

University of Nebraska - Lincoln

DigitalCommons@University of Nebraska - Lincoln

Library Philosophy and Practice (e-journal)

Libraries at University of Nebraska-Lincoln

5-31-2020

Changes in Web Content in First 20 NIRF Ranking Institutes During 2010-19: an Analysis

Subrata Gangopadhyay

Banaras Hindu University, Varanasi, India, ganguly.subrata@gmail.com

Follow this and additional works at: <https://digitalcommons.unl.edu/libphilprac>



Part of the [Library and Information Science Commons](#)

Gangopadhyay, Subrata, "Changes in Web Content in First 20 NIRF Ranking Institutes During 2010-19: an Analysis" (2020). *Library Philosophy and Practice (e-journal)*. 4254.

<https://digitalcommons.unl.edu/libphilprac/4254>

Changes in Web Content in First 20 NIRF Ranking Institutes During 2010-19: an Analysis

Subrata Gangopadhyay

Research Scholar, DLIS and Professional Assistant, Central Library

Banaras Hindu University, Varanasi, India

ganguly.subrata@gmail.com

Abstract: *Web content is an important source for education and research. At present it is a mandatory requirement for higher learning institutes of India to present information on their institutional home page. Due to dynamic nature of web content and increase use of emerging technology, the new ways of presenting information on higher education web sites become complex. In this paper, we try to study the changes in web content during last decade in first 20 NIRF ranking institute. The Internet Archive's Wayback Machine has been used to get the web site update dates and the content of archived web pages.*

Keywords: *Web content, Archive website, Archives, Websites, Wayback Machine*

Introduction:

Nowadays websites are the primary media of any organization to disseminate information to their client. Since last two decades, it has been seen that every educational institutions of higher learning are building World Wide Web (WWW) sites to reach to their students, faculty, and scholars. Information on the web is becoming wider and popular rather in some cases essential. Content in printed document is constant but the nature of web content is dynamic, it changes regularly (Adar, Teevan, Dumais, & Elsas, 2009). Due to dynamic nature of web content, different organization like national libraries, national archives and various consortia of organizations has been involved in archiving important web content (Retrieved from: https://en.wikipedia.org/wiki/Web_archiving on 20/03/2020). Web archiving is a challenging task due to the massive size and amount of information on the web. A number of studies have been done on the change over time to web page content and structure. Kim and Lee (2005) made an empirical study using various change behaviour of the web pages, in another study Adar et al., (2009); Lim et al., (2001) have investigated the nature of dynamics of web content. Gomes, Miranda and Costa (2011) carried out a study to draw a picture on web archiving initiatives throughout the world. They noticed that most of the web archiving initiative had been grew in developed countries after 2003. Koehler (2002) analysed the nature of structural changes of websites over time. Matthews and Owings (1999) examined the content changes in science journal web sites. Hackett and Parmanto (2005) analysed to study higher education web sites for the years 1997-2002 using Internet Archive's Wayback Machine. They find that due to increase use of emerging technology, the new ways of presenting information on higher education web sites become complex. At present it is a mandatory requirement for any higher education institutes in India to use online platform to reach to their students, faculty, and scholars. In this paper, we try to study the changes in web content during last decade in first 20 NIRF ranking institute. The archive versions of web sites are visited using the Internet Archive Wayback Machine (www.archive.org),

it is a service that allow researcher to see older version of the publicly accessible websites, as well as the dates and the web content of the subsequent updates.

Objectives of the study:

The main objective of the study is to analyse the changes in web content to disseminate information in academic institutes from 2010 to 2019. In specific we will see how frequent does an Institute's web content change, which type of data or MIME-types are gaining more importance during last 10 years in academic institute websites.

Data and Methodology of Study:

The quality of any academic institute of these days depends on how they are using their web portal to disseminate information. Each and every academic institute's websites are an important tool to serve their students, faculty, and scholars. Through this study we are going to analyse how web content has been changing in academic institutes during the last 10 years. We have selected academic institutes for the study on the basis of NIRF ranking. First 20 NIRF (Table 1) ranking institute have been selected for the present study so that we can get an overall view of changes in website information of academic institute during the last 10 years.

The data have been collected using the service of Wayback Machine of Internet Archive (<http://web.archive.org/>), it is a digital archive of the World Wide Web. We have collected data in the nomenclature of

- number of times Institutional Website changes since its development,
- year wise development,
- MIME-types changes during the last 10 years, to know what type of documents are gaining importance in web information,
- Visual sitemap in the form of a Radial-tree Graph past and present, to know the changes in the level of information in the past and present.

Limitations of the Study:

The present study is confined only to the first 20 NIRF ranking academic institutes in India. The analysis is based on web content only. The study is also limited to those data, which is collected from the Internet Archive Wayback Machine, which is a free internet service. So the findings could be affected by the archiving policy adopted by the Internet Archive.

Data Analysis:

Table 1 shows 20 academic institutes with their website address, year of establishment and NIRF ranking. First 20 NIRF ranking institute have been selected for the present study so that we can get an overall view of changes in website information of academic institute during the last 10 years.

Table 1: First 20 NIRF Ranking Institutes

Institute Name	State	Website Address	Year of Establishment	NIRF Rank
Indian Institute of Science(IISC)	Karnataka	https://www.iisc.ac.in/	1909	1
Jawaharlal Nehru	Delhi	https://www.jnu.ac.in/	1969	2

University(JNU)				
Banaras Hindu University(BHU)	Uttar Pradesh	http://www.bhu.ac.in/	1916	3
University of Hyderabad(HU)	Telangana	https://www.uohyd.ac.in/	1974	4
Calcutta University(CU)	West Bengal	https://www.caluniv.ac.in/	1857	5
Jadavpur University(JU)	West Bengal	http://www.jaduniv.edu.in/	1955	6
Anna University(AU)	Tamil Nadu	https://www.annauniv.edu/	1978	7
Amrita VishwaVidyapeetham(AVV)	Tamil Nadu	https://www.amrita.edu/	1994	8
Manipal Academy of Higher Education(MAHE)	Karnataka	https://manipal.edu/mu.html	1953	9
SavitribaiPhule Pune University(PU)	Maharashtra	http://www.unipune.ac.in/	1948	10
Aligarh Muslim University(AMU)	Uttar Pradesh	https://www.amu.ac.in/	1875	11
JamiaMilliaIslamia(JMI)	Delhi	https://www.jmi.ac.in/	1920	12
University of Delhi(DU)	Delhi	http://www.du.ac.in/du/	1922	13
Bharathiar University(BU)	Tamil Nadu	https://www.b-u.ac.in/	1982	14
Institute of Chemical Technology(ICT)	Maharashtra	https://www.ictmumbai.edu.in/	1933	15
Andhra University(AU)	Andhra Pradesh	https://www.andhrauniversity.edu.in/	1926	16
HomiBhabha National Institute(HBNI)	Maharashtra	http://www.hbni.ac.in/	2005	17
JamiaHamdard(JH)	Delhi	http://jamiahamdard.edu/	1989	18
Vellore Institute of Technology(VIT)	Tamil Nadu	https://vit.ac.in/	1984	19
University of Madras(MU)	Tamil Nadu	http://www.ideunom.ac.in/	1857	20

Table 2 shows the number of times institutional web sites updated since their development i.e. website age and number of updates. Table 3 shows website update frequency- year wise. The data indicates that University of Hyderabad updated its website 617 times in 8 years whereas HomiBhabha National Institute, Maharashtra updated its site 49 times in the same time span. Delhi University has made maximum number (1765) of changes in its website since its inception with an average 92 times yearly change. It is also be noted that Anna University started to use online platform since last 21 years, to serve to their students, faculty, and scholars. Out of 1st 20 NIRF ranking institute only 3 institute (HU, BHNI and ICT) has started to use online platform which is less than 10 years old and 17 out of 20 institutes' websites are more than 10 years old. So we can say that Indian higher education institutes are mostly ICT dependent.

Table 2: No of Times Institutional Website Changes since Its Development (data collected from Internet Archive, dated: 22/01/2020)

Sl. No.	Name of the Institute	No of Times Site Saved	Date of 1 st change	Date of last change	Time Span(Years)	Average Yearly Change
1	IISC	584	27/06/2007	04/11/2019	12	49
2	JNU	1042	27/11/1999	08/01/2020	20	52
3	BHU	1023	09/02/2001	02/01/2020	19	54
4	HU	617	31/01/2012	22/12/2019	8	77
5	CU	906	16/04/2003	26/12/2019	16	57
6	JU	622	05/05/2010	12/01/2020	10	62
7	AU	1535	05/12/1998	12/12/219	21	73
8	AVV	1206	28/11/1999	01/01/2020	20	60
9	MAHE	891	23/08/2000	12/12/2019	19	47
10	PU	624	16/02/2009	17/12/2019	10	62

11	AMU	1071	20/10/2001	12/01/2020	18	60
12	JMI	616	24/07/2003	05/01/2020	17	36
13	UD	1765	23/05/2001	01/01/2020	19	93
14	BU	750	11/02/2004	07/12/2019	15	50
15	ICT	347	20/05/2010	05/11/2019	9	50
16	AU	270	02/11/2007	16/11/2019	12	23
17	HBNI	49	21/07/2011	16/08/2019	8	6
18	JH	754	30/03/2001	29/10/2019	18	42
19	VIT	1686	03/12/2001	18/01/2020	19	89
20	MU	764	06/08/2002	20/01/2020	18	42

Source: Wayback Machine - Internet Archive (<https://archive.org/web/>)

Table 3: No of Times WebPages Updated during Last 10 Years

Institute Name	Year-wise Update									
	2019	2018	2017	2016	2015	2014	2013	2012	2011	2010
IISC	61	124	264	123	3	-	-	-	-	1
JNU	93	58	70	56	73	21	43	108	70	27
BHU	114	117	118	70	42	15	19	64	38	10
HU	49	75	135	126	60	46	63	63	-	-
CU	129	119	91	76	53	42	54	45	33	30
JU	106	153	86	80	42	11	33	55	41	10
AU	97	118	34	31	35	16	52	78	79	84
AVV	88	119	57	47	27	26	54	51	29	31
MAHE	87	60	88	177	145	23	-	-	-	-
PU	83	135	130	92	41	14	28	57	39	4
AMU	95	57	51	85	63	54	96	41	25	32
JMI	83	113	86	51	37	15	31	38	52	24
UD	402	416	262	250	242	166	-	1	-	11
BU	56	146	85	54	32	21	19	51	36	18
ICT	62	84	83	57	19	11	12	9	8	2
AU	24	94	47	17	64	16	6	-	-	-
HBNI	7	5	4	7	6	4	7	6	3	-
JH	69	111	67	89	57	28	32	64	40	19
VIT	308	255	92	134	177	45	80	72	63	7
MU	67	109	71	71	33	20	7	14	35	31

Table 4 summarises different types of file format that has been used in institutional home page and how it changes during last ten years. MIME type is a label used in online platform to identify the file type, so that software can know how to handle the file. From the table it has been seen that most of the file are “text” i.e. html document which can be open internally by the site itself but in case of “application”, an application handler is needed to extract the data.

Table4: MIME-types Changes After 10 Years.



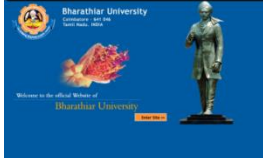

Institute Name	Text			image			application			video		
	2010	2019	Change	2010	2019	Change	2010	2019	Change	2010	2019	Change
IISC	-	-	-	-	-	-	-	-	-	-	-	-
JNU	1622	4690	3068	15	6558	6543	90	900	810	-	2	2
BHU	4941	18282	13341	6693	14240	7547	1971	8484	6513	15(audio)	13(v), 11(a)	9
HU	13132	11958	(1174)	-	124	124	10(2012)	75	65	-	-	-
CU	-	-	-	-	-	-	-	-	-	-	-	-
JU	6256	3815	(2441)	411	401	(10)	309	2506	2197	-	-	-
AU	10156	351	9805	5686	1148	(4538)	2862	364	(2498)	-	-	-
AVV	9119	1635	(7484)	16280	10582	5698	2257	1620	(637)	-	10	10

MAHE	-	-	-	-	-	-	-	-	-	-	-	-
PU	2619	779	(1840)	738	1007	269	3160	273	(2887)	8(audio)	-	(8)
AMU	1426	31122	29696	-	22635	22635	64	22651	22587	-	6	6
JMI	3838	1758	(2008)	807	1832	1025	2543	509	(2034)	-	-	-
UD	9835	9884	49	316	3018	2702	3380	2736	(644)	-	-	-
BU	-	-	-	-	-	-	-	-	-	-	-	-
ICT	57	651	594	-	696	696	5	217	212	-	-	-
AU	-	-	-	-	-	-	-	-	-	-	-	-
HBNI	64	28	(36)	25	88	63	4	-	(4)	-	-	-
JH	1712	354	(1358)	467	270	(203)	455	181	(274)	-	-	-
VIT	2912	2550	(362)	2178	640	(1538)	635	388	(247)	-	4	4
MU	119	81	(38)	89	151	62	99	39	(60)	-	-	-

Table 5 shows the first 20 NIRF ranking institutes' home page at the beginning and the present view. Table 6 displays the site structure of web pages of each institute at the beginning and at present in the form of a Radial-tree Graph. The centre circle is the root of the website and successive rings moving out from the centre represent other link pages. As we know that whenever a new institute, department or any new online services started, the parent body needs to add a new webpage to the existing institute's websites. From this site map of past and present, we can get a clear idea about how structurally changes occurred in each institute. It is quite clear from the present Radial-tree graph that structurally Banaras Hindu University (BHU) is quite large in comparison to other NIRF ranking institutes.

Table 5: Institute Home Page Past and Present

Institute Name	At the Beginning	Year 2020
IISC		
JNU		
BHU		
HU		
CU		

JU		
AU		
AVV		
MAHE		
PU		
AMU		
JMI		
UD	Don't have permission to access.	
BU		
ICT		








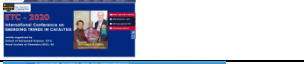







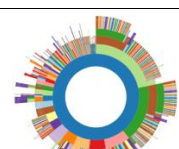
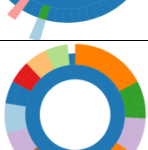
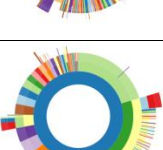
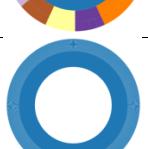
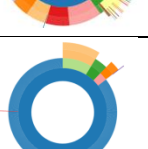
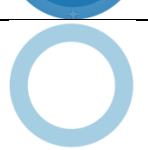
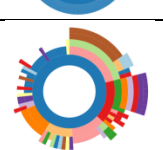







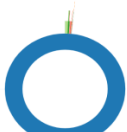


















AU		
HBNI		
JH		
VIT		
MU		

Table 6: Visual Site Map in the form of a Radial-tree Graph Past and Present

Institute Name	At the Beginning	At Present
IISC		
JNU		
BHU		
CU		
JU		
AU		

AVV		
MAHE		
PU		
AMU		
JMI		
UD		
BU		
ICT		
AU		
HBNI		
JH		
VIT		

MU		
----	---	---

Conclusion:

Though, the Internet Archive's Wayback Machine can't get access to all the websites available on internet but the archived websites and number of times website updated may facilitate researchers to study websites development over time (Murphy & Connor, 2008). The present study shows an overall view of changes in 20 NIRF ranking institutes' website during last 10 years. It has been quite clear from visual site map in the form of a Radial-tree graph (Table 6) past and present that structurally academic institutes' websites become more complex.

References:

- Adar, E., Teevan, J., Dumais, S. T., & Elsas, J. L. (2009). The Web Changes Everything : Understanding the Dynamics of Web Content. *Second ACM International Conference on Web Search and Data Mining*. Barcelona.
- Gomes, D., Miranda, J., & Costa, M. (2011). *A Survey on Web Archiving Initiatives*. 408–420.
- Hackett, S., & Parmanto, B. (2005). *A longitudinal evaluation of accessibility : higher education web sites*. 15(3), 281–294. <https://doi.org/10.1108/10662240510602690>
- Kim, S. J., & Lee, S. H. (2005). *An Empirical Study on the Change of Web Pages* g. 632–642.
- Koehler, W. (2002). *Web Page Change and Persistence — A Four-Year Longitudinal Study*. 53(2), 162–171. <https://doi.org/10.1002/asi.10018>
- Lim, L., Wang, M., Padmanabhan, S., Vitter, J. S., & Agarwal, R. (2001). *Characterizing Web Document Change*. 133–134.
- Murphy, J., & Connor, P. O. (2008). *Take Me Back : Validating the Wayback Machine*. 13(Cmc), 60–75. <https://doi.org/10.1111/j.1083-6101.2007.00386.x>
- Web archiving. https://en.wikipedia.org/wiki/Web_archiving. Accessed on 20/03/2020