Anti-plagiarism E-Tools as a technology

protecting privacy and the intellectual work

Alec Broadhurst, email: a0803623@unet.univie.ac.at, MatrikelNr: 0803623
Karolina Libront, e-mail: a0804038@unet.univie.ac.at, , MatrikelNr: 0804038
Michal Marcinkowski, e-mail: a0804194@unet.univie.ac.at, MatrikelNr: 0804194
Anna Samel, e-mail : a0806089@unet.univie.ac.at, MatrikelNr: 0806089

University Vienna - Institute of Government & European Studies

Abstract: Nowadays, with the advent of the internet there is increased accesibility to a wide range of resources from all over the world. In this internet medium access to work is limitless with no control or regulation. With this there is an increased need to protect intellectual work, with the necessity of checking the originality of academic work. The new threats to plagiarism created by the internet and the information society, there is a greater need to protect our privacy and original intellectual work.
1. **Plagiarism – definitions, concepts, legal base.**

According to the Law Encyclopedia, plagiarism is legally defined as ‘the act of appropriating the literary composition of another author, or excerpts, ideas, or passages, and passing the material off as one's own creation’.\(^1\) Plagiarism is therefore understood as a theft of another person's writings or ideas. Generally, it occurs when someone steals expressions from another author's composition and makes it appear to be his own work. Plagiarism is not a legal term; nonetheless it is often utilized in lawsuits. Courts in the majority of countries recognize acts of plagiarism as violations of copyright law, specifically as the theft of another creator's intellectual property. Because copyright law allows a variety of creative works to be registered as the property of their owners, ‘lawsuits alleging plagiarism can be based on the appropriation of any form of writing, music, and visual images’\(^2\). At its simplest and most extreme, plagiarism involves putting one's own name on someone else's work; this is commonly experienced in schools when a student submits a paper that someone else has written. Schools, colleges, and universities usually have explicit guidelines for reviewing and punishing plagiarism by students and faculty members.

With the increased ease which plagiarism can take place on the Internet we shall examine the anti-plagiarism tool technology as a deterrent against intellectual theft. We shall conduct a survey to examine student opinion on plagiarism in order to assess the degree of understanding in this area, and in order to examine the seriousness of the problem. After this we will evaluate the particular anti-plagiarism e-tools focusing on how their work and assessing their effectiveness.

2. **Questionnaire – results of the survey.**

During our research, we wanted to find out whether the commitment of plagiarism is common nowadays among students. We have prepared a questionnaire using similar questions as those used by professional scholars during survey conducted at the University of Texas at Austin

---

\(^1\) Law Encyclopedia: Plagiarism, Available from: http://www.answers.com/topic/plagiarism

in April 2006³ (sample of our questionnaire is enclosed in our paper). At the end of November and the beginning of December 2008 we sent our questionnaires via email to approx. 100 students from Poland, Great Britain and Austria in the age 19 – 25. We have collected 50 filled-in questionnaires. Because of the limited extend of the survey, only simplistic picture is given and the extensive conclusions cannot be drawn. Nonetheless, the results were useful. The general outcome is that students are aware of the growing abuse of coping intellectual work and increased incorrect use of referencing among young people. Majority of them understands how scholars define plagiarism but they seem to be convinced that professors usually are unable to detect plagiarism. Furthermore, they believe that using plagiarism detection tools by professors would restrain students from fraud. That means that common use of the plagiarism detection tools at universities is needed and could help in detection of the problem.

Below we analyze in a greater detail the results of our survey:

Are you aware that the following activities are against the regulations of your University?

![Bar chart showing the results of the survey](chart.png)

From the results of the survey, it seems that the majority of the students are aware of the plagiarism guidelines yet the difference of views between ‘copying exact sentences without
proper referencing’ and ‘paraphrasing without referencing.’ 14% more people, when surveyed believed that paraphrasing without referencing is worse than copying exact sentences. Although we only carried out a small sample the result still can give us a broad idea of opinion. With this result it seems surprising.

36% of respondents plagiarised and 62% knew of someone who has plagiarised. These results seem high, and represents a general acceptance of the stealing intellectual property.
The majority of students acknowledge the importance of their intellectual property right, and understand the immorality of this violation.

84% of students believe that plagiarism cannot always be detected by markers. This result is key in our investigation of the necessity of usage of anti-plagiarism.
62% of students are aware of the existance of any plagiarism detection tool and 64% of them believe that those can help to reduce plagiarism what seems to be quite optimistic results.
Usage of the anti-plagiarism tools by professors would considerably reduce the occurrence of the plagiarism and the general acceptance for that, as 86% of students claim it would restrain them from committing plagiarism.

According to the majority of the respondents (88%), usage of the plagiarism detection tool does not threaten their privacy.

3. **Case study - the on-line tools detecting unauthorized plagiarism.**

In the age of widespread Internet usage, committing plagiarism is easier than it is used to be even 20 years ago. Each student with access to the Internet can very easily find hundreds of relevant answers for his or her assignment in a few seconds. And even if the professor suspects plagiarism, the enormousness of sources necessary to check works in favor for the plagiarist. Very often it is said that, first of all, universities should educate students on the problem of plagiarism and that it is forbidden to copy someone’s intellectual work without proper referring. Educational approach to prevent plagiarism focuses on “*teaching students how to write and cite appropriately*”¹, but even though students still have a problem with analyzing quotations and

---

drawing conclusions of them. It is easier just to copy someone’s ideas and sign below it, trusting that nobody would ever suspect plagiarism. One can say that plagiarism is just the issue concerned with honesty between a professor and a student. But it was proved not only in our research but in many others scientific surveys that nowadays students become more and more dishonest towards their lecturers. That is why inventing tools which will help in detecting unauthorized ‘borrowing’ seems to have become necessary process in the age of the Internet.

The concern about finding the solutions which would prevent plagiarism arose in the beginning of 21st century. That was the time when many companies invented and provided electronic tools designed to identify cases of suspected plagiarism. These tools differ on the basis of several features. Some of them are licensed software, others are on-line used applications which charge users by the paper or for a specific time period. These tools often use different databases – the majority of them are search engines which match submissions against content found on the Internet. Others submissions against a proprietary database may contain other student submissions and papers from term paper mills or academic sources. But the technological solution is based on the idea to compare textual documents in the respect of their similarity to other texts stores in the system database or existing in the Internet sources.

It was scientifically proved that “the occurrence of words in English and other European languages follows a Zipfian distribution. This means that some words occur frequently (function words like “the”, “to”, “of” and so on) but most words are relatively scarce. Quite ordinary words actually have a low frequency. As an example, in the Brown corpus of 1 million words, taken from representative samples of everyday texts, 40% of the word forms occur only once (Kupiec, 1992). Now, this distinctive distribution will be more pronounced for word bigrams (two consecutive words), and even more marked for trigrams (three consecutive words). (...) Any article will on average have 77% of its trigrams unique. Even in a very large corpus in a limited domain most trigrams in independently written articles do not match those in other articles.”. In practice, anti-plagiarism tools combine a number of triples or longer word strings and match them to the identical “clones” in other papers occurred in the database. Of course, there is still a small

---

5 Compare with: http://www.strikeplagiarism.com/
possibility that the software will not detect plagiarism but simply match similar “clones” which seem to be identical but when used in other contexts are quite different. That’s why anti-plagiarism tools are not perfect and still need human verification.

In our work we would like to focus on the one of the most popular among academic institutions tool detecting plagiarism called iThenticate. The American company – iParadigms, which created iThenticate (and its twin-tool called Turnitin) was set up in 1996 by the group of the researchers from the University of the California at Berkeley. The iThenticate first was created in 1998 as a Plagiarism.org – “the world’s first internet-based plagiarism detection service”7. In the next six years that tool evolved and developed and now it is recognized worldwide as a iThenticate (because of the increased demand for the anti-plagiarism e-tool, iParadigms created also a twin-sister for iThenticate called Turnitin. Although we will use the name iThenticate for both of the tools as both of them use the same technology solutions.). This on-line tool “is able to clearly identify matching texts between two text based documents of any language or size. The search engine examines a chunk of text, eliminates words that are too common, and turns the other words into numbers. It also converts Internet content into numbers. Consequently, it can compare patterns of numbers from the sample with patterns of numbers from the Internet. To do this, the search engine uses a number of complex mathematical algorithms.”8.

Everyone (professors, students, enterprises, etc.) can become a registered user of the company’s service, although “most users seem to be high school teachers”9. The registered user, after identification procedure (including submitting his name, institution, department, course) and paying the fee, can submit on-line any paper (in .txt form) which then is tested for plagiarism among the database which consists of over 10 billion Web Pages, over 70 million student papers (all papers submitted by registered users are also archived to the database), over 10,000 major newspapers, magazines and scholarly journals and thousands of books including literary classics. After submitting the paper for detecting plagiarism, a registered user receives the report indicating the probability (in percentage) of whether the paper was plagiarized. “Instructors can

---

7 Compare with: http://www.ithenticate.com/company.html
Click on links to direct them to the source of the possibly plagiarized material. This process takes twenty-four hours on average but can take up to two days, depending on the length of the text and the level of demand. The number of faculty users - more than 450,000 - is impressive. The institutional licensees were purchased by over 6,500 institutions in over 100 countries. The iThenticate seems to be the most popular and most trusted solution for originality checkings.

In addition to online resources, iThenticate compares documents to a publications database comprised of over 10,000 major newspapers, magazines, and journals. Unfortunately, it cannot be downloaded – it is a purely web-based application. Users can access their accounts from any Internet-ready computer by logging in. The company maintains that it has proved extremely advantageous for customers who have multiple locations, various dispersed departments working on projects, or employees working remotely.

At the same time, the users may feel safe. The tool does not include submitted documents in databases shared with other users. Publishers, authors, research facilities, government organizations, legal firms, financial institutions and more submit hundreds of documents to iThenticate each day. The nature of this content is often sensitive, and great emphasis is placed on security and discretion. At no time are submitted documents cached in a database that is shared with other licensed iThenticate users. It makes the website a safe tool.

iThenticate is also integrated with other modern technologies. The company’s development team has made available an Application Programming Interface (API). The API allows organizations to implement the iThenticate solution and maintain workflow efficiencies. Future development plans include deeper integrations with more commonly used Manuscript Tracking Systems and Content Management Systems.

Users view iThenticate as a useful tool. John M. Barrie in his work clearly states that it works against academical plagiarism. He also stated that it can contribute to improvements in

---

10 Ibidem.
11 iThenticate FAQ, Available from: http://ithenticate.com/
practices of publishing. He sees iThenticate as a tool for pre-publishing due diligence and as a tool to detect post-publishing intellectual property theft.

To portray the workings of iThenticate, we are adding an annex to our paper with examples of how the tool works.


Nowadays, in the era of the Internet, the problem of protecting intellectual property is more grievous than ever before. As our research clearly showed, plagiarism has become a common occurrence in academic field. Students’ assignments, articles or even whole passages in books are often copied without authors’ consent or knowledge. What is more, our research proved that a lot of people seem not to realize what plagiarism exactly is or how they can protect themselves against it.

Nonetheless, such protection is possible. Online anti-plagiarism tools are an effective weapon in the fight against theft of intellectual property.

They are easily accessible through the Internet and not hard to use. It is our conclusion that such systems should be popularized so that as many people as possible know about them and are able to protect their creations.
Annexes:

1. Sample of the questionnaire used during the research.

2. Detection of sentence or paragraph addition by iThenticate.

Bibliography:


