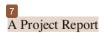
22pp.pdf

Submission date: 11-May-2020 04:27PM (UTC+0530)

Submission ID: 1321595845 File name: 22pp.pdf (1.22M)

Word count: 3619

Character count: 21442



On

PLAGIARISM CHECKER

Submitted for the partial fulfillmentof the requirement

for the award of the Degree of

B.Tech

In

Computer Science and Engineering

by

Harshit Agrawal (170102314)

Prathmesh Jain (180102906)

Utkarsh Saxena (170102174)

Satyam Bhatia (170102167)

Under the Guidance of

Mr.Kushal Gupta

Asst. Professor, DIT Univeristy, Dehradun



April,2020



DECLARATION

This is to certify that the Project entitled "Plagiarism Checker" in partial fulfillment of the requirement for the award of the Degree of Bachelor of Technology in Computer Science and Engineering, submitted to DIT University, Dehradun, Uttarakhand, India, is an authentic record of bonafide work carried out by me, under the guidance of Mr. Kushal Gupta (Asst. Prof. ,DITU).

15

The matter embodied in this Project/ Thesis/ Dissertation has not been submitted for the award of any other degree or diploma to any University/ Institution.

Harshit Agrawal Utkarsh Saxena Prathmesh Jain Satyam Bhatia

Students Name & Signature:

Mr. Kushal Gupta (Asst. Prof. DITU,Dehradun)

Supervisor Name , Designation & Signature:

7 Dr. Vishal Bharti Head of Department Signature

Date:24 April,2020

Place: Dehradun

6 ACKNOWLEDGMENT

The entry level position opportunity I had with DIT University was an unreasonable possibility for learning and general proficient improvement. Thus, I view myself as an exceptionally fortunate individual as I was given an event to be a piece of it. I am likewise obliged for having an opportunity to meet such a Faculty, superb individuals who drove me through this period.

I express my warm on account of Mr. Kushal Gupta, Module Lead for taking part in valuable choice and giving fundamental advices and supervision and organized all assignments to make the work simpler. I pick this minute to perceive his commitment thankfully.

I see this open door as a major pivotal in my profession advancement. I will endeavor to utilize the earned arrangement of aptitudes and information in the foremost conceivable way, and I will keep on working on their upgradation, with a specific end goal to accomplish foreseen vocation targets. Want to proceed with help with every one of you later on.

1 Abstract

To detect plagiarism of any form, it is essential to have broad knowledge of its possible forms and classes, and existence of various tools and systems for its detection. Based on impact or severity of damages, plagiarism may occur in an article or in any production in a number of ways. These surveys presents taxonomy of various plagiarism forms and include discussion on each of these forms. Over the years, a good number tools and techniques have been introduced to detect plagiarism.

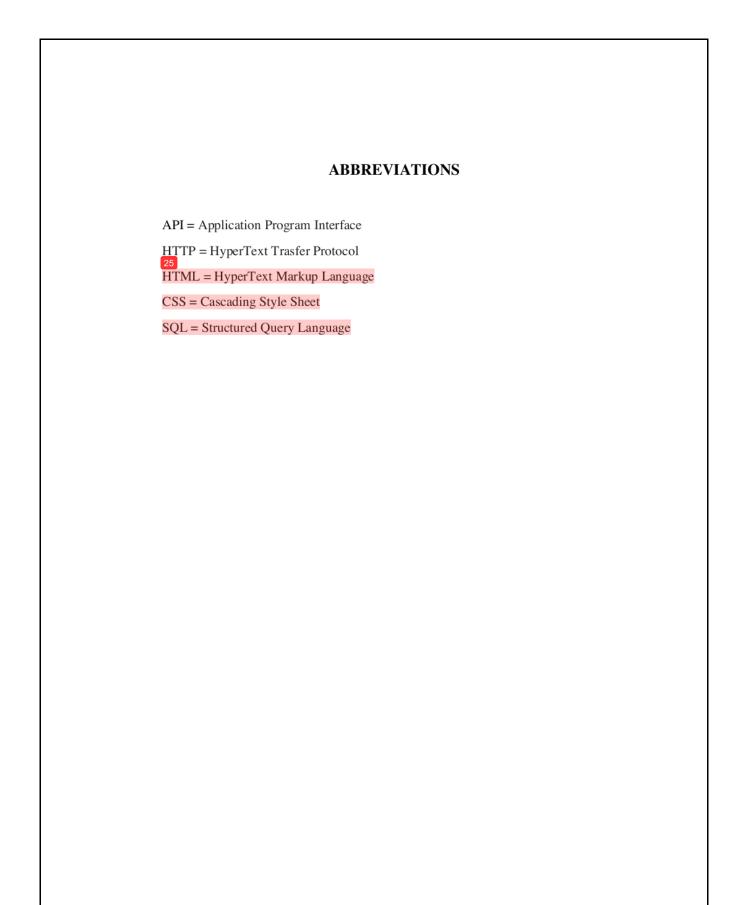
TABLE OF CONTENTS

Ti	tle	Page No.
DI	ECLARATION	::
וט	ECLARATION	11
A(CKNOWLEDGEMENT	iii
Al	BSTRACT	iv
TA	ABLE OF CONTENTS	v
LI	ST OF FIGURES	vii
Al	BBREVIATIONS	viii
CITA	PERD 1 INTRODUCTION	
CHA	PTER 1 INTRODUCTION	
1.1	Introduction	
СНА	PTER 2 Progress	
2.1	Progress	
	1 105.000	
СНА	PTER 3 Gallery	
3.1	Images	
СНА	PTER 4 Technologies Used	
6 .1	Node JS.	
.2	Express	
.3	MongoDB	
.4	Bootstrap	
.5	HTML	
.6	Javascript	
.7	CSS	

CHA	PTER 5	Code
5.1		el
5.2	Server Fil	e
34		
CHA	PTER 6	Future Scope
<i>C</i> 1	Entre C	cope
CHAI	PTER 7	Plagiarism Detection Methods
7.1	Plagiaris	m detection Methods
7.2	Similarit	y Percentage Calculation
REF	ERENCES	

LIST OF FIGURES

Figure No.	Title	Page No.
19		
3.1 Home Page	·	x
3.2 Sign Up Pa	ge	X
3.3 Login Page		xi
3.4 Dashboard	Page	xi
		xi
3.6 Courses Pa	ge	xi
		xii
3.8 Submission	n Page	xiii
3.9 Profile Pag	e	xiv



1 INTRODUCTION

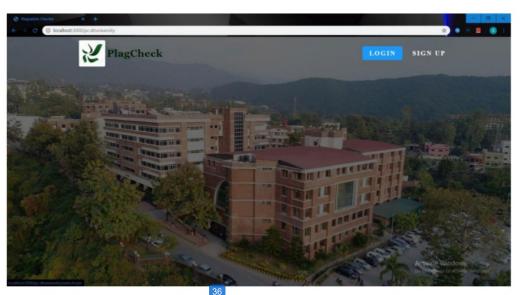
Due to the modern world of digital era, the volume of digital resource has been increasing in the World Wide Web tremendously. Today creation of such digital resources and their storage and dissemination is very eary. With the rapid growth of these digital resource, the cannoes of copyright violation and plagiarism has also been increasing simultaneously. To address this issue in depth, researchers started working on plagiarism detection in different languages. It was pioneered by a copy detection method in digital documents. However, the misuse of software detection was initiated even much earlier, in 1970 by detecting plagiarism (copying) among programs. Since then, a good number of methods and tools have been developed on plagiarism detection which are available online. But it is very much chaotic when one wants to choose the best plagiarism detection method or plagiarism detection tool. It may be due to lack of controlled evaluation environment in plagiarism detection research. Plagiarism is the presentation of another's words, work or idea as one's own. It has two components, viz., Taking the words, work or ideas from some source and presenting it without acknowledgment of the source from where word, work or idea are taken. Plagiarism can appear in different forms. However, there are mainly two types of plagiarisms typically found to occur, such as "textual plagiarism" and "source code plagiarism". Plagiarism may occur in various languages such as within same natural language or it may appear between two or more different languages. Many researchers or software companies are still trying to provide an efficient method or tool for plagiarism detection. There are mainly two types of plagiarism detection approaches available based on whether external resources or references are used or not during plagiarism detection, such as "intrinsic plagiarism detection" where no external references are used and "extrinsic plagiarism detection" where external references are used.

.

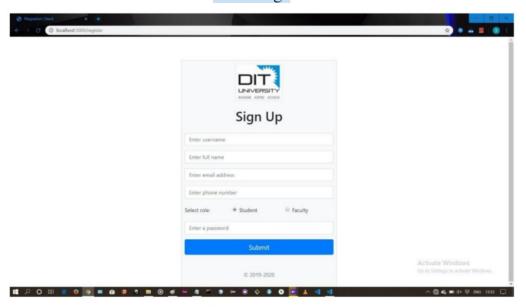
Progress

Till now we have worked on the front end part of our project. We have developed a landing page i.e.,when we go to site the landing page will be shown which will contain the 3 options- about, sign up, sign in. Our landing page is shown in the below picture. If we already have an account on this site we can directly log in to our account and find the option to upload our assignment which can be checked for plagiarism else we need to create our account using the sign up option available at right corner of the page. While signing up for new account it will ask your role as in student signup or the faculty sign up, we need to choose any one according to our role. Once we are successfully logged in we are taken to the page where we can upload our file for plagiarism detection.

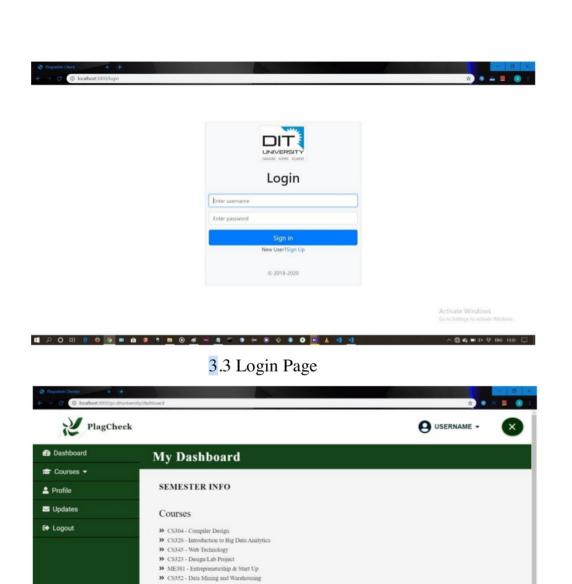
Gallery



3.1 Home Page



3.2 Sign Up Page



3.4 Dashboard Page

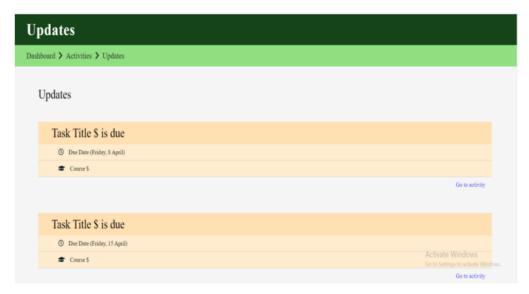
Faculties

> Zeya Umayya

> Mayank Sah

> Anurag Srivasta

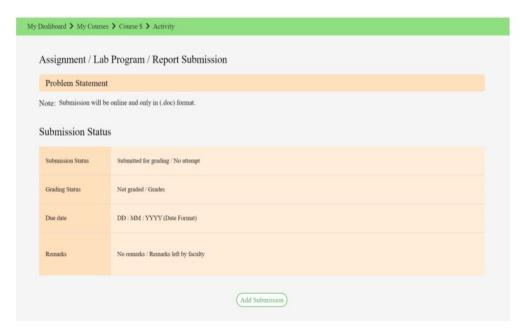
> Kushal Gupta



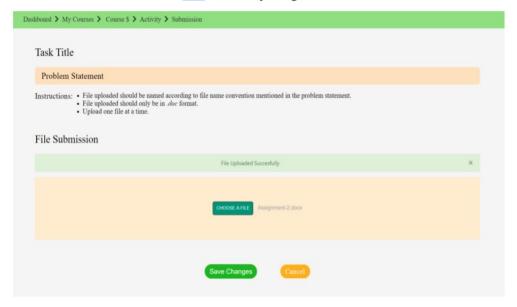
3.5 Updates page



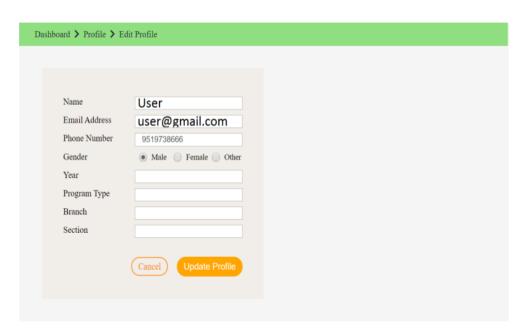
3.6 Courses Page



3.7 Activity Page



3.8 Submission Page



3.9 Profile Page

Technologies Used

18 Node JS (Language):

As an asynchronous event-driven JavaScript runtime work, Node.js is designed to build scalable network application. It is similar in design, and influenced by systems like Ruby's Event Machine and Python's Twisted. Node.js takes the event model a little further. It presents an event loop as a runtime construct instead as a library. In other systems, there is always a blocking call to start the event-loop. Typically behavior is defined through callbacks at the beginning of a script & at the end a server is started through a blocking call like EventMachine::run(). Here, there is no such start-the-event-loop call. Node.js simply enters the event loop after executing the input script. It exits the event loop when there are no more callbacks to perform. This behavior is like browser JavaScript, the event loop is hidden from the user.

HTTP is a first-class citizen in Node, js, designed with streaming and low latency in mind. This makes Node, js well suited for the foundations of a web library or framework. Node, js, which is being designed without threads doesn't mean we can't take advantage of multiple cores in its environment. Child processes can be spawned by using our child_process.fork() API, and is designed to be easy to communicate with. Built upon that same interface is the cluster module, which allows you to share sockets between processes to enable load balancing over your cores.

Express (Server Side):

11

Express is a web application framework for Node.js that allows us to spin up robust API and web server in a much easier and the aner ways. Express is a lightweight package that does not obscure any of the core Node.js features.

The author, TJ Holowaychuk, described express as a Sinatra-inspired server[citation needed], meaning that it is relatively minimal with many features available as plugins. Express is the back-end component of the MEAN stack, together with the MongoDB database software and AngularJS front-end framework.

MongoDB (Database):

MongoDB is a cross-platform document-oriented database program. MongoDB is Classified as a NoSQL database program, which uses JSON-like documents with schema. MongoDB is developed by MongoDB Inc. and licensed under the SSPL.

MongoDB supports field, range query, and regular-expression searches. Queries can return specific fields of documents and also include multiple user-defined JavaScript functions. These Queries can also be configured to return a random sample of results of a given size. Fields in a MongoDB document can be indexed with primary and secondary indices.

It provides high availability with replica set. A replica set consists of two or more copies of some data. Each replica-set member acts in the role of primary or secondary replica at any time. All writes and reads are done on the primary replica by default. Secondary replicas maintains a copy of the data of primary using built-in replication. When the primary replica fails, the replica set automatically conducts an election process to determine which secondary should become the primary. Secondary can optionally serve read operations, but that data is only eventually consistent by default.

10 Bootstrap:

Bootstrap is an open source toolkit for developing with the help of HTML, CSS, and JS. It quickly prototype ideas or builds entire app with our Sass variables and mixins, responsive grid system, extensive prebuilt components, and powerful plug-in built on jQuery.

Npm libraries/dependencies used:

- 1) body-parser
- 2) express
- 3) express-fileupload
- 4) express-session
- 5) mongoose
- 6) passport
- 7) passport-local passport-local-mongoose

HTML:

2

HTML is the standard markup language for documents designed to be displayed in a web browser. It can be assisted by technologies such as CSS and scripting languages such as JavaScript.Web browsers receive HTML documents from a web servers or from local storage and renders the document into multimedia web pages. HTML describes the structure of a web page and originally included cues for the appearance of the document.

HTML elements are the building blocks of HTML pages. With HTML constructs, images and other objects such as interactive forms may be embedded into the rendered page. HTML provides a means to create structured documents by denoting structural semantics for text such as headings, paragraphs, lists, links, quotes and other items. HTML elements are delineated by tags, written using angle brackets. Tags such as and <imput/> directly introduce content into the page. Other tags such as surround and provide information about document text and may include other tags as sub-elements. Browsers do not display the HTML tags, but use them to interpret the content of the page.

HTML can embed programs written in a scripting language such as JavaScript, which affects the behavior and content of web pages. Inclusion of CSS defines the look and layout of content. The World Wide Web Consortium (W3C), former maintainer of the HTML and current maintainer of the CSS standards, has encouraged the use of CSS over explicit presentational HTML since 1997.

JavaScript:

JavaScript is a lightweight, interpreted, object-oriented language with first-class functions. It is best known as the scripting language for Web pages, but it is used in many non-browser environments also. It is a prototype-based, multi-paradigm scripting language that is dynamic, and supports object-oriented, imperative, and functional programming style.

JavaScript runs on the client side of the web, which can be used to design / program how the web pages behave on the occurrence of an event. JavaScript is an easy to learn and also powerful scripting language, widely used for controlling web page behavior. Contrary to popular misconception, JavaScript is *not* "Interpreted Java". In a nutshell, JavaScript is a dynamic scripting language supporting prototype based object construction. The basic syntax is intentionally similar to both Java and C++ to reduce the number of new concepts required to learn the language. Language constructs, such as if statements, for and while loops, and switch and try ... catch blocks function the same as in these languages (or nearly so).

It functions as both a procedural and an object oriented language. Objects are created programmatically in JavaScript, by attaching methods and properties to otherwise empty objects at run time, as opposed to the syntactic class definitions common in compiled languages like C++ and Java. Once an object is constructed it can be used as a blueprint for creating similar objects.

CSS:

8

CSS, is a simple design language intended to simplify the process of making web pages presentable. CSS handles the look and feel part of a web page. Using CSS, you can control the color of the text, the style of fonts, the spacing between paragraphs, how columns are sized and laid out, what background images and colors are used, layout, designs, variations in display for different devices and screen sizes and variety of other effects.

CSS is easy to learn and understand but it provides powerful control over the presentation of an HTML document. Most commonly, CSS is combined with the markup languages HTML or XHTML.

Advantages of CSS

- CSS saves time
- Pages load faster
- Easy maintenance
- Superior styles to HTML
- Multiple Device Compatibility

CODE

```
User Model:
var mongoose = require("mongoose"),
  passportLocalMongoose = require("passport-local-mongoose");
//SCHEMA SETUP
var userSchema = new mongoose.Schema({
  username: String,
  name: String,
  email: String,
  phone: String,
  role: String,
  password: String
});
userSchema.plugin(passportLocalMongoose);
module.exports = mongoose.model("User", userSchema);
Server File:
var express = require("express"),
  app = express(),
  bodyParser = require("body-parser"),
  mongoose = require("mongoose"),
  passport = require("passport"),
```

```
LocalStrategy = require("passport-local"),
  passportLocalMongoose = require("passport-local-mongoose"),
  upload = require('express-fileupload'),
  User = require("/models/user"),
  port = 3000;
//Application Configuration
mongoose.connect("mongodb://localhost/DITUniversity");
app.set("view engine", "ejs");
app.use(bodyParser.urlencoded({ extended: true }));
app.use(express.static(__dirname + "/public"));
app.use(require("express-session")({
  secret: "Plagiarism Checker is running.",
  resave: false,
  saveUninitialized: false
}));
app.use(passport.initialize());
app.use(passport.session());
passport.use(new LocalStrategy(User.authenticate()));
passport.serializeUser(User.serializeUser());
passport.deserializeUser(User.deserializeUser());
```

```
app.use(upload());
// app.use(function(request, response, next) {
// response.locals.currentUser = request.user;
// });
//----ROUTES----
app.get("/", function(request, response) {
  response.render("landing");
});
app.get("/home", isLoggedIn, function(request, response) {
  response.render("files/home", { error: "", filename: "", message: "" });
});
//Authentication Routes
app.get("/register", function(request, response) {
  response.render("files/register");
});
// app.post("/register", function(request, response) {
    User.register(new User(request.body.user), request.body.password,
function(error, user) {
//
       if (error) {
//
          console.log(error);
```

```
//
         //alert("Something went wrong. Please try again!");
//
         response.redirect("/register");
//
       }
       passport.authenticate("local")(request, response, function()\ \{
         response.redirect("/home");
//
//
      });
// });
// });
app.post("/register", function(request, response, next) {
  User.register(new User(request.body.user), request.body.password, function(error,
user) {
    if (error) {
       console.log(error);
       //alert("Something went wrong. Please try again!");
       response.redirect("/register");
     }
     next();
  });
}, passport.authenticate("local", {
  successRedirect: "/home",
```

```
failureRedirect: "/home"
}));
app.get("/login", function(request, response) {
  response.render("files/login");
});
app.post("/login", passport.authenticate("local", {
  successRedirect: "/home",
  failureRedirect: "/login"
}), function(request, response) {});
app.get("/logout", function(request, response) {
  request.logout();
  response.redirect("/");
});
app.post("/upload", function(request, response) {
  if \, (request.files) \, \{ \,
     var file = request.files.upfile,
```

```
name = file.name,
       type = file.mimetype;
    var uploadpath = __dirname + '/uploads/' + name;
    file.mv(uploadpath, function(err) {
       if (err) {
         console.log("File Upload Failed", name, err);
         response.render("files/home", { error: "true", filename: "", message:
"Something went wrong. Please try again!" });
       } else {
         response.render("files/home", { error: false, filename: name, message: "
uploaded Successfully!" });
       }
    });
  } else {
    response.redirect("files/home", { error: "true", filename: "", message: "No file
selected!" });
  }
});
function isLoggedIn(request, response, next) {
  if (request.isAuthenticated()) {
    return next();
```

```
response.redirect("/login");
}

app.listen(port, function() {
   console.log("Server is running on port " + port);
   console.log("Press CTRL + C to stop the server.");
});
```

Future Scope

After starting the server, landing page is loaded from where only 2 requests can be made. One to the login page, second to the sign-up page for new user,. We see 2 more requests in future about page showing information about DIT University and last to the about page showing info about the project. During registration of a new user, user is asked about his details and it is also marked whether the user is faculty or a student. Then, the user logs in by his username and password.

After login, homepage appears based on the role of user, different page for faculty and different for students. Where students are only allowed to see and edit their profiles and upload the assignment, report files to their respective faculty while faculties have have the authority to see all the files uploaded by the students and operate on those files. Operations like checking files for plagiarism and generating results, remarks for each submission, grading each student. Later students can check these grades or remarks made by the faculties on their submission.

1 Plagiarism Detection Methods

Detection of plagiarism in text document with high accuracy is a challenging task. In the past two decades, a large number of methods have been reported by researchers to handle this task. These methods can be classified into eleven distinct categories. Some prominent methods under each of these categories are discussed next.

- 1. Character-Based Methods: Mostly plagiarism detection method belongs to this category. These methods exploit character-based, word-based, and syntax-based features. It utilizes these features to find similarity between a query document and existing documents. However, the similarity between a pair of documents is estimated using both exact matching and approximate matching. In exact matching, every letter in both the strings must be matched in the same order. Our survey reveals that most detection techniques are developed based on n-gram or word n-gram based exact string similarity finding approach. F use character 16-gram matching, whereas the authors of use word 8-gram matching. Similarly, some researcher has made an effective use of approximate string matching approach. This string matching shows degree of similarity/dissimilarity between two strings.
- 2. Vector-Based Method: Here, lexical & syntax features are extracted and categorized as tokens instead of strings. The similarity can be computed using various vector similarity measures like Jaccard, Dice's, Cosine, Euclidean and Manhattan coefficients. Our observation is Cosine coefficient and Jaccard coefficients are popular and effective in finding similarity between two vectors. Cosine coefficient in detecting partial plagiarism without sharing documents content. Hence it is very useful to detect plagiarism in documents where submissions are considered as confidential.
- **3. Syntax-Based Methods:** These methods exploit syntactical features like part of speech of phrase and words in different statements to detect plagiarism. The elements of basic POS tag are verbs, nouns, pronouns, adverbs, prepositions, conjunctions and interjections. In , the authors use POS tag features followed by string similarity metric

to analyse and calculate similarity between texts. The authors of use syntactical POS tag to represent a text structure as a basis for further comparison and analysis.

- 4. Semantic-Based Methods: A sentence may be defined as an ordered group of words. Two sentences may be same but the order of their words may be different. The degree of similarity between two words used in knowledge-based measures by Gelbukh is calculated using information from a dictionary. This similarity between two words is used as semantic similarity between two words. In another approach, Resnik used WordNet to calculate the semantic similarity, whereas, Leacock's et al., determine semantic similarity by counting the number of nodes of shortest paths between two concepts.
- **5. Structure-Based Methods:** Unlike those methods above, developed based on lexical, syntactic, and semantic features of the text in documents to find similarity between two documents, a structure based method uses contextual similarity such as how the words are used in entire documents. However, our survey can find a few methods of this category. In the author detects plagiarism in 2 steps. 1st step performs document clustering and candidate retrieval using tree-structure feature representation and 2nd step detects by utilizing ML-SOM.
- 6. Methods for Cross-Lingual Plagiarism Detection: Cross-lingual plagiarism detection is a very challenging task. It require in depth knowledge of multiple languages. Finding appropriate similarity metric for such method is also an important issue. This type of methods work based on cross-lingual text features. Various types of these methods include cross-lingual syntax based methods and cross-lingual dictionary based method. A statistical model is used to evaluate the similarity between two documents regardless of the order in which the terms appear in suspected and original documents.

Similarity Percentage Calculation

Similarity percentage is the comparation degree of percentage similarity between two tested documents. This similarity will give a result of a score which will be a reference for determining percentage similarity degree on a these documents. The number of similarity percentage is affected by the similarity degree from these documents. If similarity percentage is higher, then similarity degree will be higher. Equation for calculating similarity percentage by dividing number of similar sentences detected with total sentences in document. There are three ways for determining similarity between documents:

- Testing results lower than 30% means these document considered have little plagiarism.
- Testing results between 30-70% means these document considered have moderate plagiarism.
- Testing results more than 70% means these document considered have heavy plagiarism.

References
• https://www.plagiarism.org/article/what-is-plagiarism
 https://nevonprojects.com/online-assignment-plagiarism-checker-project-using-data- mining/
• https://elysiumpro.in/plagiarism-checker-tool/
xxxiii

22pp.pdf

ORIGINALITY REPORT

Student Paper

63% 19% SIMILARITY INDEX INTERNET SOURCES **PUBLICATIONS** STUDENT PAPERS **PRIMARY SOURCES** arxiv.org 26% Internet Source www.filemakr.com Internet Source developer.mozilla.org Internet Source nodejs.org Internet Source en.wikipedia.org 5 Internet Source Submitted to Amity University 3% Student Paper Submitted to Symbiosis International University 3% Student Paper easyanimation.in Internet Source Submitted to RK University

10	Submitted to University of Ulster Student Paper	1%
11	www.digitalocean.com Internet Source	1%
12	Submitted to National College of Ireland Student Paper	1%
13	Submitted to University of Newcastle upon Tyne Student Paper	1%
14	Submitted to Birkbeck College Student Paper	1%
15	www.dituniversity.edu.in Internet Source	1%
16	Submitted to Higher Education Commission Pakistan Student Paper	1%
17	blog.pcollaog.cl Internet Source	<1%
18	Submitted to National University of Ireland, Galway Student Paper	<1%
19	Submitted to University of Hertfordshire Student Paper	<1%
20	Azat Mardan. "Pro Express.js", Springer Science and Business Media LLC, 2014	<1%

21	gitlab.doc.ic.ac.uk Internet Source	<1%
22	peaceplantsandpushups.com Internet Source	<1%
23	asknodejs.blogspot.com Internet Source	<1%
24	Submitted to Thammasat University Student Paper	<1%
25	publicationslist.org Internet Source	<1%
26	Submitted to Metropolia Ammattikorkeakoulu Student Paper	<1%
27	www.hackerstribe.com Internet Source	<1%
28	Submitted to Sri Lanka Institute of Information Technology Student Paper	<1%
29	Submitted to Middle East Technical University Student Paper	<1%
30	baadalsg.inflibnet.ac.in Internet Source	<1%
31	gitlab.ecs.vuw.ac.nz Internet Source	<1%

Submitted to RMIT University Student Paper dspace.cusat.ac.in Internet Source Practical Node js, 2014. Publication Submitted to Manchester Metropolitan University Student Paper	<1%
Practical Node js, 2014. Publication Submitted to Manchester Metropolitan University	<1%
Submitted to Manchester Metropolitan University	<1%
University	<1%
	<1%
Submitted to VHS Virtual High School Student Paper	<1%

Exclude quotes On
Exclude bibliography On

Exclude matches

< 5 words