|  |  |
| --- | --- |
| Internship Project Title | Question Creation Responsive Web App with in-built Plagiarism check |
| Project Title | Question Creation Responsive Web App with in-built Plagiarism check |
| Name of the Company | TCS iON |
| Name of the Industry Mentor | Nilesh Haridas |
| Name of the Institute | University institute of engineering and technology , CSJM university kanpur |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Start Date | End Date | Total Effort (hrs.) | Project Environment | Tools used |
| 28 may 2020 | 11 july 2020 | 110 hours | Microsoft VS code, Bash terminal , windows 10 , Morzilla Browser | Node.js , Express.js ,MongoDB , Bootstrap 4 , Tesseract.js , Search.js , and other javascripts frameworks |
| Project Synopsis:  The Application is a Web based application that will have a Web form used to do data entry of Multiple choice Questions into a database. Before each question is saved in the database the user is expected to do a Plagiarism check | | | | |
| Solution Approach:  1:- Structure Building : Build the basic structure of the project by using HTML 5, CSS 3 and BOOTSTRAP  Build forms for the question field with 4 options. (small SMEs)  And having a file input for the file upload which will upload the file and do OCR  Task , a result page was built for the result.  2:-APIs and Technology:- node.js and express.js to create the server , google search API used to search  over internet. Mongodb as database  3:- javascript library:- search.js is used to implement the plagiarism to search over google search  Engine platform.  Tesseract.js is used to implement optical character recognizer (OCR). This also  Javascript best library to recognize the character within the images or pdf file    4:- REST Api :- Rest api is used to build the project and it is best api for the node.js plateform  The systematic folder and file is easy to use for other open source contributor which  Can enhance the project.  5:- Frontend :- It is a fully responsive web view , used bootstrap for the frontend and some media  Query is used to make more responsiveness  6:- Backend :- MongoDb is used to make more dynamic database than that of sql database. | | | | |
| Assumptions:     * I assume that I have good skill to develop this project * There may be small bugs are there. | | | | |
| Project Diagrams: | | | | |
| Algorithms:     * **To get the result of plagiarism**   var length = tosearch.length  if (length > 100) {  return res.status(400).json({ error: 'Max sentences limit crossed!' })  }  // contains all results  var result = []  // count of total number of sentences, and of plagiarsed ones  var count = {  total: length,  plagiarised: 0  }  // iterate over every sentence, find its source, and push to results array  for (let i = 0; i < length; i++) {  const currQuery = tosearch[i]  const a = await search(currQuery)  if (a.length > 0) {  result.push({ text: currQuery, url: a[0].url })  count.plagiarised += 1  } else result.push({ text: currQuery, url: null })  }  // render result page with found results and counts  res.render('result', { result: result, count: count })  })   * **Tesseract.js**   app.post('/upload',(req, res)=>{  upload(req,res,err =>{  (async () => {  await worker.load();  await worker.loadLanguage('eng');  await worker.initialize('eng');  const { data: { text } } = await worker.recognize(`./upload/${req.file.originalname}`);  res.render("upload",{data:text} );  await worker.terminate();  })();  })  }) | | | | |
| Outcome:   * Sentences highlighted in red are plagiarised. Clicking on them will open the source. * Green highlighted are unique question. * Text extracted from images which was uploaded by the users | | | | |
| Exceptions considered:   * No hand written text is in images * Only small and medium question are in input. | | | | |
| Enhancement Scope:  The project has a very vast scope in future. The project can be implemented on intranet in future. Project can be updated in near future as and when requirement for the same arises, as it is very flexible in terms of expansion. For Enhancement of this project it would be more accurate and more type of questions. As a web Developer I used the Same topic for the quote Writing And Shayaree writing. | | | | |
| Link to Code and executable file:- <https://github.com/gauravkmr065/Plagiarism_check_TCS_iON> | | | | |