Comparative Quality Analysis of Bangladeshi University Websites

Md Kawsarul Islam

Department of Computer Science and Engineering

Chittagong University of Engineering and Technology (CUET)

Chittagong-4349, Bangladesh

Email: kawsar.cuetcse@gmail.com

ABSTRACT

In the age of internet and today's developed academic societies, universities have recognized the importance of having websites to be able to benefit from appropriate academic communications. We can say, student's satisfaction from a university website shows to some extent university's web success. The most demanding factor for a university website is good quality. The purpose of this study is to determine the quality of university websites. The system we have developed can do comparative quality analysis among top thirty-five Bangladeshi universities. As our system is highly dependent on data, we have crawled data from university websites using python scrapy. We have considered five attributes. The attributes are performance of website, loading time, contents of information, contents of faculty information and updated notice in homepage of the website. Based on five attributes our developed system evaluates a university website. After evaluating all selected university websites, the system does a comparative analysis among them. Finally, generate a ranking of university websites.

Keywords: University websites, data, performance, contents, faculty information, updated notice, loading time.

I. INTRODUCTION

University websites could be used as a medium to give necessary information to the public through better design and intense delivery of content. These websites are good sources of reference for general information about the university and thus can be used to empower users to learn about the university easily and also provide different facilities to the students. A university should give all information about their university in their website. In this project, we have tried to evaluate the quality of Bangladeshi university websites and also do comparative analysis among them. There are many existing website evaluation methods and criteria for evaluating websites quality are not able to sufficiently assess the quality of university websites, and most of them focus on usability and accessibility, html page, asthetic design, page size but not focus on the contents of information those are helpful and convenient for both students and teachers. The system we have developed, collects data from thirty-

five public and private university websites using web crawler.

We have developed a system that can:

- Analyze five attributes (Performance of website based on accessibility and speed test of pages, loading time of website, contents of information, contents of faculty information, updated notice on homepage).
- Data of performance of website and loading time are collected using Pingdom API. Contents of information, contents of faculty information and updated notice in homepage are collected using crawling data and then preprocessing them.
- After preprocessing data assign numerical value on each attribues.
- Finally, perform comparison and generate ranking.

The rest of the paper is organized as follows: Section II provides a brief review of related work. Section III discusses in details about the system architecture and design. Section IV presents implementation. Section V concludes the paper and recommends the future directions of this work.

II.   RELATED WORK

There are many related works and methodologies for quality evaluation of university websites. We have found many publications which are related to my work. We are presenting some of these in the following:

In [1] they think that there are at least two factors representing university websites; accessibility and usability. They tested three criteria of accessibility and usability. The three criteria's are

stickiness, backlink, and web page loading time. To determine the actual results of accessibility and usability, they used fuzzy number to calculate the results.

In [2] they presented an empirical survey to measure the relative performance of university website in Iran. Their study used a standard questionnaire and distributed it among 387 active website users and the results were analyzed using structural equation modeling. They considered four factors including efficiency, accessibility, achievement and security on customer loyalty via two variables of trust and satisfaction.

In [3] their study evaluated 20 selected university websites in Bangladesh from the usability perspectives. They used two online automated tools (html toolbox and webpage analyser). They also used a questionnaire and distributed it among some of the graduated students of selected universities. The questions were classified into five categories.

- Content, organization and readability.
- Navigation and links.
- User interface design.
- Performance and effectiveness.
- Educational information.

In another part of the evaluation, two automated evaluation tools (HTML Toolbox and Webpage Analyzer) were used to assess website usability. By using HTML Toolbox, they measured download time, HTML check and repair and browsers compatibility. By using Web Page Analyzer ,they examined the internal attributes of the websites including HTML page sizes, total number of images, total number of HTML files and other relevant items of websites. The objectives of their study were to identify

criteria for the evaluation of academic websites and find out the weakness of the university websites and give suggestions for improving the usability.

In [4] they proposed a methodology for evaluating the quality of Malaysian University websites. They used the following criteria to evaluate the quality of website. The criteria are load time, response time, page rank, frequency of update, traffic, design optimization, size, number of items, accessibility error, markup validation and broken link. They used web diagnostic tools to analyze data. They analyzed top 5 ranked Malaysian University websites.

In [5] their research aim was to determine the usability performance of university website in Indonesia. They ranked university websites in Indonesia on the basis of usability performance. Their research result ranked the university website using VIKOR method. They worked with 30 university websites in Indonesia. They collected data 15 times during three months and to collect data they used website tools. After collecting data, they calculated to get rating result. They determined performance based on six factors load time, response time, page size, number of request, markup validation and broken link. To determine load time, page size and number of request they used Pingdom GT Metrix, for response time they used Site Speed Checker GT Metrix, for markup validation they used W3C Validation Achecker, for broken link they used W3C Linkchecker Deadlink Checker.

In [6] they evaluated university websites from students perspective. They rated the quality of university websites in four areas. They are,

- Informational content.
- Transactional content.
- Website design.
- Students satisfaction.

They assessed the quality of top ranked five university websites in United Arab Emirates. They conducted their survey using questionnaire. They selected a group of 50 final year students for their questionnaire. For informational content they used the following factors: strategic goals and mission, personnel information, continuous education and training, research activities, accreditation, academic programmes, future jobs, student services, online library facilities, web mail service, online learning support system, universities information of privacy, current news and events. For transactional content they considered the following factors: search option, online library facilities, enrolment facilities, payment, communication and learning facilities. For students satisfaction, they considered the following factor: easy to use, effective, efficient, fast, guide (online help, documentation and messages), quality of information should be updated and expected.

III. SYSTEM ARCHITECTURE AND DESIGN

The system architecture of the system comprises four basic modules. The first module is data collection module, the second module is data preprocessing and string matching module, the third module is analysis module. The fourth module is an output module. The function of data collection module is to crawl data from pages of the websites and store them in list of python. The function of the data
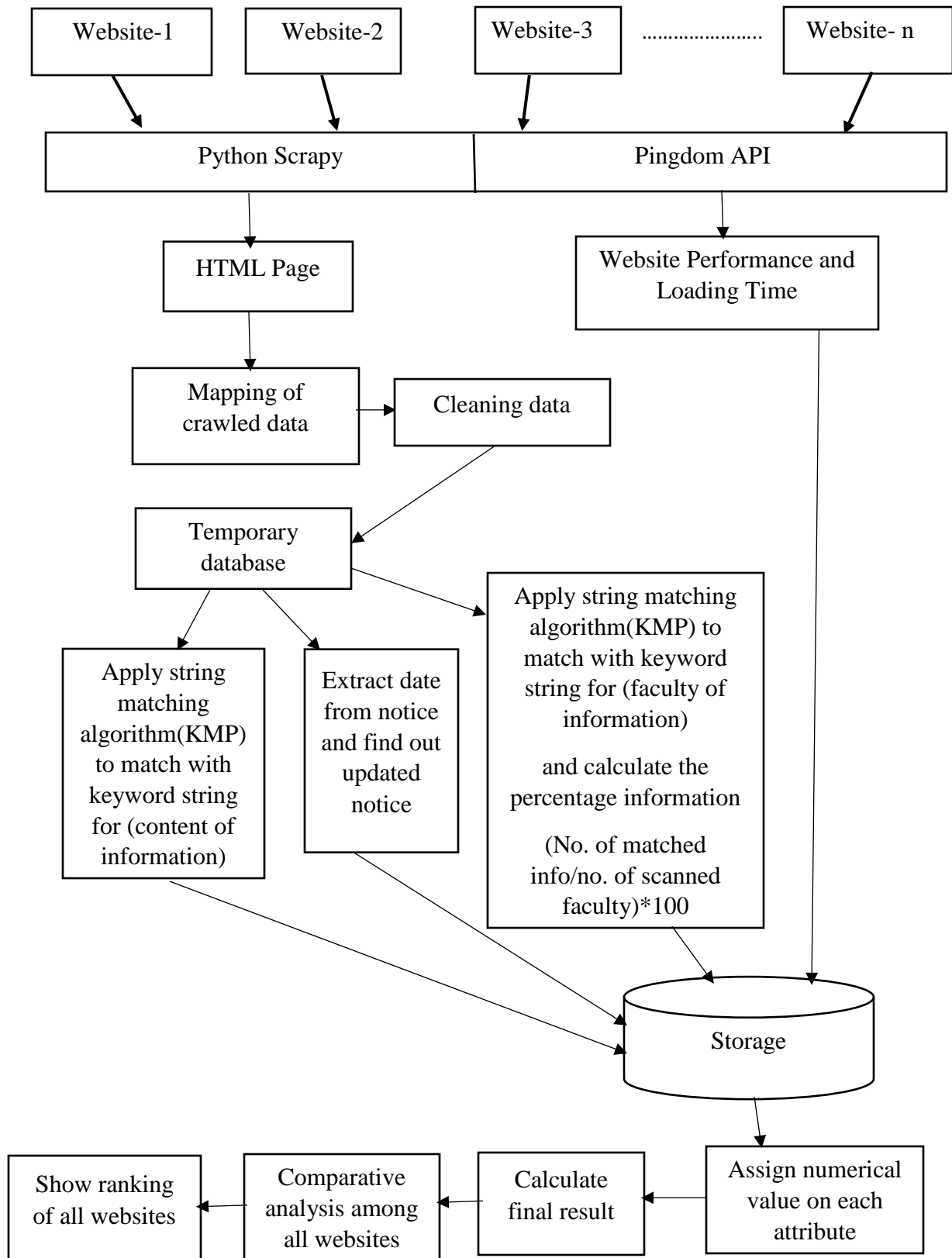
Fig 1. The system architecture of the proposed system

preprocessing and string matching module is to remove duplicate data and also match strings with keywords. Database system stores all the information. The third module analyzes data with normalize equations. The fourth module shows the ranking of university websites and also show the score of individual attributes. The architecture is shown in figure 1.

A. Data Collection Module and Preprocessing Module.

Data collection is the procedure of gathering and measuring information which is an established and systematic way. The system collects data from university websites using crawling process.

Algorithm (Crawl data of contents of information).

1. **Begin.**
2. **Input** URL.
3. **Request** the HTML page.
4. **Parse** the HTML page.
5. **Find** all 'a' tag.
6. **Find** all text from 'a' tag.
7. **Remove** Unnecessary element.
8. **Store** in temporary database.
9. **Read** key string from CSV file.
10. **Apply KMP algorithm** to find out matching string from temporary database and store matched string in main database.

We have considered 25 types of information as key string. Each type of information are scored (35/25). They are followings: Access to web mail services, university's strategic vision, mission and goals, academic member and staffs, administrative staffs, workshops and seminars, research activities, hosting conference, publications, academic programs offered, information about job and career opportunity, accommodation service, transportation service, online library facilities, online learning support system, online course registration system, online application system, publication of current notice,publication of current news and events, information about alumni, automation system (Hall automation, academic automation), annual reporting of university, academic rules and regulations, information about IQAC, library digital repository, academic calender.

Algorithm (Crawl data of contents of faculty information).

1. **Begin.**
2. **Input** URL.
3. **Request** the HTML page.
4. **Parse** the HTML page.
5. **Specify** the faculty information.
6. **Find** all 'a' tag of faculty member.
7. **Find** text from 'a' tag.
8. **Remove** unnecessary element.
9. **Store** in temporary database.
10. **Read** key string from csv file.
11. **Apply KMP algorithm** to find out the matching string.
12. **Count** no. of scanned faculty.
13. **Count** no. of matched string for each faculty information type.
14. **Calculate** (result of 12/result of 13)**.**
15. **Store** result in main database.

For faculty information we have considered 5 types of information as standard for a website.Those are the followngs:

- Faculty member list.
- Information about faculty member's publications.
- Information about faculty member's qualification.
- Information about faculty member's field of interest.

- Information about faculty member's experience.

Algorithm (Count Updated notice in homepage)

1. **Begin.**
2. **Input** URL.
3. **Request** the HTML Page.
4. **Parse** the HTML Page.
5. **Specify** the noticeboard.
6. **Find** all notices.
7. **Extract date** from notices and find updated notice for a given time range.
8. **Count** updated notice.
9. **Store** the result of count.

Algorithm (Website Performance and loading time)

1. **Begin.**
2. **Call** Pingdom API.
3. **Input** URL of website**.**
4. **Access** data in JSON format**.**
5. **Extract** data of Performance and Load Time from JSON data.
6. **Store** data in SQLite database.

The performance are determined using pingdom tool based on accessibility and speed of the pages of the website. Performance of website depends on several factors. A few of the categories are mentioned bellow:

- Fewer HTTP requests.
- Use of cookie-free domains.
- Add expires headers.
- Reduce DNS lookups.
- Compressed components with gzip.
- Avoid URL redirects.
- Avoid empty src or href.
- Put Javascript at bottom.
- Avoid HTTP 404 error.

- Make favicon small and cacheable.
- Configure entity tags.

For example, DU website avoids empty src or href: 100 %, reduces DNS lookups:70%, compresses components with gzip:56% , avoids URL redirects: 20 %, expires headers: 0, uses cookie free domains: 0, makes fewer HTTP requests: 0.

B. Data Analyzing Module

Maximum score for a website is 100. Here is score distribution for each attribute:

- Contents of information: 35
- Contents of faculty information: 20
- Updated notice in homepage: 10
- Performance of website: 25
- Website loading time: 10

Total score for the attribute contents of information for each website = (Total count*35)/25.

For example, in BUET website contents of information 88% (total count: 22) and score for this attribute: (22*35)/25=30.8.

Total score for the attribute contents of faculty information for each website = (Average value*20)/100.

Average value = summation of percentage value of all factors of contents of faculty information/5.

For example, in KUET website information about -

- Faculty member list: 100%.
- Faculty member publication: 80%.
- Faculty member qualification: 82%.
- Faculty member experience: 82%.
- Faculty member's field of interest:78%

For KUET website, average value: 84.4% and score for this attribute 16.88 out of 20.

Total score of the attribute, updated notice in homepage = 10; if count is greater or equal 10. Otherwise, score = total count.

For example, updated notice in homepage of RUET website in August-2019: 8 (total count in time range 30 days). So, score for this attribute: 8

Score of the attribute performance of website = performance in percentage/4.

For example, performance of CUET website is 69%. So, score for this attribute 17.25 out of 25.

The quality standard of the website loading time must be less than 30 seconds. If loading time greater or equal than 30 second. Then, score for this attribute is 0. Otherwise, score will be [10-(loadtime)/3].

For example, loading time of NSU website is 4.88s. So, score for this attribute:

$$(10-(4.88/3)) = 8.37 \text{ out of } 10.$$

Final score for a website = Summation of all individual scores.

C. Output Module and Comparative Analysis

In Output Module, the system shows the analyzed data in chart form. We have added two features in the output module:

- University website ranking.
- Details of each attribute for each university website.

Comparative analysis among different university websites are shown in figure 2,3,4,5 and 6 in context of different attributes.
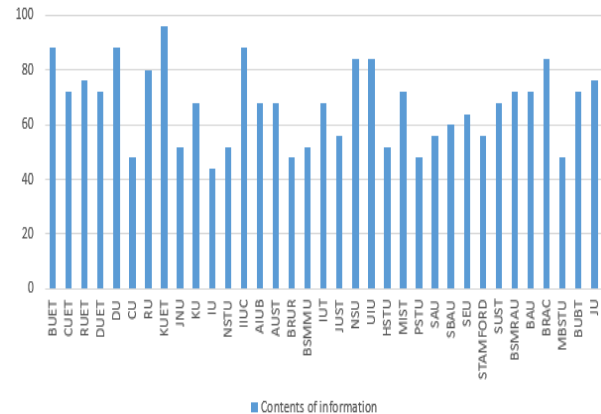

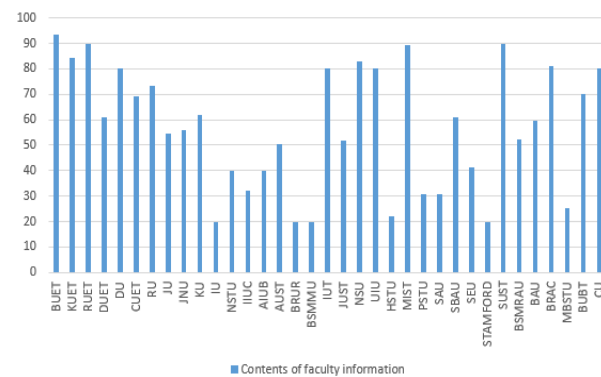Fig. 2: Comparative analysis in context of contents of information.

Fig. 3: Comparative analysis in context of contents of faculty information.
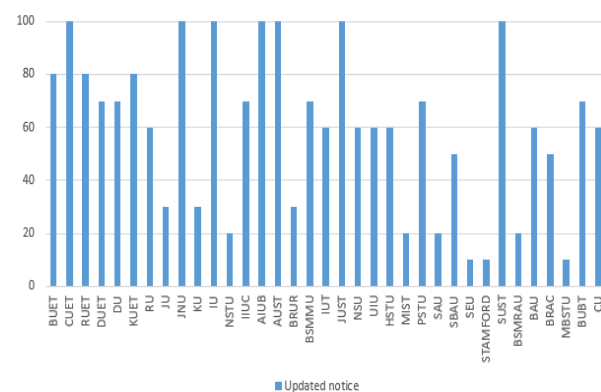

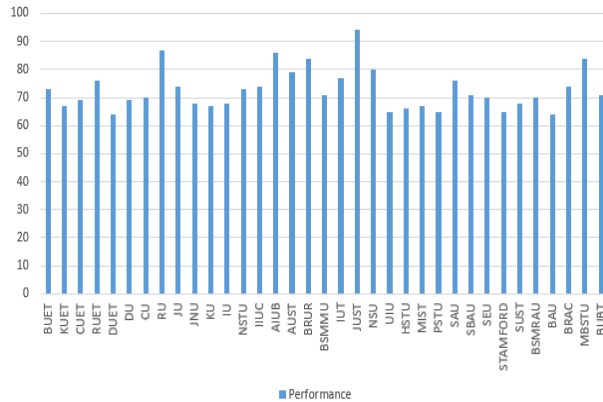Fig. 4: Comparative analysis in context of updated notice in homepage.

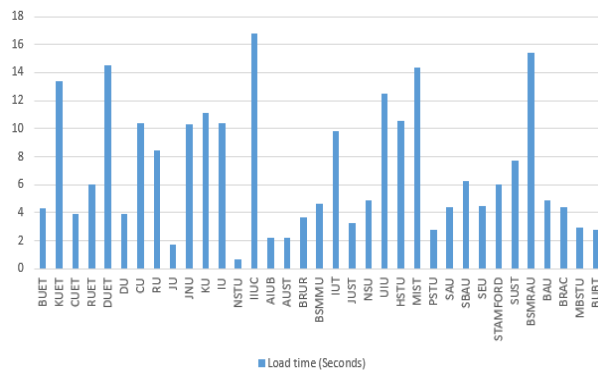Fig. 5: Comparative analysis in context of Performance of website.

Fig. 6: Comparative analysis in context of Loading Time of website.

University Website	Total score
BUET	85.25
KUET	80.76
NSU	80.37
DU	80.33
RUET	79.51
BRAC	77.92
RU	77.57
SUST	76.16
CUET	74.98
UIU	73.48
BUBT	73.07
AUST	72.82
AIUB	72.55
JUST	72.4
IUT	71.77

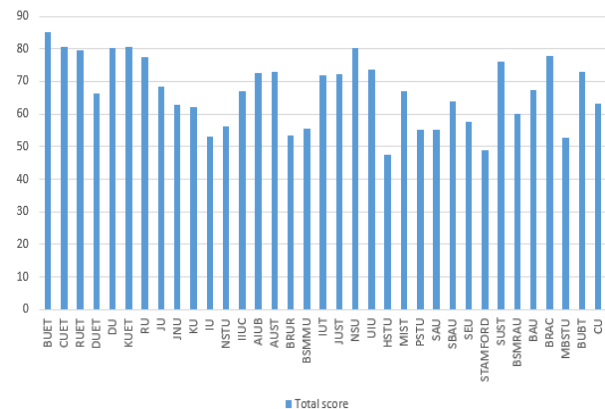| JU	68.41
BAU	67.53
IIUC	67.07
MIST	67.05
DUET	66.28
SBAU	63.83
CU	63.07
JNU	62.96
KU	62.25
BSMRAU	60.04
Southest University(SEU)	57.64
NSTU	56.21
BSMMU	55.39
SAU	55.28
PSTU	55.24
MBSTU	53.85
IU	52.93
BRUR	51.58
Stamford University	48.85
HSTU	47.59

Table. 1: Total score of each website.

Fig. 7: Comparative analysis in context of total score.