IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

SURVEY PAPER ON PLAGIARISM DETECTION

Prof. Ms. Shweta Barshe

Department of CSE
Bharati Vidyapeeth College of Engineering, Navi Mumbai, India

Harshit Chiwande
Department of CSE
Bharati Vidyapeeth College of Engineering, Navi Mumbai, India

Satyam Sharma Department of CSE
Bharati Vidyapeeth College of Engineering, Navi Mumbai, India

Vedant Mukadam
Department of CSE
Bharati Vidyapeeth College of Engineering, Navi Mumbai, India

***Abstract*— Plagiarism is a major problem for research. There are different ways to define plagiarism. The concept of plagiarism should be given importance and discussed deeply in relation to research. Plagiarism should be explained as if someone else is using our products such as texts, ideas, or results, therefore implying that it is their own and argue that this is an adequate and fruitful definition. A number of circumstances should be discussed that make plagiarism more or less grave and the plagiariser more or less blameworthy. According to a normative analysis, what makes plagiarism reprehensible is that it distorts scientific credit. In addition, intentional plagiarism involves dishonesty. There are many number of potentially negative consequences of plagiarism.**

***Keywords*— Plagiarism, cheating, similarity detection, IPR**

I. INTRODUCTION

1.1 Defining Plagiarism

There are many definitions of what constitutes plagiarism, and that we will take a look at a number of them in additional detail below. However, consistent with research resources at plagiarism.org, the items that immediately come to mind as description of plagiarism are:

• turning in someone else's work as your own

• copying work or ideas from someone else without mentioning or giving them credit

• not putting a quotation in quotation marks

• giving half truths about the source of a quotation

• copying the sentence structure but changing words of a source without giving credit

• copying such a large amount of words or ideas from a source that it makes up the bulk of your work, whether you give credit or not [Plagiarism.org 2006]

Plagiarism and research has a very broad and wide border-line and is surprisingly murky. After all, advanced research is barely possible by "standing on the shoulders" of others, because it is usually said. In some areas (such as e.g. literature or law) a scholarly paper could comprises a conjecture followed by many quotes from other sources to verify or falsify the thesis. In such case, any try to classify something as plagiarized vs. not-plagiarized just supported a count of lines of words that are taken literally from other sources is absolute to fail. In other areas (like in a very paper in mathematics) it's going to be necessary to quote standard literature just to create sure that readers have enough background to grasp the important part, the proof of a brand new result whose length would be below one third of the paper! In other disciplines like engineering or computing the 000 value of a contribution is also within the device or algorithm developed (that might not even be explicitly included within the paper) instead of the outline of why the device or algorithm is vital that will rather be spelled get into variety of text books. In summary, we believe that there's no valid definition of even textual plagiarism that's not somewhat domain dependent, complicating the difficulty tremendously.

A good survey of further ideas about a way to define plagiarism, and famous samples of suspected or perpetrated plagiarisms will be found within the Wikipedia1 . allow us to now turn, however, to a shot to classify various kinds of plagiarism:

Plagiarism is derived form the Latin word "plagiarius" which suggests kidnapper. It is defined as "the passing off of another person's work as if it were one's own, by claiming credit for something that was actually done by someone else" [Wikipedia:Plagiarism 2006]. Plagiarism isn't always intentional or stealing some things from some one else; it are often unintentional or accidental and will comprise of self stealing. The

broader categories of plagiarism include:

• Accidental: because of lack of plagiarism knowledge, and understanding of citation or referencing style being practiced at an institute

• Unintentional: the vastness of obtainable information influences thoughts and therefore the same ideas may set out via spoken or written expressions as one's own

• Intentional: a deliberate act of copying complete or a part of some one else's work without giving proper credit to original creator

• Self plagiarism: using self published add another form without touching on original one [Wikipedia:Plagiarism 2006] [Beasley 2006].

There is an extended list of plagiarism methods commonly in practise. a number of these methodologies include:

• copy-paste: copying word to word textual contents. • idea plagiarism: using similar concept or opinion which isn't general knowledge.

• paraphrasing: changing some words and grammar, or re-ordering sentences in original work.

• artistic plagiarism: presenting some one else's work using different media, like text, images, voice or video.

• code plagiarism: coping or using someone else program codes, algorithms, classes, or functions without giving credit.

• forgotten or expired links to resources: addition of quotations or reference marks but failing to supply information or up-to-date links to sources.

• no proper use of quotation marks: failing to spot exact parts of borrowed contents.

• misinformation of references: adding someone else references to original sources.

• translated plagiarism: cross language content translation and use without relation to original work.

1.2 Impact

A survey (released in June, 2005) conducted as a part of Center of educational Integrity's Assessment project reveals that 40% of scholars admitted to engaging in plagiarism as compared to 10% reported in 1999 [CAI 2005]. Another mass survey conducted by a Rutgers University professor in 2003 reports 38% of scholars involved in online plagiarism [Rutgers 2003]. These alarming figures show a gradual increase. The new generation is more alert to technology than ever before. Plagiarism now's not confined to mere cut and paste; synonymising and translation technologies are giving a replacement dimension to plagiarism.

Plagiarism is taken into account to be a most serious scholastic misconduct; academia everywhere is undertaking efforts to teach the scholars and teachers, by offering guides and tutorials to clarify kinds of plagiarism and the way to avoid it.

This growing awareness is forcing universities and institutes all around to assist students and college understand the meaning of educational integrity, plagiarism and its consequences. Since plagiarism is commonly connected with the failure to reference or quote properly, many institutions suggest following one in all the recognized writing styles as proposed by major publishing companies.

II. RESPONSE OF ACADEMIC INSTITUTIONS

Although plagiarism is fairly well defined and explained in many forums, the penalty for cases detected varies from case to case and institution to institution, Many universities within the u. s. have well defined policies to classify and cope with academic misconduct. Rules and data regarding it are made available to students during the enrolment process, via information brochures and also the internet sites. Academic dishonesty may be restrained at teacher-student level or institute-student level. The penalties which will be imposed by teachers include written or verbal warning, failing or lower grades and further assignments. The institutional case handling involves hearing and investigation by an appropriate committee, with the accused aware and a part of whole process. The institutional level punishments may include official censure, academic integrity training exercises, social service, transcript notation, suspension, expulsion, revocation of degree or certificate and possibly even referral of the case to legal authorities. To be specific, we've got collected variety of examples: Stanford University: Stanford University provides its students with a well defined academic misconduct policy (Honor Code, operative since 1921) and a decent collection of copyright and enjoyment resources [Stanford Copyright 2006]. in line with a piece of writing within the Stanford daily, the Stanford's office of judicial affairs saw 126 percent increase in honor code violation from 1998 to 2001. This aroused the increasing usage of anti plagiarism software among people at individual levels [Stanford Daily 2003]. As per the Stanford Honor Code "The normal penalty for a first offence includes a one year suspension from the University and 40 hours of community service. Additionally, most school members issue a "No Pass" or "No Credit" for the course within which the violation occurred. the quality penalty for multiple violations (e.g. cheating over once within the same course) may be a three-quarter suspension and 40 or more hours of community service" [Stanford Honorcode 1921]. Yale University: Yale College Executive Committee Yearly Chair Reports [Yale 2005] indicate that the committee had to cope with a sizeable number of plagiarism cases once a year. They show some concern about increase in anysort of plagiarism. There are discussions about its causes and possible preventive measures mentioned within the reports. Punishments vary from case to case ranging from reprimands, probations and lengthening to suspension. Despite clear academic misconduct policies there have been cases of accidental or mistaken plagiarism, which suggests that there's a necessity of more practical ways of communicating details to students. Teachers are encouraged students to avoid plagiarism in any form and should teach citation rules and writing styles to students. U.C. Berkeley: This university also has clear rules and preventive procedures against academic dishonesty and cheating. Instructors are encouraged to resolve the matter personally and issue academic sanctions; just in case an accused person doesn't trust allegations or sanctions, the matter is handed over to student judicial affairs for further investigations and backbone. Teachers are encouraged to coach students about permissible academic conduct. MIT's online writing and communication center [MIT Writing 2006] provides a platform to enhance writing abilities and explains various aspects of plagiarism. in line with a report available at MIT News Office portal, usually the discipline committee must handle 12 to fifteen cases annually with a bent of increase in number of cases in recent years [MIT News 2003]. The penalties follow an identical trend as in other universities, ranging from reduced grades, warning letters, redo of exam or assignment and in extreme cases with recommendation of the discipline committee, suspension or expulsion. In Europe, UK is maybe prior the opposite countries by taking collective measures against plagiarism. Most of the schools have online guides and tutorials available for college kids and researchers, helping them

to grasp academic integrity and improving writing skills. the upper education community in UK took a collective measure by forming a plagiarism consulting service [JISC 2006] giving all UK institutes access to a web plagiarism detection service. Examiners are asked to judge and make recommendations about suspected work but they will not impose any penalty. Oxford University: In March 2005, six cases of plagiarism related had occurred in the college.

The Disciplinary Court prohibited 2 plagiarism cases; in one case the examiners were instructed to disregard the plagiarised work. The candidate failed the exam, but was allowed to reappear for the examination, and if the examiners are satisfied, permitted to re-enter the university. Within the second case, a candidate had previously been convicted of plagiarism by the Court of Summary Jurisdiction. He/she was permitted to submit new work and a few of this was subsequently found to contain plagiarised material.

Elsewhere in Europe, there's also a growing concern and individual efforts are started by teachers at departmental levels to coach researchers and students about plagiarism. At Graz University of Technology, Austria, a Commission for Scientific Integrity and Ethics defines guiding principles to handle cases of plagiarism. a list of possible academic, civil and criminal consequences are ready by end of 2006. Instructors at various institutes of the university started adding information and warnings about plagiarism it slow ago, e.g. figure 1, 2 & 3 show responses to plagiarism cases heading in the right direction websites at various institutes of Technical University Graz.

III. PLAGIARISM TYPES

The use of the net may be a blessing and a curse for the common mass. Crimes of several dimensions get disclosed through the net. One such type is plagiarism. It involves the copying or theft of an ingenious idea and publishing the identical claiming to be one's own.

The labor and creativity of another person cannot get copied so smoothly. it's an immoral act for the one that demands to showcase himself as a resourceful writer. Plagiarism is split further into two types, namely, intrinsic and extrinsic.

3.1 Know the Difference between Intrinsic and Extrinsic Plagiarism

Today, plagiarism broadly are classified into extrinsic and intrinsic plagiarism. When it involves the subject of plagiarism checking, then the most job that the plagiarism checking tools are concluding is remarking extrinsic plagiarism.

In other words, it's just a cursory check on a specific content wherein the intricate details like grammar, parts-of-speech, and other things often get overlooked. The matter gets delivered over the web, together with those flaws. However, with advanced technologies like tongue processing coming into the image, such scenarios can o.k. be handled.

The outward or extrinsic plagiarism is relatively easy to be detected, while the intrinsic plagiarism is sort of hard. the utilization of machine intelligence is significant here. It can detect what form of intrinsic plagiarism gets utilized in the chosen piece.

1. Near copies, intrinsic plagiarism could be a type where a skinny line of differentiation between the chosen text for plagiarism detection and therefore the source. it's unauthorized and unethical to not acknowledge the borrowed idea.
2. To remove plagiarism, disguised plagiarism is used to restrict a copied idea.
3. Translated intrinsic plagiarism may be a type quite clever one. it's the interpretation of a previously used idea during a foreign language, translated within the vernacular language, and copied.
4. The thought is that genre of plagiarism that discusses precisely the identical topic with a change in structure and use of words.

IV. DETECTING PLAGIARISM

Plagiarism detection methods is broadly categorized into three main categories; the foremost common approach is by comparing the document against a body of documents, basically on a word by word basis where documents may reside locally or not. the opposite two approaches aren't exploited the maximum amount, yet may also be surprisingly successful. One is by taking a characteristic paragraph and just doing a pursuit with a decent programme like Google. and also the other is by trying to try to to style analysis; during this case either just within the document in question or performing style comparison with documents previously written by the identical author. this can be usually called stylometry.

Let us look at the three approaches in additional detail:

4.1 Document source comparison:

This approach may be further divided into two categories; one that operates locally on the client computer and does analysis on local databases of documents or performs internet searches, the opposite is server based technology where the user uploads the document and also the detection processes happen remotely. the foremost commonly used techniques in current document source comparison involve word stemming or fingerprinting. This is an approach introduced by Manber [Manber 1994] in which moderately sized strings from a document are compared word to word with preprocessed indexes from other documents. The document then gives a result of similarity approximation among other documents being checked. Figure 1 shows a generic structure of document source comparison based plagiarism detection system.

The core finger printing idea has been modified and enhanced by various researchers to enhance similarity detection. Many ongoing commercial plagiarism detection services justify to have proprietary fingerprinting and comparison mechanisms. The comparison may be local or it will be across the net. Some services utilize the potentials of accessible search engines. Recent steps taken by Google to index the complete text of a number of the world's leading research libraries [Band 2006], and its well-known searching and ranking algorithm makes it a perfect choice not just for open source and free tools but is additionally employed by many commercial service providers and applications. Among some popular commercial and server based approaches justify to use their own search and querying techniques over more extensively indexed internet documents, proprietary databases, password protected document archives and paper mills. The detection services or tools usually represent the similarity findings in an exceedingly report format, by identifying matches and their sources. The findings are then utilized by users of the service to see whether the writing under question is truly plagiarized or whether there are other reasons for match detection. We come to the current later within the paper.

Returning to the problem of paper mills, this term refers to "website where students can download essays, either free or for a charge. Online paper mills usually contain an outsized, searchable database of essays. Most paper mills today provide custom writing services and charge by the page. There are variety of internet sites that even list paper mills.

4.2 Manual search of characteristic phrases

Using this approach the teacher or examiner selects some phrases or sentences representing core concepts of a paper. These phrases are then searched across the web using single or multiple search engines. allow us to explain this by means of an example.

Suppose we detect the subsequent sentence during a student's essay

"Let us call them eAssistants. they're going to be not much bigger than a mastercard, with a quick processor, gigabytes of internal memory, a mix of mobile-phone, computer, camera"

Since eAssistant is an uncommon term, it is smart to input the term into a Google query. "They're going to be not much bigger than a mastercard, with a quick processor, gigabytes of internal memory, a mixture of ... www.jucs.org/jucs_9_4/the_future_of_pcs/Maurer_H_2.html - 34k -"

This proves that without further tools the scholar has used a part of a paper published within the Journal of Universal Computer Science8 . it's clear that this approach is labor intensive; hence it's obvious that some automation will add up, as is completed in SNITCH [Niezgoda & Way 2006].

4.3 Stylometry

Stylometric analysis is predicated on individual and unique writing forms of various persons. The disputed writing is evaluated using various factors within the identical writing. Or it is cross compared with previous writings by the identical author. The detection of plagiarism inside the document range or without any external reference is well described as "intrinsic plagiarism detection" by Eissen and Stein [Eissen & Stein 2006]. This approach requires well defined quantification of linguistic features which might be wont to determine inconsistencies within a document. in keeping with Eissen and Stein "Most stylometric features fall in one amongst the subsequent five categories:

(i) text statistics

(ii) syntactic features, which measure genre at the sentence-level,

(iii) part-of-speech features to quantify the utilization of word classes,

(iv) function word sets to count special words, and

(v) structural features, which reflect text organization."

As an example of easy generic intrinsic plagiarism analysis allow us to take the subsequent paragraph..

"Our goal is to spot files that came from the identical source or contain parts that came from the identical source. we are saying that two files are similar if they contain a big number of common substrings that aren't too small. we'd wish to find

enough common substrings to rule out chance, without requiring too many in order that we are able to detect similarity whether or not significant parts of the files are different. However, my interest in plagiarism lies within academic institutions, therefore the document domain are going to be local research articles. The limited scope of domain will make it easier to see if it's same source or not."

A careful reading reveals the subsequent inconsistencies:

• There's a change in pronoun from "our/we" to "my"

• The author used the article "the" with "same source" in two sentences and missed the article in another.

The bold words show the inconsistency and thus exhibit the chance of plagiarism, where the author took text from some source not matching the general genre. This approach may be hard to use just in case of collaboratively written communication where multiple writers are contributing to one source.

Cross comparisons include a check on change of vocabulary, common spelling mistakes, the employment of punctuation and customary structural features like word counts, sentence length distributions etc. So as to further explain stylometry and another approach, we glance at a service by Glatt [Glatt 2006], which uses Wilson Taylor's (1953) cloze test. during this approach every fifth word in an exceedingly suspected document is removed and therefore the writer is asked to fill the missing spaces. the amount of correct responses and answering time is employed to calculate plagiarism probability.

V. PROPOSED METHOD

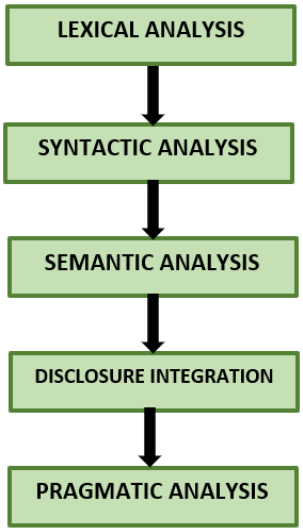
Artificial intelligence is a blessing for plagiarism checker. tongue processing may be a process to detect plagiarism hidden within. Natural Language Processing or NLP is the process to extract materials from the raw and unconstructed data.

The NLP process can make the full data complex, and also the supervised process is alleged to be the foremost used NLP process, among the opposite. the various sorts of NLP processes are explained and elaborated below:

- **NLP based on Semantic analysis:** this can be a process wont to detect plagiarism between two words or more and whether or not they are near in meaning with one another or semantically same. After comparing it gets deduced, the smaller the worth, the more is that the similarity between the words.

- **NLP based on Lexical analysis:** the tactic detects plagiarism involving the structure and grammar usage in a very sentence. In any NLP, the chosen text gets divided into tokens or words, while trying to find similarity or dissimilarity within the text. Structural copying are detected and besides problems in structures are pointed, and necessary changes are done well ahead. However, this process has its drawbacks and could be a bit imperfect.

- **NLP based on Syntactic analysis:** almost like the other NLP after the breakdown of the sentences into tokens, each portion is compared with the grammar or vocabulary used. After that, the ultimate decision depends on whether the words are used correctly and are grammatically error-free.



The final nod is given only after studying the choice tree provided by the structural scaling of the sentences. For structural analysis the machine learning algorithm is as follows:

1. **Top-down parsing**- it starts with the sentence then comes all the way down to the paraphrasing of a phrase and phrase.

2. **Bottom-up parsing**- in contrast to the above one here, the parsing starts with the primary word then emerges each sentence to make a tree-like structure.

3. **Depth parsing**- it searches for the deepest node or the fundamental unit before plagiarism check and so reaches resolute the larger areas of the tree.

4. **Repeated programming parsing**: it only reduces the discrepancy, if any, within the article and resolves all ambiguities to assist present the article efficiently.

5. **Dynamic programming**: it's a just a partial variety of repeated parsing, utilized by some plagiarism checkers as a tool.

• NLP on grammar analysing: The framing and rephrasing of a selected sentence are often done by copying it in exact from the source. The mentioned process analyzes the identical with the algorithm by breaking the structure into smaller units. Each structure or unit gets compared with

the grammars provided and whether or not they are correct or simply a copy-paste from another original piece.

The cumulative above discussion beat all aims towards a presentation of a creative piece. Plagiarism, whether intentional or not, should be avoided at all costs. All the known styles of language processing methods check and recheck the use of language, grammar, and similar words, for the improvisation of the document.

Machine learning algorithm deduces the human language by coding-decoding the provided document in smaller units and ease out the method of plagiarism check. It's necessary because specific minutes, hidden language plagiarism can be overseen by the reader.

It is unattainable to find out every possible error. For grammatical and syntactical errors, the method gets completed during a short span, and for the remainder part i.e., to test to repeat from source NLP completes it perfectly.

5.1 What history says about Natural Language Processing?

The invention of the algorithm method take back to about 1950. Alan Turing discovered it, to stress more on the utilization of computer language for correctional methods. Further improvisation was done by Georgetown experiment that involved full conversion of the Russian language into English.

However, gradually, the funding for translation machine intelligence was drastically gone down. Despite the obstacles, people understood the requirement for developing computing in a very more useful way, since plagiarism was a controversy that's today having a disruptive effect on academic and research papers. the necessity to develop something called an internet plagiarism checker sprung up from this crisis.

The idea was born back after the 1980s when the Turing method got implemented in language correction. The age-old grammatical rules developed several difficulties when implemented in machine language.

The process, too, became quite lofty. Hence, the developers processed the algorithmic calculation in such how that it sounds easier even for the pc. Hence, the results are as per the requirements of the reader.

In 2010 deep neural learning started coming to use. Further, the learning process is developed deep to correct the possible problems present in the piece of writing.

5.2 Using Reinforcement Learning for NLP

In order to know how reinforcement learning or RL gets used for NLP, one first must understand what reinforcement learning is. Well, in RL the behavioural psychology is employed on the software agent.

The trial and error method make sure that the software agents learn a specific quite behaviour over a period and increase the cumulative reward during a particular environment. If someone knows the idea behind this particular learning, it becomes easier to point out how it gets used for NLP.

Often people working with machine learning and NLP feel that RL is ideal for NLP because, within the case of NLP, the system is within the process of learning the behaviour of that of the trainer. The simulated ambiance plays a vital role here, where the trial and error method, too, includes a critical part to play.

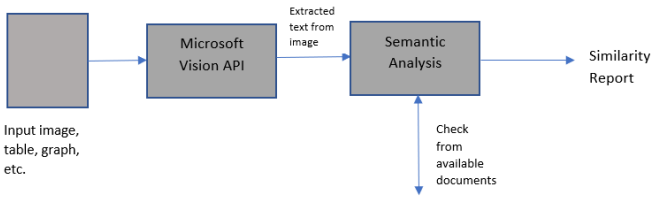
Now so as to grasp this, a specific example may be used here. Suppose during the classification of text process, where the info is there from varied domains, and there's not any training data, an environment and an agent gets created.

The agent tries to classify the text from the information, and within the beginning, it uses some arbitrary methods. After receiving the results of its action, the agent can now decide for the following step.

5.3 How can NLP help in Plagiarism Detection?

The paid or the free plagiarism checkers for college students help in detecting the duplicate content from the initial content. These checkers help in identifying an analogous text, and for that, they use the unique identifier or the structural patterns.

NLP acts as a vital link between programming language and human language, and also the same when blended with the machine and deep learning cause excellent outcomes like the one that gets implemented within the development of Chatbots.



Now, how this gets used in checking plagiarism, NLP makes use of algorithms to check plagiarism. Now, the question is, how does this algorithm add order to place a check on plagiarism?

A straightforward thanks to put it's by parsing or breaking sentences into bits or tokens and processing the identical in pieces. It follows a preferred method that's called 'Latent Semantic Analysis' or 'LSA.'

5.4 How LSA helps in Plagiarism Checking?

LSA includes a very scientific approach towards NLP based plagiarism checking. In other words, it analyses to what extent two words are similar with the assistance of cosine values of the vectors being reproduced by the words that are within the radar of comparison.

The proximity of the values results in a conclusion about the similarity between the words. the method may sound pretty straightforward, but truly, the applying of NLP in plagiarism checking involves lots of mathematical and statistical calculations involving 'Lexical Analysis, Syntactic Analysis,' and even a much-detail approach of the algorithm with keeping in mind of grammar.

5.5 The Other Algorithms of NLP

Apart from these, there are other algorithms in NLP in addition, like 'MinHash or Locality-sensitive Hashing, SimHash and Text Profile Signature' that use even better scientific techniques of checking plagiarism.

However, the fundamental approach is all about breaking and checking sentences first with the words, and so finally, the most idea gets portrayed within the matter.

The plagiarism check supported NLP can also act as a refinement tool for the content as this process removes stop-words or words that are burdening the information without adding any value in an exceedingly sentence.

So, in a way, NLP can play a pivotal role within the field of plagiarism checking and protection of holding rights within the future days to return.

VI. CONCLUSION

It is fair to mention, that current plagiarism detection tools work reasonably well on textual information that's available on the net or in other electronic sources. they are doing break down:

(1) When systematic attempts are put to avoid plagiarism tools by e.g. using extensive paraphrasing with the assistance of synonymizing tools, syntactic variations or different expressions for same contents. (NOTE: most of the higher systems are stable against the order within which paragraphs are arranged: fingerprinting is typically not done on a sequence but on a collection of information, hence order doesn't matter).

(2) When plagiarism relies on documents that aren't available electronically (Since they only are available in printed form, or in archives that aren't accessible for the tool used) .

(3) When plagiarism crosses language boundaries.

Of the three points mentioned above there's hope concerning item (2): more and more material is being digitized, and a few tools have managed to induce access to hidden material in paper mills and such. Item (3) are going to be challenge for a few time to return. We believe that almost all headway will be achieved in reference to point (1) by employing a multiphase approach:

Observe that we've mentioned that the similarity check of alittle set of documents is feasible using rather deep techniques that may determine conceptual equivalence even when heavy paraphrasing is employed. However, those techniques break down if the amount of knowledge becomes overlarge. Hence we predict that the thanks to obtain a successful system that determines whether a specific document x is plagiarized will should work as follows:
A fast algorithm scans the entire available duoverse (the set of all available documents) and eliminates all documents that 'clearly' haven't been used for the document x in question.

The remaining much smaller docuverse is now scanned by a higher algorithm to again reduce the scale of the set of still possible sources used for plagiarism. This continues, until a 'fairly small set' of documents remain that it's feasible to use deep and computing intensive techniques.

Whether the amount of 'passes' should be 2, 3 or more remains to be seen. Since all major plagiarism tools are proprietary it's not known to us what quantity this multipass technique is already in use. it's clear for us from the observations we've got, however, that there's much room for further progress.

In closing we would like to say two further details to which we return in [Zaka & Maurer 2006]:

First, plagiarism isn't confined to academia. it's rampant and still not much recognized in schools, particularly in high schools where many assignments are of the final essay type, precisely the quite stuff easily found on the web. It also appears in an exceedingly different form when government agencies or other organizations commission some 'study' or report back to be compiled: in a very number of cases they get what they require, pay quite some money for it, but what they get is simply obtained by simply copying and pasting and minor changes or additions of existing material. In those cases it's not most a matter to detect plagiarism after the actual fact, but rather have some specialists spend some hours searching on the online if the fabric requested it not available anyway before commissioning a report.

Second, plagiarism is getting many attention in academia without delay. The reaction has been that several universities purchase tools for plagiarism detection. it's our belief that to detect plagiarism at a university you would like quite a software tool: you wish a collection of them, specialists who know the way to figure with those tools, domain experts and also language experts if we ever want to travel beyond the boundary of 1 language. this means that a considerable group is critical to try to to good work, and this can't be achieved by anyone university. It requires a joint effort i.e. a middle for plagiarism detection that's run on a national or perhaps supra-national (e.g. European) level.

VII. ACKNOWLEDGEMENT

We would like to take the opportunity to acknowledge the support and help of all who have assisted us in the project. Without their contribution and advice, we would have never been able to progress with the work in the research. Firstly, we would like to sincerely acknowledge my research mentor Ms. Shweta Barshe, for her guidance, support, technical knowledge and encouragement in the whole research process and work. The information and feedback provided were extremely helpful and useful for this paper.

VIII. REFERENCES

[1] https://scholar.google.com/citations?view_op=view_citation&hl=en&user=GCecY8cAAAAJ&citation_for_view=GCecY8cAAAAJ:eQOLeE2rZwMC

[2] http://journals.resaim.com/ijresm/article/view/738

[3] https://scholar.google.com/citations?view_op=view_citation&hl=en&user=GCecY8cAAAAJ&citation_for_view=GCecY8cAAAAJ:Y0pCki6q_DkC

[4] https://scholar.google.com/citations?view_op=view_citation&hl=en&user=GCecY8cAAAAJ&citation_for_view=GCecY8cAAAAJ:Y0pCki6q_DkC

[5] https://scholar.google.com/citations?view_op=view_citation&hl=en&user=GCecY8cAAAAJ&citation_for_view=GCecY8cAAAAJ:YsMSGLbcyi4C

[6] https://scholar.google.com/citations?view_op=view_citation&hl=en&user=GCecY8cAAAAJ&citation_for_view=GCecY8cAAAAJ:u5HHmVD_uO8C