

Plagiarism Detection Test 2010

Katrin Köhler and Debora Weber-Wulff

Plagiarism in academia is not a new phenomenon. Despite all the concern that plagiarism is made effortless or even possible by the Internet, plagiarism and its siblings – ghostwriting, falsifying data, and other aspects of academic misconduct – have been hot topics of discussion for centuries.

Many schools and universities are in panic at the rising tide of plagiarism that they are seeing. From copies of Wikipedia articles or other Internet sources handed in as term papers, to ones purchased from paper mills and handed in as own work – teachers and administrators are finding more and more instances of such behavior. Since there are also many changes being demanded of the institutions and more and more learners needing taught with less and less resources, many wish for a magic bullet, a software that can quickly and effortlessly determine which papers are plagiarisms, so that punishment can be meted out and only the original papers need be scrutinized.

We have been testing so-called plagiarism detection systems since 2004. We now have a large collection of test cases, short essays both original and plagiarized, that we use to test how well the systems do at finding plagiarism. The systems are scored for each test case on a scale of 0-3 on how well they find the plagiarisms, and for not declaring the originals to be plagiarisms. Then as now, the results are disappointing. Many systems can find only exact copies at best, and even the best systems are only satisfactory. Over the years we have expanded our focus to include the usability of systems, and the current test also looks at the professionalism of the companies behind the systems.

Test Overview

In the spring of 2010 we set out to test the current crop of systems. The following table gives an overview of the number of systems available, the number of system tests we could complete, and an overview of the test cases.

Test	Systems evaluated	Completed tests	Test cases	Test Case Language	Assessment
2004	12	8	10	German	Binary decision: plagiarism found or not

2007	25	17	20	German	Plagiarism detection graded 0-3 code comparison
2008	27	19	31	German	Plagiarism detection graded 0-3 collusion, usability
2010	47	26	42	German, English, Japanese	Plagiarism detection graded 0-3 usability and professionalism Japanese and English tests no collusion, no code comparison

Due to the large number of systems, collusion detection and code comparison systems have had to be delayed until a later time in order to be able to present these findings. We included two special tests, one for the ability to deal with Japanese encoding¹ and one on request of one of the companies who has long suspected that other companies are reselling their product. They put an original essay in their database that is not on the Internet, and stored as the source a fake URL. Indeed, one of the systems reported that this test case was a 100% plagiarism of the URL in question. Legal proceedings have been initiated.

One of the systems tested in 2010 is ProfNet, a plagiarism detection service. We were requested to submit proper student papers, not test cases. We scanned in 5 of the papers from the 2001 plagiarism detection incident that sparked Weber-Wulff's own inquiries into plagiarism detection, used character recognition on the text, cleaned up the recognition to fit the original works, and submitted these to be hand-checked by the plagiarism detection service.

The other 25 systems that were tested in 2010 were then graded on the **effectiveness** of finding plagiarism, the **usability** of the system, and the **professionalism** of the company behind the software. It turned out that through an unlucky circumstance – trying to preserve the test results of 2004 had moved the directory of the test cases into the line of sight of the search machines – all of these first 10 test cases have been rigorously indexed by all search machines. Some of the plagiarism detection systems permitted us to exclude possible sources, if this was at all possible, we did so. But systems that could not exclude a site at all, or only reported on the top source, were thus marked down in the **full set of test cases**.

Evaluation Metrics

In order to evaluate the systems without this problem we also calculated the effectiveness on just the **cases 10-40** (the Japanese one was always done separately) as well as the effectiveness on just the new, **English-language cases 31-40**, as some systems have asserted that they are much better on English than on German, which has special characters called umlauts. We then grouped the software into grade levels (less than 50%, 50-60%, 60-70%, 70-

¹Many thanks to Vendula Knopfová, Tokyo, for preparing the Japanese-Language test case.

80%, 80-90%, 90-100% of the possible points).

The grouping of the results for effectiveness was very difficult to do for the total metric, as there was a continuum from 55-64% with no percentage point omitted. It was slightly easier to group the systems for the cases 10-40, with Urkund, Plagiarisma, Turnitin, and PlagAware obtaining between 60 and 66% of the points for effectiveness. iPlagiarismCheck is also ranked here, but since this system just repackages the results of Turnitin, it should not be considered at all. It did not do well in the full set effectiveness metric because it was unable to exclude sources for the test cases 0-9.

The English language test cases had only two systems, Copyscape and PlagAware, score 70% of the points, Urkund managed 66% and PlagScan, Plagiarisma, Ephorus, Turnitin, StrikePlagiarism and Viper tied for fourth place with 60% of the possible points.

The **usability metric** took into account aspects such as design, language consistency and professionalism, navigation and descriptiveness of the labelling, print quality of the reports, and how well the system fits into the workflow of a university. We also tested the support offered by sending a support question from an email address that was not identifiable as us (we often receive answers within minutes when we write with our real names) and checking if it was answered within 48 hours, as well as how completely it was answered. Top honors with just 25 out of 30 possible points was PlagScan. Only four other systems were able to score more than 20 points in this metric: PlagiarismFinder, Ephorus, PlagAware and Turnitin.

The **professionalism metric** was a new development for this test series. A university considering using a plagiarism detection system will want to have a professional partner available, even if this means paying a bit more for the service. The professionalism metric included giving a real street address, a telephone number, and the name of a real person on the home page; registering the domain in the name of the company and not through a reseller; not also running paper mills or advertising for paper mills or other ethically questionable services on the web site; answering the phone during normal business hours for the country given in the street address and being able to speak German; and not installing malware on the computer under the guise of installing the detection software. This metric had one system, Plagiarism Finder, reaching a full 15 points; PlagAware and StrikePlagiarism only missed one point; five other systems missed only two points (Turnitin, Ephorus, Docoloc, PlagScan, and Blackboard). Seven other systems were able to obtain half or more of the points assigned (Copyscape, Un.Co.Ver, Genuine Text, Compilation, Urkund, Plagium and ThePlagiarismChecker). All other systems should not be considered for university use.

This gave us five different metrics for evaluating the software. No system was in the top group for all metrics. The top value for effectiveness was 70% for the English-language test cases, 66% for the 10-40, 64% for the complete group, 83% for usability and 100% for professionalism. We decided to use a composite metric that did not work with absolute percentages, but with the relative rankings. We ranked the 25 systems in each of the 5 categories and put together

an average ranking. We then ranked by average ranking, and can group the systems now into three groups on the basis of their usefulness for finding plagiarizing students at a university: partially useful, barely useful, useless. This is because even though systems are “top ranked” in our tests, they still do not reliably find plagiarism, or only partially find plagiarism, or are rather difficult to use. There are systems, however, that can be useful for other tasks, such as looking for plagiarisms of a web page, even if they are not very effective. This test only looks at the usefulness of systems for university use.

System	Rank for all tests	Rank for Tests 10-40	Rank for Tests 31-40	Usability Rank	Professionality Rank	Aver. Rank	Rank	Effectiveness grade
Partially useful								
PlagAware	4	6	1	4	2	3,4	1	C-
Turnitin	3	3	5	5	4	4	2	C-
Ephorus	4	9	5	2	4	4,8	3	C-
PlagScan	8	8	5	1	4	5,2	4	C-
Urkund	2	1	3	13	13	6,4	5	C-
Barely useful								
Plagiarism Finder	11	12	11	2	1	7,4	6	D+
Docoloc	9	9	12	6	4	8	7	D+
Copyscape Premium	12	12	1	7	9	9,2	8	D+
Blackboard/ SafeAssign	6	9	12	19	4	10	9	C-
Plagiarisma	1	3	5	23	22	10,8	10	C-
Compilatio	6	7	21	9	12	11	11	C-
StrikePlagiarism	15	14	5	22	2	11,6	12	D
The Plagiarism Checker Free	12	14	15	7	14	12,4	13	D+

The Plagiarism Checker Premium	14	14	15	7	14	12,8	14	D+
Useless								
iPlagiarismCheck	17	5	15	19	16	14,4	15	F
Plagiarism Detector	17	19	15	23	1	15	16	F
UN.CO.VER	16	18	15	16	10	15	16	F
GenuineText	19	21	12	16	11	15,8	18	F
Catch It First	22	17	11	15	20	17	19	F
plagium	25	25	15	10	14	17,8	20	F
Viper	27	25	5	12	22	18,2	21	F
PlagiarismSearch	20	21	21	13	18	18,6	22	F
PlagiarismChecker	21	19	25	19	16	20	23	F
Grammarly	24	23	23	11	22	20,6	24	F
PercentDupe	22	24	24	16	19	21	25	F
Article Checker	25	27	27	25	25	25,8	26	F

Partially useful systems

The top-ranked system in the 2010 test is **PlagAware**. It tied for first place with Copyscape on the new test cases with 70%, however still only detecting 61,11% of the plagiarism cases 10-40. PlagAware is a German system that produces excellent documentation of the plagiarism found, highlighting the commonalities in a side-by-side presentation. However, its usefulness at university is limited, as each file must be uploaded individually - no ZIP file or student-submission is possible. The system was not designed to be used in a university setting, but rather to find plagiarisms of online texts, which is important for sites trying to optimize their search machine ranking, as plagiarism will contribute to downranking. The system has improved for university use since the last test, as previously only texts could be checked that were available online and had a logo linking back to the PlagAware site. We hope that this system might eventually develop its system to the point of being useful for educational purposes.

Coming in second is the well-known US system **Turnitin**, sold by the company iParadigms. The system has been plagued with problems in the past including inability to deal with umlauts, ignoring the Wikipedia, and having a relatively complex setup. They have managed to partially sort out many of these problems, now doing better on German texts than on English ones, getting the best results for material that they already have stored in their databases. This is a problem from a European standpoint, as it is not legal to store copies of material without the express permission of the copyright owners. Also, many theses in technical fields contain material which is under a non-disclosure agreement and must not in any way be stored in databases.

We analyzed why Turnitin has moved up from overall effectiveness place 13 in the 2008 test to place 3 in this test – the reason is that the other systems have gotten worse. Test cases such as number 18 (a highly-edited version of a paper available online) is now consistently not found by many systems, as well as a plagiarism from a blog (number 29). For the test cases that include plagiarism from multiple sources, many of the software systems were not able to find all of the sources, causing them to lose points on effectiveness.

Turnitin now has a German-language version that still has a few kinks in it, translation-wise (*Benotungsbuch* for gradebook, for example). There are still problems with Turnitin flagging spam sites that have scraped the content of the Wikipedia. Some of these sites are not safe for work, as their business market is selling pornography. This might cause a problem when used on school computers. Interestingly enough, the Wikipedia itself was often listed just after these sites. There was also a problem with the database flagging plagiarism from pages that no longer exist. Often this was, indeed, the correct page, but for a teacher it is frustrating to find evidence of plagiarism, but not to be able to find the source, which would be necessary for sanctioning the student.

Third in the overall ranking is **Ephorus**, a Dutch system that took first place in 2007, dropping to 7th place in 2008. The system has been completely redone and now ranks second in usability, up from 8th place. It is now possible for students to hand in their papers directly to the system using a hand-in code, similar to what Turnitin offers. The system still has many navigational areas, but it is not as confusing as it used to be. A nice touch is that the teacher can set a threshold for amount of plagiarism detected and have the system send an email when a report is detected to be above this value. Umlauts in file names are problematic – the system reports no plagiarism, although it does find the plagiarism in an identical file with a name that does not have an umlaut in it. It also skips some words with umlauts in them, which is problematic for German texts. We could not determine why some could be found and some not.

Highly problematic is that Ephorus still stores all of the documents in its database. There are a number of sub-databases called “pools” that are available, and schools can apply to have their papers only checked against a sub-pool, for example, all schools in one city or one country, or just restricted to the university. But the same problem exists here as for Turnitin – the author (in this case the student) must give permission for the paper to be stored in the database, even if it

is not passed on to others, according to what Ephorus states on its site.

Fourth in the group of partially useful systems is **PlagScan**, a German system. This system used to be called PlagiatCheck and came in 10th in the 2008 test. One purchases “Plag Points” (PP), a test costs 1 PP per 100 words tested. The administrator sets up users and assigns them points for use. The system had trouble with umlauts, that is to have been corrected since we conducted the tests. There are three kinds of reports, a list of possible sources with links to click on, the submitted document with the suspicious areas linked to a possible source, and a docx file with the sources in comments. Sorely missing is a side-by-side presentation that is necessary for going forward with sanctions. Despite all its problems, PlagScan was first place in usability, but only 8th place in overall effectiveness with only 60% of the points awarded for finding plagiarisms.

The last system in the partially useful systems is the new version of the Swedish **Urkund**. It ranked quite high in effectiveness (second place overall, third for the new cases), but the system is quite hard to use. Even though they now have a German version, the translation is quite bad and the system often reverts to English and then Swedish. The system has been redesigned since 2008, rather for the worse. The navigation is confusing, the layout at times catastrophic with texts overlapping fields, the printed reports could be better, the error messages are cryptic, and the link descriptions are unclear.

Extremely problematic was that our documents from 2008 were still in the database, although we sent an email in August 2008 requesting that they be removed. Even worse, an administrator from the University of Frankfurt had tested the system and used one of our test cases. We were informed that one of our test cases belonged to them! They reported that they had discovered that they were unable to remove papers from the database. This was the reason that they decided not to use Urkund, as they do not want the papers of their students stored or made available to other schools.

In summary, **none of these systems can be recommended for general use**, as the effectiveness is not good and each has some problematic area. They can only be used in situations where a first impression of possible plagiarism is needed after a teacher becomes suspicious. In such situations, however, searching with Google can be faster and more effective.

Barely Useful Systems

The next section will discuss the systems that can be considered to be barely useful – that is, they find some plagiarisms, but either miss a lot, have bad scores on usability or professionalism, or other problems.

- **Plagiarism Finder** is a German system that was one of the best systems tested in 2004, but has been plagued with difficulties since then, most specifically with the system not working stably and returning widely varying results just minutes later on a re-test. This

is a system that is installed locally on a computer, or which can be purchased on a USB stick. There are many details about the system that makes it fit nicely into the workflow of a teacher, it was tied for second place for usability. But even for the new, English-language test cases, it only placed 11th.

- **Docoloc:** This system was fifth in the 2008 test, seventh in this test and only twelfth for the English test cases. A major problem was flagging plagiarisms that could not be found in the possible sources reported! It did manage to find the sources that we expected just about half the time, but for a teacher who does not know what the source is, this is highly irritating to not be able to determine why certain sources are being given. The reports themselves are extremely hard to interpret. There is no side-by-side presentation, you have to search yourself in the links given. The markup is misleading, as the highlighted areas are not the plagiarisms found, but the samples tested. The navigation is unclear. We also experienced server problems during the test, but the company received a high professionalism rating as our (anonymous) questions were promptly answered by email.
- **Copyscape Premium:** This system took first place in effectiveness in 2008, but is only on place 7 this time. It was, however, best at finding the new, English-language test cases. It often reported no plagiarism for test cases that it found completely on the last test, something that we observed with many other systems. This is perhaps due to changes in the Google API, or to certain search engine “optimizations” that many companies attempt and which might taint the results. We did not see a method for printing of the report, beyond printing the results page. You have to click on the heading for a result in order to see the page marked up with matching text. You can switch between the markup on the original and the found source, but there is no side-by-side report. They now offer a daily or weekly check for plagiarism of a page, something that is useful for people needing to protect their online intellectual property, but not for university use. There is at least now the possibility of batch uploading of files.
- **Blackboard / SafeAssign:** We were finally able to test SafeAssign directly and not by using a system belonging to another school. However, we were only able to do this after writing an email to the CEO of the company! We repeatedly filled out the online form asking someone to contact us, but we received no answer. We even spent about an hour on the phone, calling numbers in the USA, UK, and Belgium, trying to reach someone in their call centers who could give us a name or number of someone to contact about access. When we found the web page encouraging us to write to the CEO if we were unhappy, we did. It took a week for someone to answer us, but then we were able to indeed test the system. There were many problems with the layout, the links, the labels, the error messages were at times unintelligible (“An unexpected error as occurred. For reference, the Error-ID is 266x800....”), the navigation was confusing. Over all of our tests cases, SafeAssign was in sixth place (tied with Compilatio) and would have been ranked with the partially useful systems, if the usability had not been so bad. Strangely enough, although it is a US system, it fared worse on the English-language tests than on the German ones.
- **Plagiarisma:** This newcomer system from Russia actually took first place in the overall effectiveness rating, obtaining just over 65% of the plagiarism points. The system is free,

but has a daily limit of 8 tests that can be made. There were many problems with the system, including umlaut problems, reverting to Russian, no PDF upload on a Mac, you have to solve difficult captchas in order to log in, and the system looks quite different in different browsers. The professionalism index, however, shows many problems with using this system. On the same page an automatic synonymizer and a similarity checker is offered, these could help a potential plagiarist to “fine-tune” a paper. The owner of the site also runs a pornography site and spam pages, according to a whois lookup. There is no street address listed on the site, no person responsible for the page. There is a contact page and there is an answer within 2 days, but the email does not actually answer the question asked. We feel that a university cannot work together with unprofessional companies, and thus have put this system in the “barely useful” category.

- **Compilatio** is another new system in our test. This French system conducted an interesting [study](#) in 2009 on the prevalence of plagiarism, investigating 235 student papers submitted to the study by their teachers using their system. They report finding 30% of the papers reporting plagiarism, however we are concerned that they did not hand-evaluate the results. We found the system reacting too strongly on very small sequences of characters, for example reporting 1% “plagiarism” for the phrase “Stieg Larsson was born in 1954 in”, and then adding up these 1% hits for a total plagiarism quotient of 11% – this for a document that is 100% original, one that we wrote ourselves. For our older German-language tests the system placed 6th (tied with SafeAssign) with an overall effectiveness of 60%, but for the newer, English-language test cases it was only in 21st place. Although the structure of the system fits nicely into a teacher’s workflow, the annoying tendency to speak French (and to send French emails, even if we had English set as the language) led to a lower usability grade. It would have been ranked with the partially useful systems, if the usability and professionalism index (no name of a contact person given on the web site, many grammar errors on the web site, phone not answered during working hours) had not been so low.
- **StrikePlagiarism** is a Polish system that we have tested twice before. They had improved their overall effectiveness in the 2008 test, but have now dropped to just 52%. They had similar problems to Copyscape, not flagging plagiarisms they found in 2008, and reporting plagiarism in original papers that they had not previously reported. On the English-language new test cases they did, however, reach 5th place. The usability was, however, again problematic. The page loads extremely slowly; we could not find a way to upload a PDF; a title and an author had to be entered by hand for every paper uploaded; there were pages that had English, German, and Polish on them, the language would change at times without notice; horizontal scrolling was necessary; we still do not understand what the difference in the coefficients is; once it finds a source (Wikipedia!) with 7 words matching, it does not seem to try and extend the matching, thus reporting only a small amount of plagiarism in a 100% copy of the Wikipedia. We were, however, able to reach someone who was able to speak English with us on the phone.
- **The Plagiarism Checker** (Free / Premium) is another new system in our tests. It is offered by the University of Maryland. We tested both the free and the premium version

(8\$/month), strangely enough, the free version was slightly better than the premium version. It does not quantify the plagiarism, but notes “possible plagiarism”. When you click on the link, you are given the Google or the Yahoo page with the search results (we used the default, Google). There is no side-by-side reporting done, so it is really not clear what this system does, beside selecting the sample. We were not successful in contacting anyone responsible for the system, our emails were consistently returned with “temporarily local problem please try later”. The overall effectiveness was 56%, resulting in a D grade.

Useless Systems

The following systems have been deemed useless for university plagiarism detection, for various reasons.

- **iPlagiarismCheck** (also called checkforplagiarism.net) had been determined in the [2007 test](#) to have results that bore a marked similarity to the results of Turnitin. The company insisted, however, that these just were the results that any system would give on the plagiarisms – although none of the other systems came even close to having an identical order and plagiarism amount to the turnitin results. 2008 we were not able to obtain a free test account, and thus paid for only 5 documents. These results were inconclusive. We decided this year to invest in a full 30-day license, and Turnitin put a honey pot paper into their database. iPlagiarismCheck – and only iPlagiarismCheck reported that this was a 100% plagiarism and gave as the source the non-existent link stored in the Turnitin database. This proves that there is no system here, just the crooked business of making money using other company’s software.
- **Plagiarism Detector** installs 32 MB of data locally and inserts a button in Microsoft Word for starting the test on the text opened. This system reached second place in the 2008 test with a version that they sent to us on a CD. However, a number of correspondents noted that the download version attempted to install a trojan virus, so we removed it from our test. We find it irritating that they are advertising with our school logo, although we have asked them to remove this. The system crashed numerous times, and overwrites the result report each time it gets called. Word became more and more instable, also crashing on occasion. The company was contacted and updated the system, now it hung instead of crashing. The reports could not be see inside the system, but had to be selected and viewed from the explorer. Since the system did not reach even 50% of the effectiveness point overall, it has been ranked useless.
- **UN.CO.VER**, Unique Content Verifier is offered by a German marketing company that also offers ghostwriting services. The system is intended for their freelance writers to test their systems before submitting their marketing texts. The freeware system installs 3.5 MB in 10 files. The reports are utterly useless, reacting to 3-4 words in sequence, followed by another 3-4 words in sequence: “has to be ... in case the ... in the early ... to cover the ... weather the heat gets ... in hot weather the ... is needed to ...” reports a 4% plagiarism – on a complete plagiarism! Interestingly enough, the software used

iText as its basis, which is under the LGPL license. This means that UN.CO.V.E.R should also be under the LGPL license, which it is not. It is freeware, but the source code is not available from the web site, apparently one can write to the company and obtain a copy. Even awarding generous points for listing the correct sources, even if the reports are extremely problematic, this system was not able to get even 50% of the effectiveness points.

- **GenuineText** is a Swedish system that supposedly offers an English-language version, but which slips back into Swedish annoyingly often. A nice touch is that the teacher can set the status for each report to approved, doubtful, or not approved. The user interface is in need of a thorough testing, we had trouble with windows freezing, information not being visible, buttons missing on a Mac in Firefox that are present in Safari. We were not able to complete the tests at first try, the company said that they had suffered a denial of service attack and asked us to wait a few days. There was supposedly a new version then available, but it would not let us log in. Another new version showed up and had fixed many of the design problems, but the reports were reported as being in progress. That kept up all day, four days later the reports were all finished – and not one reported any plagiarism! We retried a number of tests that the very first test had found sources for, but the new version reported no problems with any of these. On the basis of the tests completed the first time, GenuineText only found 46% of the plagiarisms, which was not sufficient for more than useless. Finding no plagiarism at all is much, much worse. The system may perhaps still be in development, but during the 17 days that we were able to work with the system it was not usable.
- **Catch It First** always returned “100% original” in the 2007 test, and did not prepare reports in the 2008 test. This time we actually had reports that could be evaluated. The first 10 test cases always listed us as the source, there was no way to eliminate one of the given sources. Overall the system only reached 35% of the possible points. Attempts to contact the company were not successful, we only had an automatic answer to our email. There is no street address or telephone number given. The reports were unusable, as they could not be printed in Firefox, and just underlined the supposed plagiarism, giving a collection of links at the top.
- **plagium** is a new system that is still in beta testing. It uses Yahoo as its search engine. It takes a long time to check, our first attempt crashed after 15 minutes of reporting 98% finished. Numerous other tests crashed as well, but we were able to get all test cases finished in about 3 hours. Only one test can be done at a time, and the reports just give some text with a link and a “rank” with a percentage. There is a “Text Usage Timeline” offered, which has red balls of different diameters plotted on a grid of months vs. time, which makes no sense to us as teachers, but apparently can be useful for people tracking plagiarism of their sites. On the older tests it would often report a partial plagiarism with a page that was a complete copy, on the newer tests only results were found when at least one of the sources was the Wikipedia, otherwise nothing was found, although this may be just a chance result. Overall, only a dismal 26% of the effectiveness points on all test cases were obtained.
- **Viper** installs a 1.5 MB client on a PC system. The hissing red snake logo is not very

pleasant to look at. The system cannot deal with umlauts and produces a complicated online report and an unintelligible printed one. Writing to the company at the address given when we called on the phone bounces, the address is strangely the same as for a paper mill / essay writing service. We checked the street address – it is the same for both companies – and the telephone numbers, that only differ in the last digit. Reading through the “Terms and condition” we find that by submitting a paper to Viper we give the company “All Answers Limited” the right to keep a copy of the paper and may use it for marketing purposes, either for Viper or for any “associated website”. This proves what many have often suspected: some supposedly plagiarism detection services are just harvesters for paper mills. Even if this were not the case – Viper has the distinction of being the worst system for detecting plagiarism, coming in last with just 24% of the effectiveness points on all test cases.

- **PlagiarismSearch** answered their emails promptly, but would not give us a free account to test. We decided to purchase an “advanced academic staff” license and were able to start immediately after the PayPal purchase went through. Only one paper could be tested at a time, and then only using copy and paste into a text box. The reports are incomprehensible, reporting matches with text that not only is not identical, but not even in the same language! It did find some sources, however, so it is not in very last place. The terms of service here, too, give themselves rights to do whatever they want to with the text submitted.
- **Grammarly** permits a 7 day free trial, but you have to purchase it on credit card and cancel before the 7 days are up. The site is focused on writing and English grammar checking and proofreading, and includes a plagiarism testing component. There is only a window for pasting in text, and only one test can be done at a time. It found about 30% of the plagiarism, but the reports could not be stored. We wrote an email to complain, and received a reply that they are working on it – we never heard back from them.
- **PercentDupe** is a system registered by a Mexican in Mexicali, but the address given is in New York. Google, however, does not know of any address such as is given on the web page. The system is free, but only 15 tests per IP-Address without signing up. There were some bizarre results, for example, the source for a previous test was given as a source for a later one. Umlauts confuse the system, and the numbers reported are unclear. There was no answer to our inquiry on the contact page, no telephone number given. The telephone number given in the domain name registration database is the private number of a woman in New York. The reports are difficult to decipher, and the results are consistently poor. In addition, the terms of service grant the company the right to use your text in any way they please. “Dupe” does seem to be an appropriate name for this service.
- **PlagiarismChecker** is a system that seems to have been produced in 2006. It has a box for pasting in text, or a URL can be given. It just looks up phrases in Google and redirects you to the Google results page, so you can check it yourself, there are no reports prepared. It will truncate sentences at 32 words, which appears to be the limit for the Google API. It runs into an occasional PHP error, or an internal server error, but we were able to complete the tests, although it only managed to achieve 42% of the overall

effectiveness points.

- **Article Checker** consistently reported 0 hits with Google, although following the link there was, indeed, something to be found. You can select Yahoo or Google as the search machine to be used, and post some text or give a URL. There ads on the page for other dubious plagiarism detection systems. We debated breaking off the test, but worked through all of the test cases – this system comes in last, no matter how you look at the results.

ProfNet

Uwe Kamenz, professor for business studies at the FH Dortmund offers a plagiarism detection service through his Institute for Internet Marketing. We had wanted to test this service in 2008, but we were denied access. We requested access again for 2010, and Kamenz permitted us to submit 5 test cases, although the requirement was that they be real student papers, not our test cases.

In 2001 Weber-Wulff began working on plagiarism after a class of 32 students submitted papers, of which 12 turned out to be plagiarisms. All of the papers have been kept, so we chose 4 papers that were known to be plagiarisms and 1 paper that had been suspected of being a plagiarism, but for which there was no source found 2001.

We scanned in the papers and ran a character recognition over the pdfs. We replaced the names of the students with fictitious names, and set up freemail accounts in these names, as we had to put an email address on the submission form of about 20 fields that has to be filled out before the paper could be submitted online. We wanted to see if the students are informed that their papers are being tested. They are not.

We submitted the following papers:

1. A paper with a few unsourced quotes from a book that are properly quoted elsewhere, and which uses an English word (“inculcate”) that even many native speakers do not know.
2. A paper for which two pages had been found to be plagiarisms from two sources in 2001
3. A paper that had been determined to be a complete plagiarism 201
4. A paper that used long passages from a book that Weber-Wulff had recognized and found without search machine help
5. A paper that was suspicious because of the language used and the extensive but old literature list

We submitted the papers on September 4, 2010. The reports are marked as tested on Sept. 8, and the report produced Sept. 16. That means that it took two weeks for just these five papers to be tested, an unfeasibly long time for general university use. We then did a thorough analysis

of the reports sent to us.

The reports look professional, with many tables, numbers, and a glossary, but it is often unclear what exactly the numbers mean. On closer inspection the reports are overly long, using a half a page for each phrase found, and the phrases could often be combined. The results for the five papers will be discussed in the following section.

- **Report 1** gives a 5% probability of the entire text being a plagiarism. It is not clear why this would be an interesting number, as even just lifting a paragraph without attribution is a clear case of plagiarism. The numbers in the tables are “-15% for the subject area for text analysis” and “-80% for the subject area for text comparison”. We have no idea how to interpret these numbers.

There are three possible plagiarisms reported. One is a 60-word excerpt from a book that is not quoted in the paper and properly quoted in the source listed, and which includes the word “inculcate”. This is correctly marked “100 % plagiarism possibility”. The second is a list of 9 words and is listed as “50% plagiarism possibility”, although there is a larger portion lifted, but words were dropped from the source, so the phrase identified is just between two words that were dropped. The third is a 19-word phrase also given as “50% plagiarism possibility” that is taken from a book and properly quoted in the source given by ProfNet. The student paper had taken a 130 word paragraph from the book and had dropped or replaced words in every sentence, this was just the longest unchanged passage.

- **Report 2** gives a 52% probability of the entire text being a plagiarism, and again has tables with numbers that are not understandable, “+ 30%” and “87%” for the subject area. This time texts were reported as being from the test candidate that were not exact copies of the candidate: “andpractices”, “itcontributes”, “thewhole”, etc. are reported as being in the candidate, with “and practices” (with a blank), “it contributes”, “the whole” in the source. The candidate did, however, include a blank as well, so it appears that the reports are not generated but created by hand, perhaps by copying from pdfs. 30 possible plagiarisms are reported, although some are multiple reports of the same text. This is not easy to see in the report. The smallest reported plagiarism is 8 words. Larger amounts are also reported, but break when a word is missing or added, when we missed an error during character recognition, or at page breaks.

All of the portions are reported as being “100% possible plagiarisms”, it would be better to know how many words have been copied and perhaps what percent of the total document this is. The URLs for the sources are not always readable, long URLs just end in “...”, one has to google the text in order to find the given source.

One of the sources found, however, was golden: The CIA World Fact Book was the basis for many other plagiarisms on the Internet that were listed as sources in this

report, and it turned out that the entire paper was a plagiarism of the fact book, and not just 2 pages.

- **Report 3** gives a 70% plagiarism possibility overall. It was known that this paper was almost completely taken from an online source, we spent time measuring the exact amount of the plagiarism. The source, which was listed as a source for 29 of the 31 possible plagiarisms listed, was actually the basis for 82% of the paper, based on word count. Again, any time the student changed or dropped a word, or there was a page break, the ProfNet report stops the reporting on this phrase and resumes after the changed or dropped word.
- **Report 4** gives a 55% plagiarism possibility overall, with -24% and -25% in the subject area. Twelve possible plagiarisms are reported, most are from an excerpt from the book that is published online, and again the report restarts at word changes. Four of the possible plagiarisms are not reported as 100% plagiarisms, but as 50, 60 (twice), or 80%. They consist of only 14, 13, 16, and 21 words. The one reported as 50% similar actually has only one word different (“terms” instead of “reference”) and one of the 60% includes a “.”, but the phrases are otherwise identical. This makes the numbers even more confusing.
- **Report 5**, the one that was suspected of being a plagiarism but no sources found by hand, gives a 6% overall plagiarism possibility. Three passages are reported that are, indeed exact copies and not sourced in the paper. Each of the passages is from a different online source. Attempting to find one of the sources in preparing this report led to another online source that gives its source properly: Microsoft Encarta 1999. And indeed, all of the references in the paper are older than 1999, so it must be assumed that major portions of this paper were lifted from the Encarta, which is unfortunately not online. The paper was 26 pages long, three found paragraphs would not have led to a failing grade, but only to a lowered grade.

In four of the five cases, the search by hand using Google would have been sufficient to find enough plagiarism to fail the student, and would have been much faster. For the fifth case, the system did turn up a minor bit of plagiarism that was not found by hand.

Japanese

We were asked to test whether any of the systems were able to deal with Japanese texts. There are two encodings for Japanese, JIS-Shift and UTF-8. We had 4 different texts, the first 3 were available in both encodings:

1. A complete plagiarism from the Japanese Wikipedia,
2. The first paragraph and the last were original, the rest copied from the Japanese Wikipedia,

3. Words were replaced or re-sorted throughout the text, and
4. A translation from the English Wikipedia to Japanese in UTF-8.

We evaluated all of the systems, however, we provoked errors in the systems, hit internal server errors, or had complaints that the text was not long enough, because a sentence - a sequence of Japanese symbols - looked like just one word to the system. None of the systems that were installed on a PC were able to deal with the texts at all.

Of the online systems, there were 4 that were able to find any of the plagiarisms:

1. Turnitin was able to flag all of the plagiarisms in 1-3 in both encodings, but could not find the translation.
2. Plagiarism Search was able to flag the first and the second one in the UTF-8 encoding.
3. StrikePlagiarism and PlagScan were both able to flag the first one in the UTF-8 encoding.

Summary

Plagiarism detection is not easy to do with software. Software can detect copies, but not plagiarisms, which can extend to paraphrases of a text, edited versions of a text, or even translations. Software in general just tests a portion of a text, not the entire document, as this would take far too much time. It is not always clear what portion of the text has been tested, and the numbers given for plagiarism found is an approximation at best, a random number at worst. Some systems flag proper quotes or properly referenced work as plagiarisms, others miss clear plagiarisms that for some reason have slipped under the radar of the system.

We cannot recommend any of these systems for general use. The ones listed partially useful could be used for situations in which a professor or teacher is suspicious about a text and cannot quickly find a source using a search machine. But in general, 3-5 longish words from a suspicious paragraph suffice for finding sources that are indexed by a search machine.

Instead, we suggest teaching students specifically about plagiarism, and focusing more on education and avoidance than on detection and punishment.