Evaluation of cheating detection methods in academic writings

Ahmed Patel

*Faculty of Computing Information Systems and Mathematics,
Kingston University, Kingston upon Thames, UK and
Centre of Software Technology and Management (SOFTEM),
School of Computer Science, Universiti Kebangsaan Malaysia, Bangi, Malaysia,
and*

Kaveh Bakhtiyari and Mona Taghavi

*Centre of Software Technology and Management (SOFTEM),
School of Computer Science, Universiti Kebangsaan Malaysia, Bangi, Malaysia*

Abstract

Purpose – This paper aims to focus on plagiarism and the consequences of anti-plagiarism services such as Turnitin.com, iThenticate, and PlagiarismDetect.com in detecting the most recent cheatings in academic and other writings.

Design/methodology/approach – The most important approach is plagiarism prevention and finding proper solutions for detecting more complex kinds of plagiarism through natural language processing and artificial intelligence self-learning techniques.

Findings – The research shows that most of the anti-plagiarism services can be cracked through different methods and artificial intelligence techniques can help to improve the performance of the detection procedure.

Research limitations/implications – Accessing entire data and plagiarism algorithms is not possible completely, so comparing is just based on the outputs from detection services. They may produce different results on the same inputs.

Practical implications – Academic papers and web pages are increasing over time, and it is very difficult to capture and compare documents with all available data on the network in an up to date manner.

Originality/value – As many students and researchers use the plagiarism techniques (e.g. PDF locking, ghost-writers, dot replacement, online translators, previous works, fake bibliography) to cheat in academic writing, this paper is intended to prevent plagiarism and find suitable solutions for detecting more complex kinds of plagiarism. This should also be of grave concern to teachers and librarians to provide up to date/standard anti-plagiarism services. The paper proposes some new solutions to overcome these problems and to create more resilient and intelligent future systems.

Keywords Anti-plagiarism, Cheating, Plagiarism detection, Plagiarism prevention, Ghostwriting, Source code plagiarism, Computer crime, Research work

Paper type Research paper

1. Introduction

University assignments are one of the important, but time-consuming, sometimes boring, and confusing requirements of the lectures. Lecturers mostly evaluate the assignments as a part of the final mark, which usually can be assessed up to 90 per cent of the total mark based on the course and total number of assignments. Every lecturer assigns a given topic for the assignment and stipulates a specific time of deadline for submitting the results. This is when the students should start preparing their papers to submit before the deadline. Almost for every assignment, there are some students who may not be able to write or complete their own works (Hawley, 1984).

As previously seen in universities, students attempt to copy previous works in the belief that the lecturer would not compare all the papers with one another. They may even try to copy from the internet on the assumption that the professor has not memorized all the web pages over the internet (Harp and Taietz, 1966; Marsden *et al.*, 2005; McCabe, 1999).

The previous concept was true until anti-plagiarism services were established. Search engines appeared to reveal plagiarisms in intellectual properties and academic papers. Search engines can search certain keywords in documents on the internet. The first idea for plagiarism detection was searching paragraphs using search engines to find the original corpus of the text, and it failed in the cases where text was mixed with multi source corpuses or mixed mode (image, text, audio) documents. Thereafter, plagiarism detection services came into existence primarily to identify sentences and phrases in a huge amount of documents. Many factors are taken into account and evaluated for plagiarism detection based on the search scope, detection algorithm, database type, database hosting, as well as precision and recall mechanisms (Carrol, 2002). One of the first practical mechanisms was done in 1995 under the name of SCAM in Stanford University (Shivakumar and Garcia-Molina, 1995). In this project, a signature based schema, and duplicate detection had been applied. Signature based documents are equipped with a special signature for each author in the document such as a watermark (Boneh and Shaw, 1995). Duplicate detection system detects and deletes same documents in the system to improve accuracy and detection performance.

The other concept in plagiarism is "Software plagiarism" or "Source-code plagiarism". Parker and Hamblen in 1982 described software plagiarism as "a program which has been produced from another program with a small number of routine changes" (Parker and Hamblen, 1989).

A wealth of research with numerous proposed solutions and algorithms for source code plagiarism have been done in the last two decades as can be observed by these references (Ahtiainen *et al.*, 2006; Belkhouche *et al.*, 2004; Cosma and Joy, 2006; Culwin *et al.*, 2001; Niezgoda and Way, 2006; Parker and Hamblen, 1989; Prechelt *et al.*, 2000; Tahaghoghi and Tahaghoghi, 2006; Vamplew and Dermoudy, 2005; Verco and Wise, 1996; Wiedemeier, 2002; Wise, 1996).

But still there are a few daunting questions:

· Did anti-viruses prevent viruses from being created?

· Did anti-viruses decrease the raising amount of viruses?

· Did firewalls block the hacking activities?

There are so many more similar questions in IT and non-IT fields. Obviously, it is clear that the answers to the previous questions are "NO". These answers do not oppose the

positive efforts done in security, but also it declares that all these efforts can be broken as the time goes on. Similarly, breaking into the anti-plagiarism systems follows the same natural rule.

University authorities, students, tutors, lecturers, editors, librarians, system developers and service providers should be aware of the techniques used to avoid detection or break into anti-plagiarism services. They should understand the benefits and negative points of services to assist students properly. Alternative cheating methods in academic writings appears to be a continuous challenge that can only be met by sophisticated detection anti-plagiarism systems and services. This research work provides an evaluation of alternative plagiarism and detection systems, which can also assist librarians in particular as follows:

- to establish and to invoke plagiarism policies of institutions; and
- to clarify plagiarism detection service reports in order to determine and explain plagiarism mark-ups and percentages to their users in a more congenial setting.

2. Methodology
There are nine methods described in this paper, which are used as tools to prepare quick academic assignment papers to bypass plagiarism checkers and to submit them to the lecturer. These techniques cannot be solely detected by plagiarism detection service providers, so they must be predicted, to catch plagiarism, in the writings through other means, like lecturer/teacher experience.

2.1 PDF locking
PDF is one of the most useful formats for documents, which can be opened on every computer in recent years, and it is widely supported all over the world. This format is introduced by Adobe Corporation in Adobe Acrobat software, and has been approved as a document standard at ISO in, 2008 (ISO, 2008).

There are some security features in PDF format, which enable the publishers to limit document access and activities. They also let users select features through enable or disable options according to their intentions. Some of these security features are as follow:

- printing;
- document assembly;
- content copying;
- content copying for accessibility;
- page extraction;
- commenting;
- filling of form fields;
- signing; and
- creation of template pages.

All the previous security features can be modified or over-ridden by using various applications or Adobe Acrobat Professional.

When a paper is to be submitted to a lecturer, it is a requirement that PDF files can be opened and be readable without any password prompt. Also, the lecturer may impose the condition that the paper should be printable, so the student needs to enable the printing option.

There are essentially two features, "Content copying" and "Content copying for accessibility" that can be disabled, so the lecturer will not be able to copy the sentences and phrases to search over the internet. Most of the anti-plagiarism services (e.g. turnitin.com) raise an access error without any elaboration of the problem, which has occurred. It would raise "Password Protected Document Error". Therefore, it is frustrating for the lecturer to understand the origin of the problem as the document can be opened locally without any password prompt.

Online plagiarism detection systems should be equipped with a more comprehensive error handling system and descriptions to inform the user about the possible solutions to overcome the problem.

2.2 Ghostwriters

A ghostwriter is an expert author who is secretly paid for writing on behalf of the others. Ghostwriters write papers, books, articles, songs, speeches and so on. They are mostly hired by politicians, celebrities, and famous authors, as well. The transcript would be published in the magazines, books, or even delivered in a speech on behalf of the speaker, or the book's author. They are mainly contracted to avoid being revealed for their ghostwriting role (Burnham, 2003; Rennie and Flanagin, 1994).

Ghostwriters are also hired by students to write their essays, papers, articles, dissertations and theses. They research widely to write their texts, but in some other cases where the text content is not really important, they write something in academic format even if it is illogical. The fee that is charged for their services depends on the type of writing, required research type, research requirements and the deadline. The fee has a range (Giombetti, 1992) from $10 to $100s per page.

Table I shows a sample of costs for academic ghostwriting by one of the service providers during the first quarter of 2011.

Table II illustrates the fee differences among the levels of writing required for the text.

Machine learning methods can be used in online plagiarism services to learn the author's style of writing, and to detect others' works. This solution is elaborated in section 5.

Time	Cost ($) per page
Within three hours	42.95
Within six hours	39.95
Within 12 hours	32.95
Within 24 hours	29.95
Within 48 hours	22.95
Within three days	19.95
Within five days	19.55
Within seven days	19.15
Within ten days	18.55
Within 14 days	17.75
Within one month	14.95

Table I.
Ghostwriting fees based on the deadline

2.3 Dot replacement
Anti-plagiarism algorithms primarily work on sentence similarities with the other papers. This similarity percentage varies among algorithms. Sometimes in a phrase, which consists of ten words, 40 to 60 per cent and higher with similarity is assumed as plagiarism. This method is applicable to sentences with more than three words; but in the sentences with less than three words, the algorithm stops processing, because processing a single word for plagiarism is illogical.

Dot replacement method can help to stop anti-plagiarism service from processing copy-pasted texts. It is only required to replace SPACE with DOT "." character, then change DOT color to white. This trick works! In this situation, a lecturer will not understand whether SPACE or any other character have been used at the first sight. Lecturers should copy-paste the text into a plain text box or change all the text color to a single color to identify the used trick. Moreover, as mentioned before, if the paper is in the PDF format, content copy feature can be disabled to prevent the user from making a copy of the text inside.

Anti-plagiarism services identify each word as an independent sentence, and they skip further processing on 1-word sentences, so when SPACEs are replaced with DOTs, all the words would be considered as one-word sentences.

Most anti-plagiarism services do not process the text in a rich-text format to identify whether dots in the paper are in the background color. They mostly convert the rich-text to the plain text for easier and faster processing.

2.4 Word replacement
As described previously, Dot Replacement is only required to decrease the similarity of the sentences. Also replacing the words of the sentence with their synonyms can decrease the similarity percentage. Combining this method with Dot Replacement increases the trust level of a safe plagiarism. Free software for this method is easily accessible.

Processing and comparing the text with synonyms is a suitable technique for detecting the similarities. This technique is used by internet search engines for more accurate information or document retrieval. This process is computationally massive which generates tones of additional sentences, and nowadays online plagiarism detection systems do not support this function.

2.5 Online translators
Paraphrasing is a very time-consuming technique, which basically demands a good command of English for the paper author. Unfortunately/fortunately, no sufficient paraphrasing software application is available on the market at the time of writing this paper; but there are other applications that can be used to paraphrase through applying some minor tricks.

Level	Surcharge fee
Undergraduate and below	Same as Table I
Graduate	+$3/page
Corporate research	+$3/page
PhD	+$6/page

Table II.
Ghostwriting texts fees based on the level

Online translation services, like Google Translator (Google, 2011) and SYSTRAN Soft (SysTranSoft.com, 2011), can be employed to paraphrase. Hence eastern and western languages have different structures; new text would be generated, if a recursive iterated translation is done.

For example, assume the original text is in English. If first, this text is translated to Chinese, and then back to English, the final text is a paraphrased version of the original one. These procedures can be applied repeatedly. Figure 1 shows the logical recursive steps for online translation.

Using multi languages for translation can be used to obtain a better or more sophisticated result of the created work with the likely proviso to avoid detection. For example, from English to Spanish, Swahili, Persian, Malay, French, Chinese, and then back to English. As it is clear from the previous example, switching between western and eastern languages can drive us to our desired destination in greater comfort.

As an example, follow the following sample text translation by Google Translator (Google, 2011):

English (original text)
"Hybrid systems have a particular attraction in that they link two types of elements that are prominent in reactions to emotion-articulate verbal descriptions and explanations and responses that are felt rather than articulated" (Cowie *et al.*, 2001).
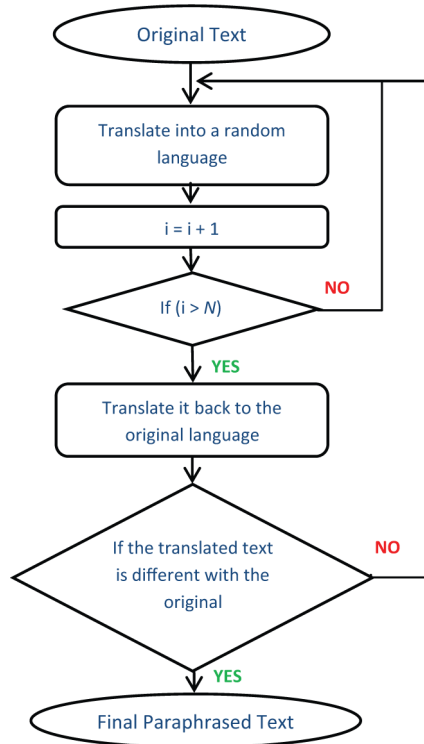

Figure 1.
Illustrating the recursive translation flow

Translated to Spanish
"Los sistemas híbridos tienen un atractivo muy particular en que se vinculan dos tipos de elementos que ocupan un lugar destacado en las reacciones de emo-ción, articular las descripciones y explicaciones verbales y las respuestas que se hacen sentir en lugar de articulados".

Translated to Swahili
"Hybrid mifumo ya kuwa na kivutio hasa katika zilizounganishwa mbili aina ya mambo ambayo ni maarufu katika reactions ya hisia, kutoa maelezo ya maneno na maelezo na majibu ya kwamba ni badala ya kujisikia ilitoa".

Translated to Persian

سیستم های هیبرید دارای جاذبه، به خصوص در دو
نوع مربوط به چیزهایی میشود که محبوبیت خود را در
اثر هیجان را برای شرح مفصلی شفاهی و توضیحات و
پاسخ است که به جای احساس فراهم شده است.

Translated to Malay
"Hybrid sistem dengan graviti, khususnya dalam dua perkara berkaitan yang popular dalam kegembiraan Rabray keterangan terperinci dan penjelasan dan merespon secara lisan daripada perasaan yang disediakan".

Translated to French
"Les systèmes hybrides avec la gravité, en particulier dans les deux questions sont très populaires dans la description Rabray fun et des explications détaillées, et de répondre aux sentiments de vive voix fourni".

Translated to Chinese (simplified)

"混合动力系统与重力有关，尤其是在作为一个细节和解释和答复热情，
它的普及问题的两种类型，提供口头感情。"

Translated back to English
"Hybrid system with gravity, in particular in the detail and explanation and as a warm response, and its popularity are two types of problems, provide oral feelings".

Now, compare the first English text with the last translated text back to English. Not only the sentence structure and words have been changed, but also there are many grammatical mistakes, meaning and idiomatic expressions changes, and also the text concept may have been changed.

In spite of the fact that this technique may change some meanings of the original text, it can also make it difficult to understand. If a student is an English native speaker, he/she should avoid this method, because the output text would have many grammatical errors that are unusual in a native speaker's text. As mentioned before in the introduction section, only a few lecturers spend an adequate amount of time on reading the assignment papers to catch plagiarism at this level.

2.6 Papers in other languages

Sometimes it is easier to search for a similar research paper in other languages. Having the translated papers or using the online translators would facilitate paper submission without any plagiarism. At the moment, plagiarisms are only considered in the original languages of the paper on the translations that are published.

Nowadays, many conferences and competitions are being held to invent new methods to detect plagiarism across languages (CLEF, 2011).

2.7 Fake bibliography

References are one of the most important evaluation factor in academic writing to determine whether the candidate has performed thorough research and investigation of the subject or not. Depending on the type of writing, references are viewed from different perspectives by assessors to determine the quality of the submitted writing.

Manuscripts with more references, papers and books generally give the impression that they are of better quality. They also lengthen them in terms of the number of lines and pages to give the impression that size means better work. Moreover, usually producing manuscripts with more length helps to reduce the percentage of plagiarism. It is a "catch 22" scenario that one has to guard against. Assume a two-pages manuscript with one page of bibliography: if all the text is copied, just 50 per cent of plagiarism would be reported, because the other 50 per cent will be the fake references. It means that the fake bibliography by the forged authors' names, paper titles, conferences/journals, and published year, which does not exist, in reality, can be inserted in the reference list.

For this case, some of the anti-plagiarism services such as iThenticate Cross Check service (iThenticate.com, 2011a) ignore the bibliography automatically, and some others such as Turnitin (iParadigms, 2011) process the bibliographies as well.

2.8 Text image

All the previous techniques and methods described for cheating an anti-plagiarism service were about the text-based papers. There is a noticeable weakness on processing the text inside images, and so far the images will be ignored on processing. A new initiative, which becomes very suitable with a high level of success, is using images and pictures instead of texts. Only creating images with good resolution consisting of the text we are willing to type is enough to overcome and cheat anti-plagiarism systems.

Using an accurate OCR system can be useful to identify the texts in images and process them for plagiarism detection (Holley, 2009; Tappert *et al.*, 1990).

2.9 Plagiarism removers – grammar checkers

Besides the previously declared techniques, there are some other applications and online services that can be applied to remove plagiarism from papers. Among the online services, "RemovePlagiarism.com" (RemovePlagiarism.com, 2011) can detect plagiarized text and tries to paraphrase it. These types of online services can be used to change the texts easily in a few seconds or minutes, but on the other hand, they may charge the user for such services.

Grammar checking applications are other useful methods that try to alert one to the best structure and words for a given paragraph. In some cases, they may even suggest some word replacements, and new grammar structures. This software does not change

all the sentences necessarily, but it may help the authors and students to change some sections or even make better and more suitable grammar structures. "White Smoke" system is one commercial sample of these applications (WhiteSmoke-Inc., 2011).

3. Plagiarism detection services

There are many anti-plagiarism and plagiarism detection online services that vary with each other based on their techniques, fees, reporting times, academic papers access, official supporters, and other features. The most popular ones are turnitin.com (iParadigms, 2011), ithenticate.com (iThenticate.com, 2011b), and plagiarismdetect.com (PlagiarismDetect, 2011). These services compete with each other by making their own databases more comprehensive, as well as improving the algorithms for plagiarism detection. Among these web sites, PlagiarismDetect.com provides a plug-in for Microsoft Word. This web site analyzes the plagiarism online while you are typing your text. All the previous services charge the users a fee based on the usage and membership. ACM, IEEE, Emerald, Elsevier, Springer and many other official academic manuscript centers support these online plagiarism detection services as part of the reviewing and typesetting functions.

4. Pros and cons of plagiarism detection services

4.1 Turnitin.com

Turnitin.com is an old and primary online plagiarism detection service on the internet. It holds, or can access a huge database of academic papers, university assignments, dissertations and theses. It has many members and clients from universities and institutes all over the world to check plagiarism in their own academic papers (iParadigms, 2011; turnitin.com 2010a, b).

Despite this popularity and huge database, its detection algorithm is old enough to be broken by the previously mentioned techniques. Besides, it may not exclude the paper's title and bibliography for document processing, thus causing plagiarism detection in all references, including the author's name. Users should mainly calculate the real plagiarism percentage by removing the author's name, title and bibliography. Also many common phrases such as algorithm names, people names, places, addresses or even popular academic phrases (e.g. "table 1 illustrates x", "as mentioned before") are considered as plagiarism. However, this should not be the case unless a whole complex sentence or paragraph is taken verbatim from another source.

One other major problem in Turnitin is failing with multi-uploads of the same document in the system. When a document is uploaded for processing, it saves it into its database. For the second time after correcting and revising the plagiarism section, your entire document is considered as a copy version, because the previous version of the document was not deleted automatically. It is not very easy to delete it manually either. The web site itself is not intelligent enough to recognize the same authors for the same papers.

4.2 iThenticate

This is a new online anti-plagiarism service, which is mostly supported by IEEE for document submissions. This web site provides a huge database of academic papers, and it is equipped with a system called "Cross-Check" to check the similarity with other papers all over the world (Harrick, 2011; iThenticate.com 2011a, b).

This system uses more flexible algorithms in comparison with Turnitin.com, and it excludes the title, author's name, bibliography and quoted sections automatically before processing the document. As a result, the plagiarism percentage is more accurate than Turnitin.com (IEEE, 2011).

4.3 PlagiarismDetect.com

PlagiarismDetect.com is another service, which is not as popular as the previous ones. One prominent advantage of this service at the time of writing this paper is providing a trial use and a free-downloadable plug-in for Microsoft Word to identify the plagiarism in your text while you are typing your text without switching to the web site and uploading your document (PlagiarismDetect, 2011).

This service is mostly useful for personal usage and a primary plagiarism detection step before submitting the document.

5. Proposed solution

One of the recent challenging problems in plagiarism detection is "author identification" which has been provided in the plagiarism detection systems (CLEF, 2011). In addition with searching and matching the text and sentences in the database, artificial intelligence may come to help in the case of plagiarism detection.

Every author throughout the world follows a specific writing pattern and style. Sometimes it can be recognized as a fingerprint for the documents originality. These patterns can consist of the used number of prepositions, verbs, grammars and so on. Basically, these patterns are consistent throughout the paper, and you would not see anybody write a whole paper in a multi-style writing. Even if one did, it would be extremely rare and a unique piece of genius work.

Patterns matching in a document can help to recognize whether it has been written by one person or not, and also it can be matched with the other documents to find the most probable original author of the text, even if it is not the direct copy-paste of the original one.

Hashing code cryptography methods (Kumar *et al.*, 2010; Seungjin *et al.*, 2010) can be applied to hash the pattern to be matched and find the other similar style papers. In this case, it can even identify the fake authors by finding irrelevant linked texts to them, but plagiarism cannot be detected 100 per cent precisely. This problem can be solved, by using Fuzzy Logic methods (Hájek, 2011; Zadeh, 1965). Fuzzy Logic can assist to determine the probabilistic of the plagiarism in the documents with a hidden source.

Figure 2 illustrates the procedure of proposed solution by using knowledge base and inference engine. Artificial Neural Network is recommended for self-learning section, and Fuzzy Logic can be hired to process the probability of plagiarism in the system.

As it is shown in Figure 2, a new document is given to Plagiarism Detection System (PDS). PDS passes the document to the Inference Engine for pattern recognition in the text. The Artificial Neural Network (ANN) subsystem is self-learning by analyzing documents for each writer and stores the processed data into a Knowledge Base (KB). This KB consists of writing style pattern, text anthology, word semantic and ontology. Writing style pattern is a feature extraction technique used by Natural Language Processing (NLP) processes. It attempts to determine the general style of the text and
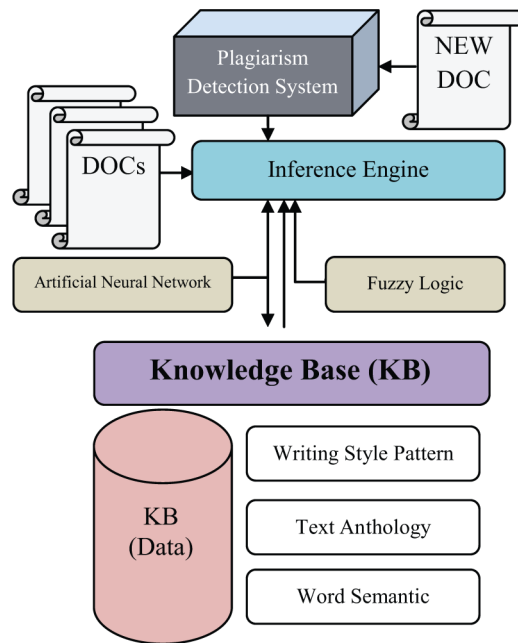
specific words which are used in the text and works by experiencing and self-learning of patterns to infer rule-based expressions. These rules are stored in KB and subsequently utilized to perform matches for better plagiarism detection.

The next action is for inference engine to gather the data based on the anthology and semantic of words and phrases of the text. In this phase, the inference engine can detect the plagiarism in cross language translations. As mentioned before, phrase concepts, semantics, and anthologies of the phrases will be changed, by using computer-based translators to translate recursively. Translators mostly fail in translations of idiomatic expressions that can be detected by semantic processing techniques.

Anthology and semantic hierarchy of the words and phrases, which are stored in a knowledge base for each paper, and author, can be used for further processing by an inference engine and anti-plagiarism detection system. Hence, researchers and commercial companies are still gathering and constructing more advanced hierarchies of anthologies, ontologies and semantics for words and phrases, and this technique is expected to be implemented successfully in the near future. This is the main conceptual idea of processing documents to detect plagiarism with a higher level of matches and confidence by anti-plagiarism services.

6. Cheating in academic writing
The techniques discussed here can be recursively and successfully used for journal manuscripts, conference papers, academic theses, dissertations and librarians as a self-checking guarding mechanism against plagiarism.

In some cases, it is important to meet the deadline of the conference, even though the actual paper is not ready for submission. These techniques can be used to prepare a very fast paper to submit. This document may pass the plagiarism test as most of the techniques have been applied on it. Then the actual paper can be submitted in conference camera-ready form by the deadline. This is a win-win situation for the student since the paper will largely go undetected by plagiarized detecting software and the conference deadline date is fulfilled without a fuss!

7. Discussion

The previously mentioned nine techniques and methods for cheating in academic writings are briefly compared by seven factors in Table III. Level factor indicates the total trust level of the technique used for plagiarism. System Cheat refers to the cheating probability of the system, through which the online anti-plagiarism services can be cheated. System Trust indicates whether the online services are able or can be programmed to recognize these kinds of plagiarisms or not. Human Trust factor means the human ability to recognize cheatings, as well as Lecturer Trust, which identifies the lecturers' ability to reveal the problems in academic papers. Cost and time are the other important factors to assess plagiarism comparison.

As it can be observed from Table III, Online Translators, using Previous Academic Works, and Texts in Images are the top techniques with high level of trust that are used. Online translators are very difficult to be detected as well as being revealed by both machines and humans. It is not easy to understand whether the text is included with plagiarism or not. Cost and time of these techniques make them competitive techniques and easy to use for everyone.

The other techniques like dot replacement, fake bibliography, and word replacement require more time so that sometimes it is faster to write from scratch rather than plagiarizing the text. Furthermore, these techniques demand experience in writing fake bibliographies and choosing suitable words as replacement.

Ghostwriting is a method which uses the others' experiences for writing. It can work well, but considering the cost and time involved, it is not suitable in the most of urgent academic works.

PDF locking with the low trust level can be just used for academic papers in university assignments, but it will not work over a long period of time because plagiarizing detectors will eventually catch up with them quickly.

RemovePlagiarism online service (RemovePlagiarism.com, 2011) and grammar checking software may be applied on the output of online translators to proof grammar mistakes and try to polish the English sentences.

By comparing all the techniques shown in Table III, it can be safely concluded that using online translator services is the most convenient method for plagiarism according to the following reasons: time factor which can be done in just few seconds or minutes, cost factor which is free, and the other trust factors which all are set as high, because it is very difficult for machine and human to recognize that the text is plagiarized by machine translators.

Plagiarism detection service providers should be aware of the cheating techniques and detection solutions to be equipped with a proper system for more precise detection service.

Method	Level	System cheat	System trust	Human trust	Lecturer cheating	Cost	Time for ten pages
PDF locking	Low	Yes	No	No	No	Free	–
Ghost writers	Medium	Yes	Yes	Yes	Yes	10-200	From one to ten days
Dot replacement	Medium	Yes	No	Yes	Yes	Free	Couple of minutes
Online translators	High	Yes	Yes	Yes	Yes	Free	Couple of seconds
Previous works	High	Yes	Yes	Maybe	Yes	Depends	From seconds to two days
Fake bibliography	Medium	Yes	No	No	Yes	Free	Couple of minutes
Word replacement	Medium	Yes	Maybe	Yes	Yes	Free	Couple of hours
Text image	High	Yes	Yes	No	Maybe	Free	Couple of minutes
Remove plagiarism grammar check	Medium	Yes	No	Yes	Maybe	Depends	Couple of seconds

Table III.
A brief overview of the methods and comparisons of techniques

Not only plagiarism detection service providers but also all teachers and librarians should be conscious about the plagiarism techniques and available online anti-plagiarism services to enlighten the students about the consequences and advantages and disadvantages of plagiarizing. It should be noted that the numbers related to plagiarism are not very accurate because of the mentioned weaknesses. Thus, we have to know how to submit the documents to get the best and more accurate plagiarism report. For example, a 30 per cent reported plagiarism is not always a real plagiarism, as it can be mostly related to bibliography, cited sections, or popular phrases.

Among all the previous research works and solutions, Carrol and Appleton suggested practical advices to lecturers and teachers as follow (Carrol, 2002):

- creating a new series of assignments each time the course is taught afresh;
- avoiding assignments that just require information description and data collection without analysis;
- trying to suggest assessments based on the analysis and evaluations with independent critical thinking;
- if possible, to design assignments, which have multiple answers and require students to justify their choice of answers based on research analysis;
- setting a range of interrelated tasks pertaining to the assignment prior to final delivery of the whole completed assignment;
- advising and instructing students in a congenial manner about plagiarism and their negative side effects to their academic writings and personal standing;
- establishing the policy and the serious actions, which will be taken against offenders caught plagiarizing;
- using online plagiarism detection services to effectively validate and remove plagiarized material prior to delivery of assignments; and
- making mandatory that the author of assignments create a cover sheet with a physical hand written or digital signature to prove that the text is the sole work of the author.

No plagiarism is allowed
Plagiarism detection in academic papers and texts is one of the recent open research problems to defeat both copy-pasted texts and cheating techniques. Plagiarism detection in translations across the languages is very challenging to be detected.

Furthermore, sometimes the source of the paper is not identified and it can be very tedious to locate in a large quantity of documents throughout the world.

Author detection and identification is one of the recent methods used by anti-plagiarism services to identify a valid author of the text and to determine whether the author's name is genuine or not (CLEF, 2011)

In the proposed solution, the technique uses Artificial Intelligent methods and concepts to reveal the fake authors and some plagiarisms that do not have clear sources. Some techniques in Artificial Intelligence, which are applicable for this purpose, are pattern recognition, fuzzy logics, and training methods like Artificial Neural Network (ANN) or Support Vector Machine (SVM) (Farquad *et al.*, 2008; Lei *et al.*, 2010; Liu and Liu, 2010).

Manuscript writers generally make their first port of call to libraries and seek help from librarians. Librarians should be aware of the cheating schemes and weaknesses of plagiarism detection systems. They can offer wise advice to writers of overcoming intentional or unintentional plagiarism as well assist them to check the documents and academic works before submitting to anti-plagiarism systems. Librarians can also clarify to users about plagiarism detection service reports to remove illogical percentage of false or inadvertent reporting, and establish more accurate report out of the system reports.

8. Conclusion

The nine techniques and methods for cheating in academic writings and the three plagiarism detection service providers described in this paper were recently used in our university assignments for our assessment of plagiarism. This motivated our institution to re-consider and update its cheating and anti-plagiarism policy.

The review and comparison of the plagiarism techniques illustrated the key factors in determining their effectiveness. A combination of these techniques may even produce more complex results in detecting plagiarism. In addition, the proposed knowledge base for a plagiarism detection system can improve the quality and accuracy of plagiarism detection techniques in advanced systems.

Librarians should advertise the results of our research to authors to check their works against the nine plagiarism tricks mentioned in this paper before submitting it to plagiarism detection systems for assessment. They can access the problems, which are discussed in our research paper, in the final report produced by anti-plagiarism systems to assist their users in determining and interpreting the plagiarized sections and scores of their manuscripts. These results, can also be used by students, lecturers, and academicians, to prevent cheatable tricks in their own writings. Those organizations and researchers involved with anti-plagiarism can use these results to understand and improve the available policies, processes and services in order to create more advanced, efficient and trustable plagiarism detection systems.

Finally, everyone should be aware that plagiarism does not pay. Therefore, it is of utmost importance to relay this message to all authors.

References

Ahtiainen, A., Surakka, S. and Rahikainen, M. (2006), "Plaggie: GNU-licensed source code plagiarism detection engine for Java exercises", in Berglund, A. and Wigg-berg, M. (Eds), *Proceedings of the 6th Baltic Sea Conference on Computing Education ResearchProceedings of the 6th Baltic Sea Conference on Computing Education Research, Koli Calling*, Uppsala University, Uppsala.

Belkhouche, B., Nix, A. and Hassell, J. (2004), "Plagiarism detection in software designs", paper presented at the ACM Southeast Regional Conference.

Boneh, D. and Shaw, J. (1995), *Collusion-secure Fingerprinting for Digital Data*, Computer Science Department, Princeton University, Princeton, NJ.

Burnham, R.A. (2003), "Marketing yourself via ghostwriting", *Certification Magazine*, available at: www.certmag.com/read.php?in=344 (accessed 20 April 2011).

Carrol, J. (2002), *A Handbook for Detecting Plagiarism in Higher Education*, The Oxford Centre for Staff and Learning Development, Oxford Brookes University, Oxford.

CLEF (2011), *PAN 2011 – Uncovering Plagiarism, Authorship, and Social Software Misuse*, available at: http://pan.webis.de/ (accessed 5 May).

Cosma, G. and Joy, M. (2006), "Source-code plagiarism: a UK academic perspective", paper presented at 7th Annual Conference of the HEA Network for Information and Computer Sciences, Dublin, 29-31 August.

Cowie, R., Douglas-Cowie, E., Tsapatsoulis, N., Votsis, G., Kollias, S., Fellenz, W. and Taylor, J.G. (2001), "Emotion recognition in human-computer interaction", *Signal Processing Magazine, IEEE*, Vol. 18 No. 1, pp. 32-80.

Culwin, F., MacLeod, A. and Lancaster, T. (2001), *Source Code Plagiarism in UK HE Computing Schools, Issues, Attitudes and Tools*, JISC, London, Commissioned by JISC.

Farquad, M., Ravi, V. and Bapi, R.S. (2008), "Rule extraction using Support Vector Machine based hybrid classifier", paper presented at TENCON 2008 – 2008 IEEE Region 10 Conference, Hyderabad, 19-21 November.

Giombetti (1992), "UW's friendly corporate ghostwriter", *Eat the State*, Vol. 6 No. 19.

Google (2011), *Google Translator*, available at: http://translate.google.com/ (accessed 20 April).

Hájek, P. (2011), "Deductive systems of fuzzy logic", *Proof, Computation and Agency*, Vol. 352 No. 2, pp. 67-78.

Harp, J. and Taietz, P. (1966), "Academic integrity and social structure: a study of cheating among college students", *Social Problems*, Vol. 13 No. 4, pp. 365-73.

Harrick, C. (2011), *iThenticate Introduces New Plagiarism Detection Service for Individual Authors and Researchers*, prweb.com, available at: www.prweb.com/releases/2011/1/prweb8072290.htm (accessed 20 April).

Hawley, C.S. (1984), "The thieves of academe: plagiarism in the university system", *Improving College and University Teaching*, Vol. 32 No. 1, pp. 35-9.

Holley, R. (2009), "How good can it get? Analysing and improving OCR accuracy in large scale historic newspaper digitisation programs", *D-Lib Magazine*, Vol. 15.

IEEE (2011), *Cross Check User Guide*, IEEE, available at: www.ieee.org/documents/CrossCheck_User_Guide.pdf (accessed 20 April).

iParadigms (2011), *Turnitin.com*, available at: www.turnitin.com/ (accesseed 10 February).

ISO (2008), *Document Management – Portable Document Format, International Organiztion for Standardization*, available at: www.iso.org/iso/catalogue_detail.htm?csnumber=51502.

iThenticate.com (2011a), *CrossCheck*, iParadigms.com, available at: www.crossref.org/crosscheck/index.html (accessed 20 April).

iThenticate.com (2011b), *Plagiarism Checker – Plagiarism Detection Software*, (accessed 1 April).

Kumar, G., Tulyakov, S. and Govindaraju, V. (2010), "Combination of symmetric hash functions for secure fingerprint matching", paper presented at the Pattern Recognition (ICPR), 2010 Twentieth International Conference, 23-26 August.

Lei, L., Zhi-Ping, G. and Wen-Yan, D. (2010), "Fuzzy multi-class support vector machine based on binary tree in network intrusion detection", paper presented at the Electrical and Control Engineering (ICECE), 2010 International Conference, Wuhan, 25-27 June.

Liu, Y. and Liu, Y. (2010), "Incremental learning method of least squares support vector machine", paper presented at the Intelligent Computation Technology and Automation (ICICTA), 2010 International Conference, Changsha, 11-12 May.

McCabe, D.L. (1999), "Academic dishonesty among high school students", *Adolescence*, Vol. 34 No. 136.

Marsden, H., Carroll, M. and Neill, J.T. (2005), "Who cheats at university? A self-report study of dishonest academic behaviours in a sample of Australian university students", *Australian Journal of Psychology*, Vol. 57 No. 1, pp. 1-10.

Niezgoda, S. and Way, T.P. (2006), "SNITCH: a software tool for detecting cut and paste plagiarism", *ACM SIGCSE Bulletin*, Vol. 38 No. 1, pp. 51-5.

Parker, A. and Hamblen, J.O. (1989), "Computer algorithms for plagiarism detection", *IEEE Transactions on Education*, Vol. 32 No. 2, pp. 94-9.

PlagiarismDetect (2011), *Plagiarism Checker*, (accessed 20 April).

Prechelt, L., Malpohl, G. and Philippsen, M. (2000), "JPlag: finding plagiarisms among a set of programs", *Journal of Universal Computer Science*, Vol. 8 No. 11.

RemovePlagiarism.com (2011), *Detect Plagiarism|Remove Plagiarism*, RemovePlagiarism.com, available at: www.removeplagiarism.com/ (accessed 20 April).

Rennie, D. and Flanagin, A. (1994), "Authorship! Authorship! Guests, ghosts, grafters, and the two-sided coin", *JAMA*, Vol. 271 6, February, pp. 469-71.

Seungjin, L., Joonsoo, K., Jinwook, O., Junyoung, P. and Hoi-Jun, Y. (2010), "A 92m W 76.8GOPS vector matching processor with parallel Huffman decoder and query re-ordering buffer for real-time object recognition", paper presented at the Solid State Circuits Conference (A-SSCC), 2010 IEEE Asian, 8-10 November.

Shivakumar, N. and Garcia-Molina, H. (1995), "The SCAM approach to copy detection in digital libraries", *D-Lib Magazine*, November.

SysTranSoft.com (2011), *SYSTRAN – Online Translation Software*, www.systransoft.com/ (viewed 20 April)

Tahaghoghi, C.A. and Tahaghoghi, S.M.M. (2006), "Plagiarism Detection across Programming Languages", in Estivill-Castro, V. and Dobbie, G. (Eds), *ACS 48, Hobart*.

Tappert, C.C., Suen, C.Y. and Wakahara, T. (1990), "The state of the art in online handwriting recognition", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 12 No. 8, pp. 787-808.

turnitin.com (2010a), *Answers to Common Legal Questions about Turnitin*, available at: http://turnitin.com/static/pdf/us_Legal_Document.pdf (accessed 1 February).

turnitin.com (2010b), *Student Quick Start Guide*, available at: http://turnitin.com/resources/documentation/turnitin/training/en_us/qs_student_en_us.pdf (accessed 1 February 2011).

Vamplew, P. and Dermoudy, J. (2005), "An anti-plagiarism editor for software development courses", in Young, A. and Tolhurst, D. (Eds), *Proceedings of the Seventh Australasian Computing Education Conference (ACE2005)*, CRPIT, ACS, Newcastle.

Verco, K.L. and Wise, M.J. (1996), "Software for detecting suspected plagiarism: comparing structure and attribute-counting systems", *Proceedings of the First Australian Conference on Computer Science Education*, Citeseer.

WhiteSmoke-Inc. (2011), *English Grammar Software*, WhiteSmoke Inc., available at: www.removeplagiarism.com/ (accessed 20 April).

Wiedemeier, P.D. (2002), "Preventing plagiarism in computer literacy courses", *Journal of Computing Sciences in Colleges*, Vol. 17 No. 4, pp. 154-63.

Wise, M.J. (1996), "YAP3: improved detection of similarities in computer program and other texts", *SIGCSE '96 Proceedings of the Twenty-seventh SIGCSE Technical Symposium on Computer Science Education*, ACM, New York, NY.

Zadeh, L.A. (1965), "Fuzzy sets[*]", *Information and Control*, Vol. 8 No. 3, pp. 338-53.

About the authors

Ahmed Patel received his MSc and PhD degrees in Computer Science from Trinity College Dublin (TCD), specializing in the design, implementation and performance analysis of packet switched networks. He is a Lecturer and Consultant in ICT and Computer Science. He is a Visiting Professor at Kingston University in the UK and currently lecturing at Universiti Kebangsaan Malaysia. His research interests span topics concerning high-speed computer networking and application standards, network security, forensic computing, autonomic computing, heterogeneous distributed computer systems and distributed search engines and systems for the web. He has published well over 170 technical and scientific papers and co-authored two books on computer network security and one book on group communications, co-edited a book on distributed search systems for the internet. He is a member of the Editorial Advisory Board of the following international journals: *Computer Communications*, *Computer Standards & Interface*, *Digital Investigations*, *Cyber Criminology*, and *Forensic Computer Science*. Ahmed Patel is the corresponding author and can be contacted at: whinchat2010@gmail.com

Kaveh Bakhtiyari received his BSc degree in Software Engineering from Islamic Azad University in Iran and he is pursuing his MSc in Information Technology - Artificial Intelligence (Intelligent Systems) at Universiti Kebangsaan Malaysia in 2011. His interests span a wide range of topics in the area of distributed systems, computer security, cloud computing, and human computer interaction. He is currently undertaking his Master's thesis under the supervision of Professor Ahmed Patel at the Universiti Kebangsaan Malaysia in the topic area intelligent human emotions recognition.

Mona Taghavi received her BSc degree in Information Technology from Parand Islamic Azad University of Iran in 2007. Besides her involvement in several Iranian national ICT research projects, she has worked for an IT consulting and project managing company which was responsible for overseeing the Supreme Council of Information and Communication Technology (SCICT) of Iran programme. She was responsible for preparing some of the technical reports for this programme. Currently, she is pursuing her MSc in Information Systems at University Kebangsaan Malaysia and undertaking research in cooperation with Professor Dr Ahmed Patel in advanced secure web-based information systems and secure mobile agent-based e-marketplace systems.