

Shodhganga and Deterring Plagiarism in Research Outputs in Indian Universities

Manoj Kumar K

Jagdish Arora

Abstract

Theses and dissertations are known to be the rich and unique source of information, often the only source of research work that does not find its way into various publication channels. The INFLIBNET Centre is entrusted with the responsibility of hosting, maintaining and making the digital repository called "Shodhganga" accessible to all institutions and universities. It is envisaged that the ShodhGanga would overcome serious problem of duplication of research and poor quality resulting from the "poor visibility" and the "unseen" factor in research output. In order to increase quality of research, plagiarism is to be checked before accepting the content in theses and dissertations.

Under Shodhganga project, Two plagiarism detection software tools, namely "iThenticate" and "Turnitin", are being made accessible to 110 universities with funds made available by the University Grants Commission. This paper discusses the model introduced in India for creation of a central repository of theses and dissertations, its current status, guidelines for submission of these, deterring plagiarism under the Shodhganga initiative, criteria adopted by the INFLIBNET Centre for selection of plagiarism detection platform, analysis of usage, recommendation of formulating policies and guidelines to check plagiarism in universities, future plans etc.

Keywords: ETD, Shodhganga, Plagiarism, Turnitin, iThenticate, Theses and Dissertation, Indian Universities, Higher Education, Research, INFLIBNET Centre

1. Introduction

The higher education system in India is one of the largest education system in the world consisting of more than 693 universities/institutions and 37,209 colleges as on Dec, 2014. 70% of the universities along with its affiliated colleges offer doctoral programmers in various subjects. Strict guidelines and regulations are necessary to ensure good quality of education and research in the country. The University Grants Commission (UGC), an apex body of Government of India, control and monitor the higher education in India. A closure examination of procedures involved in approval of a doctoral disserta-

tion reveals that the process of scrutiny, validation and its approval is confined to few experts, identified by the university on recommendation of supervisors of theses. It is not open to scientific community at large, and therefore, quality is compromised. Physically, theses collections in most of the libraries are kept in closed access, making it difficult for other students to access them. As such, it remains an untapped and under-utilized asset, leading to unnecessary duplication and repetition that, in fact, is the anti-theses of research and wastage of huge resources, both human and financial.

Mandatory guidelines and regulations are issued by UGC from time-to-time for regulation of higher education and research programmers. The UGC



10th International CALIBER-2015
HP University and IAS, Shimla, Himachal Pradesh, India
March 12-14, 2015
© INFLIBNET Centre, Gandhinagar, Gujarat, India

Notification (Minimum Standards & Procedure for Award of M.Phil. / Ph.D Degree, Regulation, 2009) mandates submission of electronic version of theses and dissertations by the researchers in universities with an aim to facilitate open access to Indian theses and dissertations to the academic community world-wide. Online availability of electronic theses through centrally-maintained digital repositories, not only ensures easy access and archiving of Indian doctoral theses but would also help in raising the standard and quality of research ultimately by curbing the chances of plagiarism. Task of implementation of above mentioned task in the regulation and responsibility for hosting, maintaining and making the digital repository of Indian Electronic

Theses and Dissertation (called "Shodhganga"), accessible to all institutions and universities was assigned to the INFLIBNET Centre, an IUC of UGC, Gandhinagar, India. It is envisaged that the central repository of theses and dissertations (named ShodhGanga) would overcome serious problem of duplication of research and poor quality resulting from the "poor visibility" and the "unseen" factor in research output. Plagiarism is recognized as one of the greatest threats to the research in higher education system in India that may lead to deterioration in quality of research coming out from the universities.

The screenshot shows the Shodhganga website interface. At the top, it features the Shodhganga logo and the tagline 'a reservoir of Indian theses'. The main navigation bar includes 'New User', 'Login as: Shodhganga User', 'University Coordinator', and 'Administrator'. The left sidebar contains a search box, 'Advanced Search', and a 'Browse' menu with categories like 'Universities & Departments', 'Submit Date', 'Researcher/Guide', 'Title', 'Keyword', and 'University'. Below this is a 'Top 5 Contributing Universities' list and a 'Theses to be Uploaded' section.

The central content area is titled 'Welcome to Shodhganga@INFLIBNET Centre'. It includes a 'Shodhganga received e-INDIA Jury Choice award.' announcement and a 'Have Synopses? Click Here' link. A 'Research in Progress' banner is also visible. Statistics are provided: 'Universities Signed MoU for Shodhganga : 197', 'No. of Thesis in Shodhganga : 30,000+', 'University Contributed : 196', and 'No. of Synopses in Shodhgangotri : 1884'. A 'Poll your Preference for Anti-Plagiarism' section is also present.

The right sidebar features an 'Update @ Shodhganga' section with news about a national seminar and an e-INDIA award. It also includes 'Research in Progress (Shodhgangotri)', 'UGC Notification', and 'More@Shodhganga' links.

At the bottom, a table lists '40 Top Universities Contributed' to Shodhganga, including Jawaharlal Nehru University [4200+], Anna University [2220+], Mahatma Gandhi University [1980+], Bundelkhand University [1769], and Cochin University of Science & Technology [1410+].

Figure 1 : Home Page of Shodhganga

2. Shodhganga and Anti Plagiarism Check

Shodhganga was set-up in January 2010 for submission of electronic version of theses and dissertations by students / research scholars in all universities in India. As on 31st Jan 2015, the repository hosts more than 30100 full-text theses from 196 universities and 186 universities have signed MoU with the INFLIBNET Centre on Shodhganga. Home page of Shodhganga is shown in figure 1 and year wise submissions of theses are shown in figure 2.

To promote the research culture among the universities, three incentives are offered to universities that are signatory to MoU on Shodhganga with INFLIBNET Centre are: i) providing access to plagiarism detection web based platform; (ii) funds for digitization of theses based on number of theses available in a university; and iii) funds for setting-up of ETD Laboratory in eligible universities. [1]

While digitization of backlist of theses in universities will increase critical mass of data available in Shodhganga, setting up of ETD laboratories in universities would lead to separate division for supporting creation and uploading of theses as per Shodhganga guidelines and accessibility of plagiarism detection software will deter plagiarism and ultimately increase in quality of research.

For implementation of plagiarism detection under this scheme, two plagiarism detection web based platform, namely "iThenticate" and "Turnitin" are being made accessible to 110 universities that are signatory to MoU on Shodhganga with INFLIBNET Centre with funds made available by the University Grants Commission. Universities have been instructed to subject all their theses to plagiarism detection using one of the platforms made available to them before awarding and submitting their theses into Shodhganga.

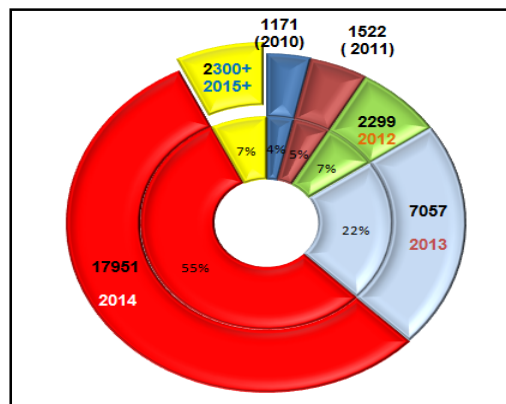


Figure 2: Shodhganga Year Wise Submission

3. Selection of Vendor for Anti-plagiarism Software

In order to find out best product and vendor, the INFLIBNET Centre constituted a special Committee for procuring Anti-Plagiarism Software in June 2013 which was earlier recommended by the national committee and UGC to provide subscriptions to the universities. The Committee finalized the tender specifications along with formats of bids, technical specifications, financial bid format, general terms and conditions etc. and the same was floated on July, 2013 on INFLIBNET website as well as in the national newspapers (All India Editions). Queries regarding any clause in tender were asked to submit by the vendor in mid July and per-bid meeting was held in the last week of July, 2013. The three vendors attended the pre-bid meeting and clarified few queries regarding the tender clauses. Based on the clarification, two tender forms (bids) were received from two vendors for two products i.e. EPhorus and iThenticate/Turnitin and the technical bids were opened in August, 2013. Subsequently, as per the request by the INFLIBNET Centre, publishers provided 3 months trial period from September-November, 2013 for evaluation of the features.

The software was required to support the following requirements:

- a) Permit site license for unlimited users in a university
- b) Off line content checking
- c) On line content checking
- d) Plagiarism check with maximum number of authentic on-line sources
- e) Integration with popular search engine e.g. Google, Google Scholar, OAlster,
- f) SCIRUS, Bing and Altavista
- g) Minimum test time
- h) Native PDF support without 3rd party software
- i) Multiple formats support: HTML, MS Word, Word Perfect, Post Script, Portable Document Format (.pdf), Rich Text, Plain Text
- j) State-of-art advanced report viewing interface
- k) Source tracing mechanism
- l) On-the-fly original to source comparison with dynamic statistics
- m) Plagiarism chart generation
- n) References automatic detection
- o) Multiple documents processing
- p) Seamless integration with the existing CMS/LMS
- q) Should cover a large number of databases for comparison

Both products were checked and technically verified for the features quoted on the products on the trial period by submitting various kinds of documents including thesis, new articles, articles from open access journal, articles from paid publishers,

aggregators etc. Based on compliance data sheet consisting of general, technical features supported and presentations made, evaluation of the product based on technical grading was done for the final selection.

4. Technical Grading of the Product Depending upon the Merits

Grades based on percentage of marks were allocated to the vendors as per tender document for each feature i.e. 25% for General section, 50% for technical and 25% for the other features supported and presentation made by the vendor in presence of the scientists in the Centre.

General section with a weightage of 25% for the criteria of selection include number of countries where the software is used, number of institutional customers using the software globally, number of universities in India subscribing to the software and number of top 200 universities in the world using the software covered under university ranking scheme by AIWU ranking, Times Higher Education World University ranking and QS World University ranking, number of colleges in India subscribing to the software, publisher's content (articles) available in their database from end-to-end under agreement signed between publisher and the vendor from UGC Infonet Journals, publisher's content (articles) available in their database from end-to-end under agreement signed between publisher and the vendor from NLIST under INFLIBNET/MHRD for E-Journals/ E-Books, number of paid databases of popular publishers covered by the software, number of aggregators sources and annual turnover of the company and also their willingness to provide trial access for universities.

Technical section (50%) included Internet checker database size in billion, student papers database in millions, Plagiarism checker database size with web pages in billion, scholarly content items in million, updation frequency, no. of newspapers covered, magazines and books in million covered, number of documents scanned and added every year, number of documents added from publishers every year, number of articles added in last 3 years etc.

The third section for features supported and presentation (25%) include payment and bank guarantee, trial access for 2-3 weeks to all mentioned universities after placing an order, Integration capabilities (API Availability for Learning Management Systems like Moodle etc; content repository databases like DSpace etc; allow custom development of API toolkit), integration capabilities with proxies/ single sign-on features Shibboleth/EZProxy, originality check, interface support for ZIP upload, periodic table of all uploaded documents per Guide/ Teacher, ease to upload multiple documents, facilitate personal database per researcher/guide, check against researcher's older work, choice of email-notifications, possibility to check with database of other subscriber. [2] The product also should have support for generating report within 24 hours, tracing original sources > 99%, Similarities with paid databases, partnership with CrossRef, to support multiple file types, manuscripts comparison, support for national wide administration panel for INFLIBNET for viewing usage statistics and other reports, number of sources at report, originality check should permit re-phrase, similarity check, results in three steps or lesser, reporting feature similarity score, reporting feature similarity report, reporting feature content tracking, reporting feature summary report, customized training in all universities, ability to provide

grading such as GradeMarks facility, support for automated peer review, support page sizing/document limit, language support, provide contractual guarantees that protect the privacy of the submitted documents. [3]

Based on the technical bid, evaluation done by the percentage of marks scored in the three areas in the technical bids, financial bids opened for those who score minimum qualification marks which was decided by Committee constituted for this purpose by the INFLIBNET Centre (e.g. 70%). On evaluation of the features of both products, anti-plagiarism software Turnitin/iThenticate found to be more suitable for the universities. It is seen that Turnitin/iThenticate is the most used products in Indian education institutions as well as world-ranked universities. Ephorus is lacking its presence in Indian markets and universities as well as the important criteria for database access ie. Crossref is missing in this product. There are many deviations shown by Ephorus against the major features required as per the tender document. Marks scored by both parties are varying significantly and on 30th December 2013 the committee finally decided to procure iThenticate/Turnitin software after negotiation meeting with vendor.

5. Content Database of Turnitin/iThenticate

It is highly pertinent that any anti-plagiarism software need to crawl through the maximum digital content to check for plagiarism within the short time. Digital fingerprints of all manuscripts are to be stored in a database or accessible to the software for complete similarity checking. Both the products use almost same databases for the similarity checking except Turnitin has an additional database of student paper archives. Many a times, students paper, term papers etc. are copied from student papers themselves. Therefore regular addition of content in student papers to the archival repository is

essential. Due to the copyright issues, all published content cannot be archived to a repository and the software provider need to enter into agreement with library databases, textbook publishers, digital reference collections, subscription based publications, homework helper, site and books. [2] As on date, over 45 billion web pages, 337 million student papers and 130 articles from various publishers are available in the database and built-in web crawler of Turnitin/iThenticate regularly crawls the Internet as done by other search engines like Google or Bing and index the content from various sources by indexing into searchable forms. The software also have language support for 30 foreign languages but regional language support is poor. Hindi manuscripts are successfully tested in the products, but due to lack of online regional content, the result is cannot be taken as final and authenticated.

6. Originality and Similarity Check by Turnitin/iThenticate

In order to check plagiarised content, originality check is to be done by using the anti plagiarism software. Similarity check report can be generated by both software with exact number of matches of words. About 30000 documents are submitted by researchers and faculty members in iTheniticate (10600) and Turnitin (19246) till December 2014 for the similarity check. Details are given in Table.

Table 1: Originality and Similarity Check by Turnitin/iThenticate

Sr. No.	Month	Documents Submissions in iThenticate	Documents Submissions in Turnitin
1	December 2014	2042	3234
2	November 2014	680	2773
3	October 2014	1033	2971

4	September 2014	1483	1463
5	August 2014	1236	1437
6	July 2014	440	1613
7	June 2014	980	796
8	May 2014	1168	728
9	April 2014	664	1773
10	March 2014	679	2458
	Total	10405	19246

Both administrative interface gives comprehensive results in different style. The administrator of Turnitin can generate university wise and nation wise report from Turnitin online, while iThenticate reports are generated by system on demand from the user by the technical team of the software. The result of usage and similarity found in Turnitin from all 110 universities till Dec 2014 is given in the Table below.

Status of Turnitin usage by Universities (Mar 2014-Dec 2014)

1	Total No. of instructors added by Universities	947
2	Total No. of students added	657
3	Total No. of submissions	19400
4	Originality Reports generated	18728
5	No. of articles with 75-100% similarity	1749
6	No. of articles with 50-74% similarity	1942
7	No. of articles with 25-49% similarity	4565
8	No. of articles with 0-24% similarity	9620
9	Articles with no matches found	852

Percentage of similarity reported by Turnitin in Universities from 19400 submissions is given in chart. This indicates that 75-100% similarity is reported during checking in 25% of the article submissions. Students get a chance to improve and remove similar content from their articles or theses. Out of the originality report generate about 9% of articles are showing similarity between 75-100%. About 10% of articles are having similarity between 50-74% and 23.6% of articles are having similarity between 25-49%. 50% of articles are having similarity between 0-24%. Only 4.4% (ie.852) articles out of 19400 are not having any match with any content which is acceptable for its originality.

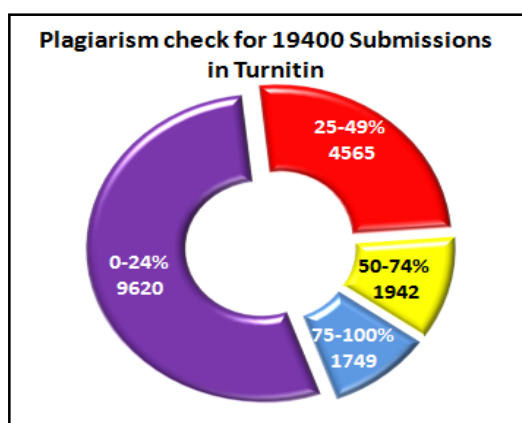


Figure 3: Similarity Reports in Various Category (% and Nos)

Top 20 universities in using iThenticate are listed below to show the extensive use of it within nine months of implementations.

Sl No	University Name	No of Submissions
1	Guru Nanak Dev University	1471
2	Jamia Hamdard University, Hamdard Nagar	947
3	Cochin University of Science and Technology	920

4	Aligarh Muslim University, Aligarh	865
5	University of Kashmir	708
6	University of Delhi, New Delhi	706
7	Tezpur University, Tezpur	578
8	Central University of Rajasthan, Ajmer	521
9	Jawaharlal Nehru University	465
10	Tilak Maharashtra Vidyapeeth	457
11	Pandit Ravishanker Shukla University	445
12	Shri Mata Vaishno Devi University (SMVDU), Katra	390
13	Indraprastha Institute of Information Technology	363
14	Gauhati University	308
15	University of Mysore, Mysore	235
16	Dayalbag Educational Institute, Agra	219
17	INFLIBNET Centre, Shodhganga	200
18	Doctor Harisingh Gour Vishwavidyalaya, Sagar	193
19	University of Hyderabad	191
20	Calicut University	183

Top 20 universities in using Turnitin are listed below to show the extensive use of it within one year of implementations.

Sl No	University Name	No of Submissions
1	Tata Institute of Social Sciences, Mumbai	10521
2	Banasthali Vidyapeeth, Jaipur	5296
3	Jawaharlal Nehru University	5047
4	Pondicherry University, Puducherry	3623
5	Jawaharlal Nehru Technological University, Kakinada	2616
6	Dayalbag Educational Institute, Agra	2447
7	Delhi Technological University, New Delhi	2193
8	University of Pune, Pune	2122
9	University of Delhi, New Delhi	893
10	Jawaharlal Nehru Technological University, Hyderabad	805
11	University of Hyderabad	651

12	Avinashilingm Institute for home Science and Higher education for women, Coimbatore	564
13	Central University of Gujarat	506
14	Tezpur University, Tezpur	387
15	Gauhati University	321
16	Jawaharlal Nehru Technological University, Anantapur	291
17	Maharshi Dayanand University	250
18	Doctor Harisingh Gour Vishwavidyalaya, Sagar	246
19	Institute of Chemical Technnology	218
20	Guru Gobind Singh Indraprastha University	208

7. Plagiarism Check by INFLIBNET Centre: Case Study

INFLIBNET Centre do not have any mandate to check plagiarism for the awarded theses or theses under submission. A case study is shown how the software/tool is used to find out plagiarised content in the articles received. The Centre is using the software for checking plagiarism for invited papers for various conferences like CALIBER and PLANNER etc. Centre provide the plagiarism report to reviewers or the reviewers generally submit the articles to the software before starting the review. The pragmatic approach for selection of a paper is first done with similarity check report by using the software. The case study is not done on theses or articles in universities due the lack of mandate, but from the article received for an International Conference to justify how quality of a paper is initially checked for its similarity report and decisions are made. In universities same policy may be adopted with better policy guidelines.

For presenting in the International Conference CALIBER 2015, 177 papers were received before 3rd February 2015 which are taken for the study. Out of 177 papers, only 46 papers are reported for minimum (0-10%) copied contents. Three papers are in the category of 75-100% which were straight away rejected by the reviewers. 26 papers reported about 51-75% plagiarism which also reported to be rejected after a glance of the content. 42 Papers which reported similarity report of 26-50% in its content which are rechecked for originality and authors are instructed to modify the the paper after giving proper citations. 60 Papers which reported similarity report of 11-25% are accepted in principle for review as per the guidelines after checking for the originality and its relevance to the theme. Details are given in Figure 3:

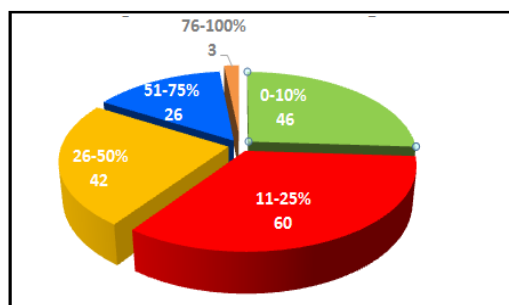


Figure 3: Plagiarised Content in %age

% of copied content	No. of Papers
0-10%	46
11-25%	60
26-50%	42
51-75%	26
75-100%	3
Total	177

8. Findings and Recommendations

There was no tool for checking plagiarism in many Universities till INFLIBNET Centre provided the antiplagiarism software in 2014. The Centre monitors usage of anti-plagiarism percentage through the interface provided by the developer of the software. Through this analysis it is ascertained that research articles or these chapters are being subjected to copied content in most of the universities.

- ▶▶ The software is provided to only UGC affiliated universities (Universities under section 12(B) of UGC Act). ie. Universities eligible for getting grants from UGC.
- ▶▶ The software usage has boosted universities to detect plagiarism in the articles.
- ▶▶ The analysis in usage of Turnitin software shows that there was a considerable amounts of plagiarised content in articles/theses chapters, which is more than permissible limit set up by many universities (10-15%).
- ▶▶ More than 50% of the articles submitted in Turnitin are having 25% and more copied content.
- ▶▶ Tendency of copied content is visible to the extent of 75-100% in Universities from various sources.
- ▶▶ Plagiarised content is reported from all types of universities irrespective of its status as central, state, deemed etc. which used the detection tool. Official records are not available for private, self financing universities.
- ▶▶ A number of universities have developed policies for anti-plagiarism.

Based on the above findings, following recommendations are made:

- ▶▶ Research work in all universities should be brought under the plagiarism check policy.
- ▶▶ Universities should develop proper anti-plagiarism policies for research.
- ▶▶ Proper citation rules and content excluded from plagiarism checking should be formulated by each department with the consent of University.
- ▶▶ Researcher should check the articles/theses for plagiarised content by a software mentioned by UGC/INFLIBNET Centre and submit a report to the Research Supervisor.
- ▶▶ Research Supervisor should certify the plagiarism check report and justify any content copied from other source based on fair use, which is generally required in literature review.
- ▶▶ INFLIBNET Center should make draft antiplagiarism policy for the Universities/ Institutes under UGC.
- ▶▶ The software should be made available to other private/self financed universities under associate membership.
- ▶▶ Proper Hands-on and awareness programme should be conducted in universities for the research scholars.
- ▶▶ If plagiarism is reported in theses which are already awarded and later uploaded into Shodhganga, the decision of withdrawal of the theses should be left to the Universities.

In this way, the academic community can cultivate the awareness about anti plagiarism among the researchers for their articles and students for their

assignments. This model can enhance the quality of research in the nation especially in universities, so that an anti-plagiaristic environment in the higher education system would be emerged in India.

9. Conclusion

Open repositories are increasing day by day to support research and academic content. Research outputs are published to get wider coverage and lead to its collaborative use for new research avenues. Copying the content of others, knowingly or unknowingly, create suspicion about the quality of research. In order to bring out quality research and articles with originality, awareness and ethics of doing research should be made important. Tools and software will help to detect similarity reports and matching of words in articles. But onus of creation of original content is ultimately the responsibility of the researcher. Copying of intellectual property of original author is not fare and advisable in research. Shodhganga maintains the quality of its theses content and it is also important that content created by the research scholar should be verified before awarding the PhD. Antiplagiarism tools made available to universities would play an important role in curbing the plagiarism in universities, thus, enhance the quality of research in universities.

References

1. M. Kumar and J. Arora, User's Guide and tutorials : Shodhganga Repository of Indian Electronic Thesis and Dissertation, Ahmedabad, 2012.
2. Turnitin, "Originality Check, Online Grading, Peer Review," 3 February 2009. [Online]. Available: http://turnitin.com/en_us/features/originalitycheck/content. [Accessed 6 February 2015].
3. M. Kumar K, "INFLIBNET Centre Home Page - Tender," INFLIBNET Centre, 13 July 2014. [Online]. Available: http://www.inflibnet.ac.in/tenders/Antiplagiarism-2013_10_07_2013.pdf. [Accessed 15 January 2015].
4. "Shodhganga Home Page," UGC, 11 June 2009. [Online]. Available: http://shodhganga.inflibnet.ac.in/moredetails/UGC_Guidelines_for_Shodhganga.pdf. [Accessed 15 January 2015].
5. M. Kumar, N. Chand and S. Gandhi, "Ontological Mapping for Semantic Search in Shodhganga : A National Repository of electronic Thesis and Dissertations (ETDs)," in Conference Proceedings PLANNER 2012, Gangtok, 2012.

About Authors

Mr. Manoj Kumar K, Scientist-D (CS), INFLIBNET Centre, Gandhinagar, Gujarat.
Email: manoj@inflibnet.ac.in

Dr. Jagdish Arora, Director, INFLIBNET Centre, Gandhinagar, Gujarat.
Email: director@inflibnet.ac.in