Differential Impact of Web Technology on Various Dimensions of Students’ Academic and Non-Academic Activities (a Case Study)

Rozita Jamili Oskouei1*, Mohsen Askari2 and Phani Rajendra Prasad Sajja3

1, 3 Computer Science & Engineering Department, Motilal Nehru National Institute of Technology, Allahabad, UP, India; rozita2010r@gmail.com, sajja.phani@gmail.com
2 Computer Engineering Department, Islamic Azad University, Ramsar, Iran; mohsen.askari65@yahoo.com

Abstract

Internet and computing have been advocated as one shot solution to many existing constraints and difficulties of education system. Our work aims to study exhaustively differential impact of Web technology on various dimensions of students’ academic performance, co-curricular, extra-curricular and non-academic activities. We analyzed the Internet usage behavior of students in terms of average time spent in Internet per day and category of visited Websites by them along with the effects of these Internet usage behaviors on their academic performance (CPI) and other co-curricular and extra-curricular activities. For this analysis we used proxy server access log files of an engineering college in India which are collected during 30 months continually. We have classified student activities (with extending ODP classification scheme) under four categories: Curricular, Co-Curricular, Extra-Curricular and Non-Curricular. Only 59% of female students used Internet continually as against 81% of male students. This analysis led us to conclude that a significant gender gap exists in terms of Internet usage.

Keywords: Website Classification Scheme, Open Directory Project (ODP), Internet Usage Behavior, Social Behaviors, Academic Performance, Navigation, Pedagogical Issues.

1. Introduction

The growing use of computing and communication technology has both good and bad effects. Kraut and his colleagues [1–2] claimed that greater use of Internet was associated with the negative effects on individuals, such as diminishing social circle, and increasing depression and loneliness. Also, many quantitative studies confirmed that loneliness was associated with increased Internet use [3–6]. High frequency and larger duration of Internet use which interfere with daily life have been linked to a range of problems, including decreased psychosocial well-being and relationship breakdown [7–9]. All of these studies lead one to hypothesize that a controlled use of technology with a purpose can result into increased productivity and better quality of life. However, if one becomes the slave of the technology, then it adversely affects both productivity and quality of life.

Many of academic community believe that, students use Internet and World Wide Web (WWW) as stress buster and are increasingly becoming addictive to it-to the extent that it hurts them. Some of studies [10–11] have been made to study the pattern of Websites usage by students and to validate the perception held by the academic community, but it will be unreasonable to generalize these results. These perceptions have prompted some of the academic
administrators to block access to all Websites which do not contribute directly to curricular activities as a remedial measure. Further, majority of students at different levels of education including high school and colleges have access to Internet and spend some minutes/hours daily on different category of Websites. Therefore, a growing number of researchers and academic coordinators believe that, it is important to provide perspectives about main effects of Internet usages on students’ personal and social behaviours along with impact of these usages on their academic performances. There is no significant research activity for these issues.

Most of the research studies have been focused to understand the behaviour of the users based on the usage patterns of their products and services with aim to in-crease either the customer base or profit or both. These studies did not focus on impacts of their goods services for their effect on individuals’ productivity, health and social behaviours. This is more evident in the case of student community. Not many serious studies have been made in the context of under developed and developing countries like India to understand the effect of Internet on learning experiences of students and faculty, their productivity and social behaviours. It needs to be emphasized that such studies are related to the economic, social and cultural environments of academic community. The results of study for one community from a geographical region should not be and cannot be generalized for communities from other regions.

Major contributions of this investigation are:

- Augmentation of classification hierarchy of Websites of Open Directory Project (ODP) (36) by concepts which are related to activities of student communities. The concepts which have been embedded in ODP include: Curricular, Co-Curricular, Extra-Curricular, Non-Curricular, Media, General, Professional, Undesirable, Adult, Webcam, Sharing Websites and Special Communities. These concepts are either generalization or specialization of existing concepts.
- The distribution of Websites used by student community at Motilal Nehru National Institute of Technology (MNNIT) Allahabad, India. The relationships between Internet usage behaviour and various activities are studied.

This paper is organized in seven sections. Section 2nd presents related works. Section 3rd shows our proposed Website Classification scheme. Section 4th describes data collection step. Section 5th presents a case study on the use of Internet. The results of analysis are presented in Section 6th. Finally, section 7th includes the conclusion of this research.

2. Related Works

2.1 Website Classification

Websites are created to serve specific purpose and functionalities. Search engines classify these Websites on the basis of their content, structure and functionality to assist the Internet users in getting the desired Information. Quality of classification is defined in terms of precision and recall. Precision is ratio of the number of correct categories assigned and the total number of categories assigned. The higher precision has the smaller amount of false positives. Recall is the number of correct categories assigned divided by the total number of known correct categories. Higher recall means a smaller amount of missed categories. In theory, scores of 1 is desirable for both precision and recall. In practice even human assigned classifications may only achieve scores between 0.7 and 0.9, depending on the classification task. This is because to some extent classification is a subjective task and there are usually “grey areas” in a classification scheme.

Earlier researchers [12–13] reported that lack of homogeneity in Websites based contents [14–17]. Dong [14] presented the hierarchical classification of Web content based on the combination of both textual and visual features. His results shows when limited texts are available, combined use of image and text may greatly improve the categorization performance than using single classifier. Tian [17] proposed a multi-scale tree structured representation model for Web sites. They used multi-level granularity to model Websites. Lindemann [15] explored relationship between structure and functionality of Websites. They considered eight functional classes namely Academic, Blog, Community, Corporate, Information, Non-profit, Personal and Shop. They also established relationship between functionality and Website structural contents. Contents of Websites, it is challenging to extract key features (topic) for the purpose of classification. Search engines are in general unable to aggregate features belonging to different part of the hierarchies.
2.2 Internet Usage Pattern Discovery

Extracting users’ behaviour pattern in Internet is an important and challenging research topic of Web usage mining. Web usage mining aims to discover interesting and frequent user access patterns from Web usage data. The discovered knowledge is very useful for modelling users’ web access behaviours. Several research efforts [10, 18–21] have proposed different algorithms for extracting Internet usage pattern of users or attempted to apply usage patterns on various applications. Butakov [10] studied the Internet usage behaviour of students at the American university of Nigeria and concluded that the Internet resources were not used for academic related requirements and were mostly used for Entertainment by both genders (female 60%, male 70%). The number of visited Website by male was more than female. This explained that female users spent more times on particular Websites, especially Social Network Websites. Bidgoli [18] predicted the students’ final grades based on their Web-use features, which are extracted from the homework data. They improved the prediction accuracy at least 10%. Machado [20] studied approach for the evaluation of educational site design as a Web Usage Mining (WUM) application. Merceron [22] discovered pedagogical relevant knowledge contained in databases obtained from web-based educational systems and proposed a model based on number of exercises undertaken and mistakes committed by students in their completion to classify them under different categories. Lin CH [23] explored gender differences in adolescent Internet accessibility, motives for use, and online activities in Taiwan. Their results revealed no gender difference in subjects’ view of the importance of television and the Internet.

Several researchers studied gender based effects of information technology on education [24–29]. In recent research about gender differences with respect to technology, Welty [24] found that women opted for communication and medical technologies more frequently than men. Jo Sander [25] has presented an excellent survey of relationships between technology in education and gender. One of the major observations in her study was that women are significantly underrepresented in Information and Communication Technology in most of the countries. Huynh [26] presented a focused gender based group study on perspective and experiences of female management students and their instructors in an online learning environment that used computer supported collaborative learning (CSCL) system to facilitate student-to-student interactions. Kim & Chang’s [29] studies has concluded that computer usage have differential effects on academic performance of users from the immigrant and gender groups. It advocated adaptive design of multimedia contents for different groups of students addressing their specific needs.

3. Our Proposed Website Classification

Open Directory Project (ODP) [30] uses hierarchy of topics for classification of Websites. Websites are categorized in terms of these topics and they may belong to one or more topic. Most of the search engines such as Google, Netscape, AOL and etc. use ODP scheme for their search. ODP data is made available for open content distribution under the terms of the Open Directory License, which requires a specific ODP attribution table on every Web page that uses the data. The Open Directory License also includes a requirement that users of the data continually check the ODP site for updates and discontinue use and distribution of the data or works derived from the data once an update occurs. This restriction prompted the Free Software Foundation to refer to the Open Directory License as a non-free documentation license, citing the right to redistribute a given version not being permanent and the requirement to check for changes to the license. To evaluate the suitability of ODP classification scheme to model the behavior of students’ community, we analyzed Web access log files of Motilal Nehru National Institute of Technology (MNNIT) Allahabad, India for a period of 12 months. Five thousand two hundred ten students, who were authorized to use Internet, visited approximately fifty thousand (~49376) Websites during this period. Based on this analysis the dominant visited categories were Society, References, Business and Computer.

Since the ODP categories to which Websites visited by students belong, do not relate to activities of academic environments. We need to have concepts in the classification scheme which explicitly relate to activities of students in a residential academic institution. Students’ activities are generally classified as curricular, co-curricular, extra-curricular and non-curricular.
These categories have been introduced based on the analysis of contents of the popular Websites visited by students. The contents which are related to course and were mostly grouped under science, computer and references. For example 75% of the visitors to the popular Websites visited these Websites for downloading different tools such as browsers, antivirus or software packages and etc. Accordingly we decided to have a generalized concept curricular for these three categories. The analysis of access log files indicated that approximately 3 to 4% of 27% of the visited Websites belonging to category Society were primarily for video communication. The examples of these Websites were Skype, OOVOO and etc. consequently we decided to have Webcam as a sub-category in Social and Communication Network. Similarly there are many Websites which are primarily used for sharing text, video, music or multimedia files and etc. Presently they are classified under general sub-category of Art in terms of music, video and etc. We thought that it will be appropriate to make a new category for such Websites and designated as a part of Social and Communication Websites. Further, special Communities include those Websites which are primarily used for communication between professionals. For example researchgate.org is a Website for scientists for professional discussions and communication between them. These categories are based on either related activity or usage. The Undesirable category is introduced to capture social acceptable norms. Terrorism, Criminals and Drugs related Websites are undesirable in most of the societies. However, the Adult Websites may not be in the same category. In the context of Asian society including Indian society we consider to include under Undesirable category.
Differential Impact of Web Technology on Various Dimensions of Students’ Academic and Non-Academic Activities 
(a Case Study)

Figure 1 gives the distribution of the Websites visited by students in MNNIT as per augmented ODP categories. It is evident from this table that, the majority of visited Websites (35%) belong to Extra-Curricular and Sub-category of Social Network Websites (Society) and not to curricular Websites. It is heartening to note that, there is small percentage (less than 1%) who visit drugs related Websites. Further, there is a small percentage (~10%) of students both male and female who visited Adult Websites.

4. Data Collection and Pre-Processing

For this analysis, we collected the proxy server’s access log files from the computer center of Motilal Nehru National Institute of Technology Allahabad, India which runs on “Squid 3.0 stable”. The data collection period was 30 continuous months, which includes five academic semesters. These log files are used for debugging, user and site profiling and measurement of resource utilization. The computer center also provided user identification information, which includes user-id, full-name and department in one text file. Further, users’ academic information are collected from the dean (Academic Affairs) office in one Microsoft excel file. The fields which are included in this excel file were: registration-no., full-name, program, branch, semester, gender, Cumulative Performance Index (CPI) and SPI (Semester Performance Index).

I was permitted to use the pre-processed and filtered contents of the log files for the research purposes. The pre-processing hides the actual identity of students and replaces Virtual-Id instead of real User-Id.

For this research, the experimental training set of data including the access log files of students undertaking B.Tech. Degree program is collected during the period of 30 months continuously. The simple random sampling (without replacement) is used for creating training set. The number of samples was 1200 individual users’ accounts.

Each record of the access log file has 11 fields for each visited Web page by one individual user. For our analysis, we have selected three fields out of these 11 fields of each record. These three fields are: user-id, date & time of connection and the URL of the visited Web page. For statistical analysis, we used Rapid-Miner and R-Miner tools. The records with inconsistent or missing values are removed during the pre-processing step. An algorithm is developed and implemented to compute the time spent by users on different categories of Web-sites along with the total time spent per day by him/her on Internet by using time stamp of the access log files. Minute is used as a unit of time duration to measure the time spent by one individual user on each Website during a day. This unit of measurement implies that all durations less than a minute have been rounded to one minute.

5. Internet Usage Behaviors: Case study

In this investigation, exploring included gender based differential usage patterns of Websites during 30 months included all working days, holidays, examination weeks and a day before examination weeks.

5.1 Discipline and Gender Based Differential Websites Visit Patterns

5.1.1 Internet Usage: “Regularity”

We classified all the visitors to Internet under four categories: Authorized, Regular, Non-Regular and Internet-Absentees. Authorized users are students who have login-Id and pass-word whereas Regular users are those authorized users who visit Internet every day. Non-Regular users do not use Internet every day. They use it infrequently. Internet-Absentees are group of users who never connects to Internet. Our analysis results of gender and degree program wise distribution of users show that:

- Female users are in minority in all degree programs and in all categories.
- Post-graduate degree programs have better female representation as compared to under-graduate programs.
- M.Tech. degree program is an exception as it has lesser percentage of female users.
- There are no regular users from MBA and M.Sc. degree programs.
- It is surprising to note that such a large percentage (22%) of Authorized users did not use Internet even though the facilities are available in all major units (including hostels) of the institute. This percentage of Internet Absentees is very high for female students. This leads one to believe that either there is some inhibition on the part of female users or the computing resources or their accessibility are not convenient to them.
- 34% of female and 7% of male from Authorized users used Internet infrequently.
• Both female and male users belonging to infrequent user group preferred to use Internet one day per week. Within infrequent Internet users, 59% of female and 37% of male have used Internet one day per week.

This analysis led us to conclude that a significant gender gap exists in terms of Internet usage. Only 59% of the female students used Internet as against 81% of male students. Similarly, students of post-graduate programs use Internet very sparingly compared to other programs. For example, very high percentage of authorized users from M.Tech. and Ph.D. degree programs are regular users.

5.1.2 Internet Usage: “Duration”

We analyzed the log files to discover the Internet usage duration and the time of connection in 24 hours of a day. It is evident from these results that the female which belongs to both Regular and Non-Regular groups spent significantly more time on Internet as compared to their male counterpart. Further, the number of pages visited by Regular male users is more than the female users. However, this is not true for non-Regular female users. In other words, female users spent more time per page compared to their male counterpart.

Table 1 shows gender and degree program wise analysis of Regular users’ average time spent and number of visited Webpage on various days in a semester including holidays, regular working days, and a day before examinations and final examination week. In this table T, P, F and M stand for average Time spent on Internet, number of visited Web pages per day, female and male users respectively.

Table 1. Degree Program and Gender wise Average Time Spent & Number of Web page Visited by Regular Users

From Table 1, we conclude:

• Minimum usage time and minimum number of visited pages during semester belong to a day before examination for all degree programs.
• In all degree programs, time spent on Internet and number of visited Web pages by female users decreased on holidays compared with working days whereas this is reverse for male users.
• Average time spent on working days on Internet per day by female users under all degree programs were more than male users except M.Tech. degree program.
• For all degree programs except MCA, users’ time spent on Internet during examination week was more than regular working days during semester.

• Maximum number of visited Web pages during semester by users on all degree programs belongs to examination weeks.

Further, the category of visited Website by Regular and Non-Regular users along with their time spent on these Websites justify that the dominant use of Internet is not for curricular activities. 77% of the visited Websites belong to Co-Curricular, Extra-Curricular and Non-Curricular categories. Only 23% of visited Websites belong to curricular category. Further, female users spent on an average 280 minutes per day on Websites belonging to Extra-Curricular Websites category whereas male users spent on an average 140 minutes per day on Co-Curricular Websites. Compare to these duration the time spent by female and male students on Websites belonging to Curricular category were 17 and 25 minutes per day respectively.

Some of these observations are also valid for non-Regular users. For example, female users who visited Internet only one day a week, spent 60 minutes per day on an average on Websites belonging to Extra-Curricular Websites when Compare to their male counterparts who spent 20 minutes on an average on Websites belonging to Co-Curricular Websites. These results are very similar to the results published which belong to different geography and culture [10]. It employs that students in general, irrespective of social and cultural background have similar behavior pattern when it comes to Internet usage.

We also analyzed time of connection during 24 hours of a day for both genders of the users in interval 24 hours during semester, examination weeks and a day before examination weeks. It is surprising to note that number of male users during very early hours [1 a.m.–3 a.m.] is comparable to the number of male users from [2 p.m.–6 p.m.] of the day. In contrast, there is no such surprise for female users. However, this pattern is absent on a day before examination and examination week. For both these periods the maximum numbers of users are between [6 p.m.–10 p.m.].

5.2 Outliers

The term outlier has been defined in different ways for different context. Some of the definitions are given below.

• “An outlier is generally considered to be a data point that is far outside the norm for a variable or population” [31–33].
“An outlier is an observation that deviates so much from other observations as to arouse suspicions that it was generated by a different mechanism” [34].

“Outliers are values that are dubious in the eyes of the researcher” [35–36].

In the context of our study, outliers are individuals whose activities, skill sets, features and extent of engagement to commitment are different from the majority of the members of the community to which she/he may be belong. Since the majority of users belongs to the B.Tech., we are concerned on detecting outliers among these users. The distribution of B.Tech Regular-users based on their average time spent on Internet is shown in Figure 2. We identified different ranges of average time spent based on discontinuity of the curve. These ranges are circled in the following figure.

We created clusters using k-means and density based clustering with applying DBSCAN method. For k-means clustering method k was varied from 3 to 10 whereas for density based clustering e was varied from 1 to 20 minutes. We looked for a threshold value of average time spent on Internet to detect outliers, the cluster 5 and 6 for k = 6 (Table 2) look very reasonable candidate outliers. Same results were appeared with density based clustering. Therefore it looks reasonable to define threshold of average time spent on Internet as 457 minutes for detecting the outliers based on their daily time spend on Internet. The number of outliers with this threshold is 156 Outliers out of 2473 B.Tech Regular users. In other words, 6% of total Regular users were outliers based on higher time spent on Internet. The categories of visited Websites by this group of users (Outliers) are presented on Table 3.
The relationship between outliers based on CPI and average time spent on Internet is given on Table 4. In this table, academically weak students are indicated by CPI<=4 and academically excellent students are shown by CPI>=9.7.

Table 4 shows that:

- Only 1% of outliers based on their time-spend on Internet were academically weak students (CPI<4).
- 22% of total outliers were academically excellent students. It implies that we cannot generalize that more time spent on Internet lead to better academic performance.

We analyzed records of students who participated on technical, cultural and sports events organized by the institute. These analyses were made to investigate the relation between outliers and their different activities. Based on our analysis results, it is interesting to note that:

- All participants on co-curricular activities were regular users of Internet and none of female and male participants on co-curricular activities were outliers based on average time spent on Internet. The category of visited Websites by participants on Co-Curricular Activities is shown on Table 5.

Extra-curricular activities consisted of cultural and sports activities. The total number of female and male students in this group was 69 and 196 respectively. Out of these female and male participants in extra-curricular activities, 87% and 82% respectively were regular users. In other words, 60 out of 69 female students and 160 out of 196 male students who participated in extra-curricular activities were regular users. 3 out of 60 female participants and 20 out of 160 male participants belonged to our described Outliers (Time-spend>457).

The category of Visited Websites by Participants on Extra-Curricular Activities shows difference of these students with other students who participated in co-curricular activities are, this group of students spent more time on Websites belonging to co-curricular and extra-curricular activities. Moreover, 75% of their time had been spent on Websites belonging to those categories. With gender based analyzing all participants on Extra-Curricular activities including sport and cultural we can conclude that outliers with respect to the average time spent on Internet more than 457 minutes [both genders] participate minimally in these activities. In other words, higher average time spent on Internet did result into non-participation in other activities which are very essential for the growth of these students. This non-participation in these group activities may prove to be an indicator for loneliness of individuals [3–6].

### 6. Results and Discussion

In this investigation, we argued that there is a need to validate the perception that Internet infrastructures are not being used for the purposes for which they were established and their usage have determinant effects on academic performance of students’ community. As a part of this study, we explored students’ interactions with Internet, their participation in co-curricular or extra-curricular activities and their academic performance.

We analyze reasons for large Internet absentees. One of the reasons could be that, the Authorized students who do not reside in the institute campus do not have...
access to institute Internet infrastructure from outside of the campus and they are unable to use during regular working hours of the institute due to lecture and practical classes. This is especially true for MBA and MCA and M.Sc. degree programs. However, it does not explain the Internet absentees from B.Tech. students as all of them reside in the campus and have Internet access in the hostels. Our analysis of IP addresses revealed that some of the students impersonated their teacher and used their User-Id and password to access to Internet. However, this number was very small.

41% of Authorized female students falling under category of Internet absentees should be a matter of concern for academic administrators. There may be two dominant reasons for absentees. The IT infrastructure and their accessibility to female students both in the hostel and in the institute are limited. Hours of accessibility in the institute may be more limiting their adequacy. The second reason personal inhibitions or resistance to adapt to technology due to cultural, social or religious beliefs and traditions. Many of the Authorized students from Internet absentees have very low CPI. It is very tempting to hypothesis that one of the reasons for the poor academic performance could be their absence from the Internet usage. Very large number of students belonging to some of the post-graduate programs such as MBA and M.Sc. either did not use Internet or used very in-frequently. This may be due to different orientation of their academic programs or evaluation scheme. The curriculum contents may not be compelling these students to use Internet.

All our observations are based on analysis of data from one of the technical institutions. It will be inappropriate to generalize these observations to entire students’ community belonging to different institutions. At best, these results may be applicable to approximately twenty National Institute of Technology in the country as all of them have similar infrastructure and student population. Since each institution have their distinctive features in terms of admission policy regarding to gender in academic degree programs and in Internet infrastructure and also in their location. All of these factors contribute to behavioral aspect of students vis-a-vis Internet. Therefore, it is important to validate these observations from the data from different heterogeneous group of institutions.

One of the primary motivation for undertaken this study was to identify outliers in the students’ community who are unable to cope up with academic and environmental stress and strain. For this purpose one needs to collect more input about other students’ various activities in addition to their Internet related activities. The additional input could be network of friends and their Internet and mobile usage patterns or their visits outside of the campus for different personal needs and etc. It is desirable to have a comprehensive catalog of activities and to collect data about them so that pro-active detection of outliers can be performed.
7. Conclusion

Higher average time spent on Internet did result into non-participation in other activities which are very essential for the growth of these students. This non-participation in these group activities may prove to be an indicator for loneliness of individuals [3–6].

8. Acknowledgements

I am thankful to my Ph.D. thesis supervisor, Professor B.D.Chaudhary for helping me in preparing this paper. I also acknowledge the help rendered to me by the staff of computer center and Dean (Academic Affairs) office.

9. References


