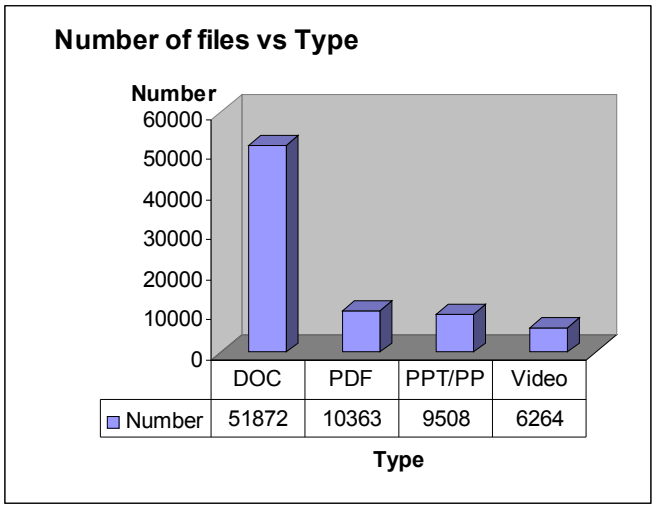

Figure 3: Total number of files vs. type

The gross number of downloads per type show that Word is prevalent. PDF format is competitive with PowerPoint and is about four times as compressed. Figure 5 shows the relative popularity of file types. On average for every PDF file the end users download 69 times where as word documents are roughly three times less popular. As 80% of all information to the human brain is seen through the eyes one would assume that PDF format is visually more appealing. The absolute popularity of video format is 30-40 times less popular than PDF. Zooming in on this issue via figure 6, one sees that it takes 15 times more time to download the average video file than the average PDF file. Relating this to bandwidth pipe size (table 3) one can see that it takes over five minutes to download a video fragment for the sort of modem that is standard on most laptops and for a large section of computer users at home.

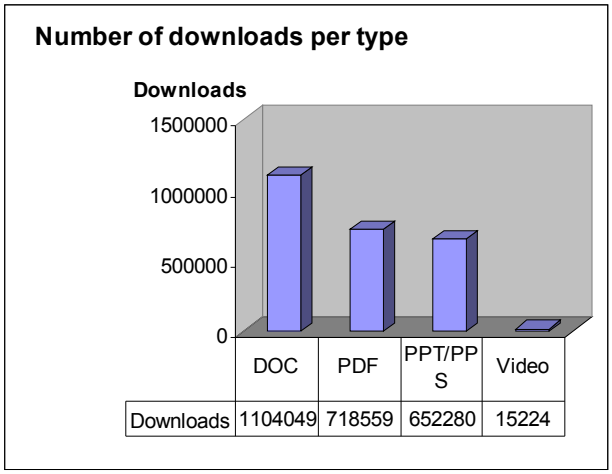

Figure 4: Total number of downloads vs. type.

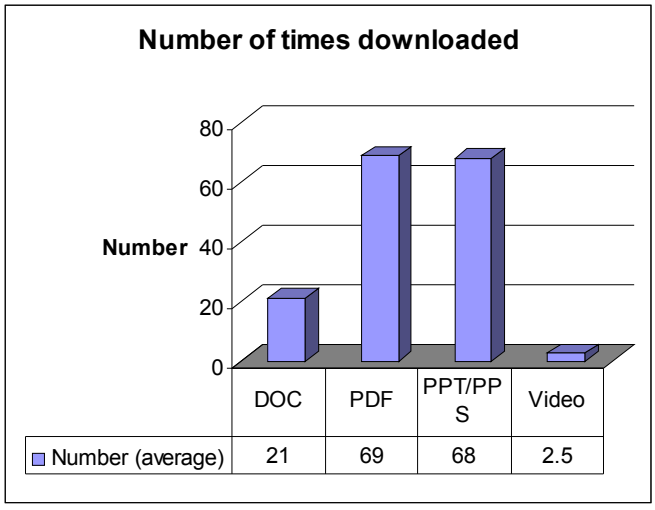

Figure 5: Average number of times each file type is downloaded

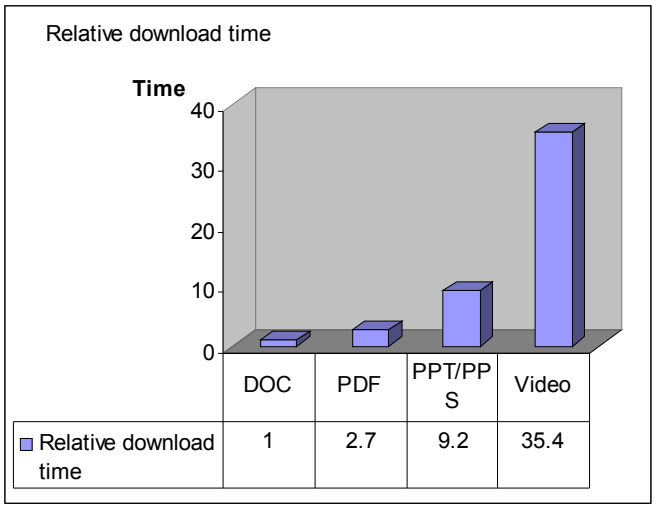

Figure 6: Relative download time per file type. It takes a video fragment 35 times longer to download than an average word document.

Type	Average download size (kilobytes)	Relative download time	Time in seconds to download for different sized bandwidth pipes (kilobits)		
			64 Modem	128 ISDN	1024 ADSL
DOC	70.3	1	8.8	4.4	0.5
PDF	190.1	2.7	23.8	11.9	1.5
PPT	648.7	9.2	81.1	40.5	5.1
VID	2488.7	35.4	**311.1**	**155.5**	19.4

Table 3: Table relating download times to per file type to type of Internet connection.

4 The effect of policy on quality

Figure 7 is an approximate force diagram for the improvement of course quality. It suggests that University wide initiatives are the extra impetus that is needed to improve the perception of quality for Blackboard over other delivery methods. Quality in this situation is relative to perception. Quality improves over time partially due to improving infrastructure and authoring applications. However that is true for content in general and not only for the Blackboard system. The extra force required to manage the change in perceptions is that of University wide initiatives. From the pedagogical point of view it must be mentioned that content is only one aspect of the process of learning. A lot of other learning activities such as group activities and improving academic skills are important parts of a complete learning environment. However it is in our opinion that improving the specific quality in which electronic learning environments handle content is also improving generic course quality and thus provides students with the desired improvement in learning experience.

Placing this thesis into our measurable framework one notes the following: In the last section we have shown statistically that video formats are not popular in the absolute terms of number of files on system and number of times each file is requested. We have related this fact to the average download time. However though the University recognizes this as a quality issue and tries to motivate content providers (teachers, instructors) and students with two projects: The first project is squarely focussed on students, to sell more bandwidth via cheap ADSL pipes [9]. The second is a streaming media project for content providers to deskill the creation of sources [5]. Streaming media is different from direct download as it delivers content on demand and at guaranteed bandwidth consumption. You effectively view as you download so the latency to the beginning of the viewers experience is significantly decreased. With our methodology we now have in place metrics that relate the effect of these projects on student usage of video content. The clicking of the streaming link adds an extra entry to the Blackboard database and this usage per course can be measured against course popularity in general. We can now zoom in decompose the elements and analysis.

The importance of the impact of the dominance of the Office format should not be underestimated. Another central computing services project is the U-drive project [1] that allows students access to storage space within and without the University. Part of the solution is to use web folders from within Office. These folders allow dragging and dropping of files over from office to a web server that uses the WebDav protocol [2]. It is possible to enable this feature within Blackboard. The authors expect the knock on effect of this project will be to set on WebDav in Blackboard. Then the student/instructor can drag and drop directly from there u-drive into Blackboard. Further the Student Homepage System [1] is not WebDav enabled, but now we have a reason to do so. The metrics from our methodology can track the various impacts of enabling on the learning environment

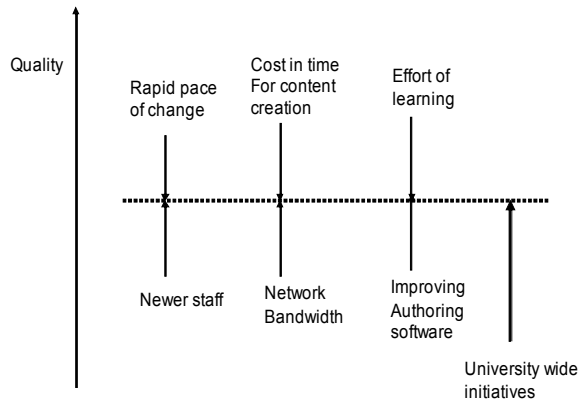

Figure 7: An evenly balanced force diagram where University wide initiatives may be the required extra impetus for quality improvements in content.

5 Summary & Conclusions

This paper highlights that moving relevant data into readily accessible databases and unifying views via a web interface will simplify the understanding of patterns. Further we have related pattern tracking to university policies as implemented via definite projects.

The authors have noted that it is also relatively trivial to use the building block framework to allow instructors to have immediate feedback on "how it is going". This feedback can only push up the general quality of courses.

Student reactions to the use the different content types are not a part of this paper. In the near future we hope to develop a model in which blackboard analyses as shown in figure 1 will be combined with statistical information based on student responses and appreciation. An example of such an outcomes based model is the work of Hassan M. Selim [10]. The development of a model will give us the opportunity to provide statements on the quality of a course from content perspective and will give us insight on how to improve the quality of blackboard courses at the University of Amsterdam.

In specific it was noted the dominance of the series of Office formats. This clearly defines the content creation era we find ourselves in. The authors humbly suggest that the use of webfolders within the Office application and enabling webdav in Blackboard may have a positive impact.

Video formats are clearly undervalued and University policy is focussing on this. Two projects that should have impact are the selling of better bandwidth and the deskilling of creation of streaming media content. Both projects should have a positive influence and will be tracked.

References

[1] Alan Berg, Frank Pinxt. "Building and securing large scale storage systems for students". Eunis 2004 .

[2] A Berg amd L. Raijmann "The building of a successful zero administration homepage system for students and its relation to the wider campus infrastructure". Eunis 2003 Proceedings Book pages 350-354

[3] Y. Goland, E. Whitehead, A. Faizi, S. Carter, D. Jensen, "HTTP Extensions for Distributed Authoring – WEBDAV" Request for Comments: 2518 "http://www.ietf.org/rfc/rfc2518.txt"

[4] Matthew Pittinsky. "*Blackboard Building Blocks* 2003 Overview White Paper". http://www.blackboard.com/common/docs/White_Paper_2003.pdf

[5] M. van Staalen. "Streaming media at the University of Amsterdam:. Eunis 2003 Proceedings Book page pages 308-313

[6] Davenport Open source project. "http://sourceforge.net/projects/davenport/"

[7] JDBC http://java.sun.com/products/jdbc/index.jsp

[8] University of Amsterdam "http://www.uva.nl"

[9] Uva-Dsl project. http://www.ic.uva.nl/uvadsl/

[10] Hassan M. Selim. "An empirical investigation of student acceptance of course websites". Computers & Education 40 (2003) pages 343-360