

## Anti-plagiarism strategies: How to manage it with quality in large-scale thesis productions

*Ken Larsson*

Department of Computer and Systems Sciences  
Stockholm University, Sweden  
[kenlars@dsv.su.se](mailto:kenlars@dsv.su.se)

*Henrik Hansson*

Department of Computer and Systems Sciences  
Stockholm University, Sweden  
[henrik.hansson@dsv.su.se](mailto:henrik.hansson@dsv.su.se)

**Keywords:** anti-plagiarism, management, peer-review, support, supervision, thesis

*A previous version of this paper was presented at the 5th Integrity and Plagiarism Conference, Newcastle-Upon-Tyne, UK, 16-18 July 2012.*

### Abstract

More than 400 students write their bachelor's or master's theses each year at the Department of Computer and Systems Sciences, Stockholm University. In order to support self-driven student thesis work and to reduce the burden on supervisors for feedback on basic skills, an IT support system called SciPro was developed. An important consideration in developing this system was to take actions to reduce plagiarism. Both prevention and detection were accomplished with the following: 1) *prevention* by policy guidelines, FAQ, face-to-face information, peer-reviews and transparency in the process of recurrent online thesis manuscript hand-ins; and 2) *detection* by automatic originality check of the final manuscript enabled by integration between SciPro and *Turnitin*. Explicit rules and regulations as well as frequent education about anti-plagiarism targeting both students and supervisors were also important parts of the prevention strategy. Current results include: 1) substantial improvements in policy development; 2) successful integration of anti-plagiarism software; and 3) recurrent educational activities for students and supervisors have raised the awareness of plagiarism issues at the department. Future development includes three new technical approaches in order to manage sophisticated anti-plagiarism controls efficiently, with a quality standard not possible by other means, in large-scale thesis production: 1) automated and integrated (SciPro/*Turnitin*) *recurrent anti-plagiarism controls* of submitted thesis manuscripts at various stages in the thesis text production process; 2) automated anti-plagiarism controls of thesis texts submitted in SciPro by *comparing consistency in style of writing* between different versions of thesis manuscripts handed in by the same student during the process of producing the thesis text; and 3) an automated check of thesis manuscripts submitted to SciPro for identification of images/figures/illustrations/graphs copied from the Internet through integration of an *image pattern recognition programme*. These measures taken together will significantly increase thesis quality by verifying authenticity to a very high degree and systematising the anti-plagiarism procedures. They will also substantially reduce tedious, boring and immensely time consuming manual work for administrators and supervisors who need to guarantee that theses do not contain plagiarised texts or illustrations.

## Introduction

Approximately 150 master's students and 300 bachelor's students write their theses each year at the Department of Computer and Systems Science, Stockholm University. In order to support self-driven student thesis work and reduce the burden on supervisors for feedback on basic skills, a digital system called SciPro (supporting the **Scientific Process**) has been developed (Hallberg, Hansson, Moberg, & Hewagamage, 2011; Hansson, 2012; Hansson, Collin, Larsson, & Wettergren, 2010; Hansson, Larsson, & Wettergren, 2009; Hansson & Moberg, 2011; Larsson & Hansson, 2011). This system contains a number of modules, which facilitate management, communication and learning. Among other things the SciPro system prevents plagiarism by providing the following: 1) clear instructions about rules and regulations for students and supervisors; 2) an online peer-review system; 3) transparent online communication and file storage of accumulated manuscripts; and 4) a final seminar module enabling automatic generation of originality reports from *Turnitin* when students upload their final manuscripts.

The use of software for detection of plagiarism in thesis work is an integrated part of the model for systematic support of the thesis production process at the department. Besides the technical requirements that the software should detect plagiarism without false indications, there are factors such as rules and regulations, usability, technical requirements etc., to consider, when deciding on what software tool to integrate in a pedagogical support system.

The design of SciPro is based on several years of study of the problem area of how to efficiently manage an ever increasing number of theses produced by more students with constant or decreasing supervisor resources (Hansson et al., 2009; Larsson & Hansson, 2011). It is also a big step for the student to change learning style from reproducing knowledge (taking courses) to creating knowledge (producing a scientific thesis), and this is an unfamiliar process that involves exploring a new territory of learning (Cohen, Manion, Morrison, & Morrison, 2007; Creswell, 2009). The aim of the SciPro system is to support both the supervisor and the student in the thesis work and increase the quality of the theses.

The procedure for designing the support for students and supervisors in this process included selecting suitable anti-plagiarism software for integration in SciPro. For a successful design it is not only the software selection that is important, but also the design of the process support. In SciPro the alignment of this support to existing rules and regulations regarding plagiarism, as well as aligning the software and process for the future of SciPro, discussing future plagiarism policy, and using software to detect instances of plagiarism were key design structures. Equally important were training, discussions and common understandings on plagiarism issues with students and supervisors. This is part of the ongoing quality assurance work in order to improve processes and policies for detecting and preventing plagiarism.

## Related studies

A large number of studies have been conducted on plagiarism. One category of studies is focused on the students' perspectives on plagiarism (such as Chuda, Navrat, Kovacova, & Humay, 2012; Dee & Jacob, 2012; Ford & Hughes, 2012; Stetter, 2012; Zafron, 2012). As concluded in the study by Chuda et al. (2012), software is needed in order to cope with the vast material at hand: "...while detecting plagiarism is necessary and important, there are too many students and too few staff members to make this easy. Automating plagiarism detection would help very much" (Chuda et al., 2012, p. 27). Ford and Hughes (2012) also concluded that software integration in courses was necessary as well as refining and raising awareness of the guidelines.

Cameron, Zhao and McHugh (2012) as well as Howard (1995) and Pecorari (2003, 2008) discuss the concept “patchwriting” to help understand and explain the ‘copy and paste’ done in articles by non-native English speakers, who may not deliberately intend to plagiarise, but struggle to write comprehensibly in a global scientific community dominated by the English language. Staff perspectives on plagiarism is less studied. In a questionnaire study by Gourlay and Deane (2012) teachers were asked to “compare current student performance to when they began teaching. Fifty two out of 80 respondents expressed a view that student writing was ‘poorer’, that the standard was ‘lower’, or that more help was required than before” (Gourlay & Deane, 2012, p. 22).

There are also studies on staff and students’ experiences and attitudes on using plagiarism tools such as *Turnitin*. A study by Jones (2008) focuses on the practical issues of using the system, such as applying procedures for submission of student work in such a way that it does not generate more work than without the system (Jones, 2008). Another perspective is provided by Savage (2004) who focuses on the experiences and perceived use of the system. In summary, both students and staff were positive towards the use of detection tools, with the reservation that it is a deterrent rather than a solution to the plagiarism issue (Savage, 2004). This is supported by Lukashenko, Graudina and Grundspenkis (2007) who put forward the view that detection requires less time to implement compared to prevention, but the positive effects of detection are short-term while the positive effects of prevention are long-term (Lukashenko et al., 2007).

Comparison of different tools for detection has been conducted from various perspectives; from broad studies to evaluations of the effectiveness of the systems (Bull, Collins, Coughlin, Sharp, & Square, 2001), evaluation of the functionality of different systems (Lukashenko et al., 2007) or descriptions of available systems (Culwin & Lancaster, 2000).

### **Prevention and detection**

There are two fundamentally different, but complementary, approaches in the efforts to reduce plagiarism: prevention and detection. It is fair to assume that to be successful the work has to include both of these in parallel. Focusing on detection only would probably result in more work in the long-term for the supervisors and administrators as the positive effects of this are of a short-term nature (Lukashenko et al., 2007). With the focus on prevention, it is probable that you will end up not knowing how successful this strategy has been unless detection is used to verify that your prevention is having the desired results by actually reducing the cases of outright plagiarism and fraud. In the context of having to handle more students with less supervision the strategy should include both sides of the coin: prevention and detection.

The type of assessment to be used is also a factor to consider in self-reported occurrences of plagiarism, where for instance programming and essay type assignments are more prone to include plagiarism in some form while presentations have less plagiarism (Alam, 2004).

When working to address plagiarism one also needs to consider the policy and regulations that are applicable on the international, national and local levels as well as how these rules should be implemented in practice. It is also important to establish a common definition of what plagiarism is among students and faculty.

#### *What is plagiarism?*

Defining plagiarism seems to be fairly straightforward, but even if you find a definition that appears to be clear and not subject to misunderstanding it is not necessarily

accepted worldwide (Pincus & Schmelkin, 2003). Yeo (2007) proposes a useful definition that is also used in our context:

Plagiarism means knowingly presenting the work or property of another person as one's own, without appropriate acknowledgement or referencing.

This definition is chosen, as it is appropriate for the type of plagiarism probably most common in thesis work, which is presenting something as your own original work. There are other definitions that probably would be more suitable for other types of examinations.

An important factor is also what is considered to be plagiarism by various stakeholders in education: students, supervisors, administrators, and evaluators from national agencies for higher education. In the study by Yeo mentioned above, students were asked to determine if a scenario is plagiarism or not. Most of the students considered copy-and-paste and copying another student's assignment to be plagiarism. While as many as a third considered copying a graph, copying text or poor referencing not to be plagiarism (Yeo, 2007), still these are important parts of a thesis work.

#### *Policy on plagiarism*

Internationally, the European Science Foundation, part of the European Collaboration on Science and Technology, put forward common viewpoints on research. In the report *The European Code of Conduct for Research Integrity* (Drenth, 2009) plagiarism is stated to be "an unacceptable form of misbehaviour, and a violation against other researchers". In a similar way, the Swedish Research Council lists plagiarism as misconduct in research (Swedish Research Council 2011). At university level there are rules of conduct for the student that are in alignment with national and EU policy documents, but these rules also regulate how disciplinary actions should be used and contain detailed instructions for professors and administrators on how to handle such cases. The Swedish National Agency for Higher Education reports that in 2011 a total of 517 students were disciplined for plagiarism, a significant increase compared with 68 cases ten years earlier (National Agency for Higher Education, 2012).

#### *Preventing plagiarism*

There are several strategies suggested on how to prevent plagiarism, from the design of assignments (this may be most applicable for hand-in assignments, but hard to apply for thesis work) to special education and information to students about plagiarism and consequences if they choose to engage in it. It is suggested that the best prevention is accomplished by training the students in what constitutes plagiarism and how to handle this in their own work (DeVoss & Rosati, 2002; Vernon, Bigna, & Smith, 2001). It is also suggested that peer review is a tool for prevention of plagiarism (Swedish Research Council, 2011). The peer review process in SciPro is described and analysed in depth in "Peer Portal: Quality enhancement in thesis writing using self-managed peer review on a mass scale" (Aghaee & Hansson, 2013). Some studies suggest that by introducing discussions in class on plagiarism the students will be more likely to be more honest in working with their assignments (Austin & Brown, 1999). For thesis work this is probably best implemented by including material on plagiarism that is available for all students. Ideally, this material can be referred to by the supervisor, as there is seldom time for the supervisor to have this type of discussion with each and every student.

#### *Detecting plagiarism*

For detecting plagiarism there are two different approaches; the supervisor uses his/her knowledge or gut feeling to detect that a submitted work is composed of minor or major parts that are not the author's own, or by using software that compares the

submitted text with other document collections in order to find similar text. When using software for detecting plagiarism there are four different phases according to Culwin and Lancaster (2001): collecting the text, analysis of the text, verification of the analysis and finally, investigation of those submissions that may be cases of plagiarism. The first two phases use some form of IT system to upload and automatically analyse the text. Verification needs to be done by a human (usually the supervisor) with some software support and this applies to the subsequent investigation as well. This implies that supervisors and examiners need to be trained in order to do the verification correctly and to be able to assess how it relates to the plagiarism policy. Proper handling of the verification step is a prerequisite for conducting the investigation in a fair and just way.

There are different motivations for using plagiarism detection. The most obvious is probably because plagiarism is suspected and a detection system is used to confirm or disprove the suspicion. A less obvious reason is to apply the detection process on all work submitted by all students. This approach is more proactive as it not only detects and deters students from misconduct, but also may help the student develop the skills of scientific writing and referencing (Barrett, Malcolm & Lyon, 2003). However, this is very time consuming, if the control is not automatic.

Strategies for using software tools for detecting plagiarism have changed over the years. From semi-manual detection with the use of search engines like Google to automatic detection with the help of specialised systems. The use of Google as a tool for detection is found to be efficient in finding word-for-word plagiarism (McCullough & Holmberg, 2005), but due to the time involved it may be more suitable for smaller groups of students and supervisors. This procedure involves finding suspicious phrases in a thesis and manually testing these in a search engine. Using special software for detection makes the process more automatic as it is not needed to find suspicious phrases; the submitted text is uploaded to the system and automatically checked for originality.

### **Aim**

The aim of this work was to evaluate the present stage of the SciPro process in regards to prevention of plagiarism in the thesis writing process. This included the integration of anti-plagiarism software, and improved workflows and processes to use the originality reports in order to increase the quality of the thesis. One important part of this was to propose a strategy which contains both prevention and detection techniques as well as suggestions for suitable training of all stakeholders in the thesis writing process.

### **Method**

The study was conducted by workshops with supervisors and students to evaluate the users' views of the process and support. As this was the second pilot run of the system, and the first to include the integration of the anti-plagiarism software, the strategy was to gather in-depth opinions on all aspects of the usability of the system, as it concerns the upload and automatic analysis of the text, and propose proper measures that will enhance the quality of verification.

The first stage of this work was to decide on which detection system to use for automatic integration in the system. In order to evaluate current anti-plagiarism software in terms of accuracy, relevance and usability for our purpose a number of studies was set up, the most comprehensive being a bachelor's thesis that was supervised by a team of supervisors and provided both technical and human interaction perspectives (Appelgren Heyman & Olofsson, 2011). The evaluation concluded that *Turnitin* was the most suitable software for the requirements that were

established for the evaluation, and also the fastest system tested. *Turnitin*, based on the aforementioned results, was selected and integrated in the IT-support system for supervision (SciPro). This thesis also had implications for which software for anti-plagiarism several other Swedish universities later selected as their option.

Another study, results first presented here, focused on the originality reports from *Turnitin* generated by SciPro in our department and how the responsible supervisors treated the reports. The material consisted of selecting a full sample of all completed theses (22) during the spring semester 2012 for analysis. The originality report generated for each completed thesis was read, and divided in three categories: 1) those with more than 90% text matching score (probably a false report); 2) those with low scores, less than 10% text matching score (no plagiarism, since the literature reference information and quote matches were not problematic); and 3) the group with a higher score than 10% text matching. The last category was analysed in detail to determine the likelihood of a genuine case of plagiarism, and this category comprised about 10% of the sample.

### Support the scientific writing

The origin of SciPro goes back to some draft project ideas in 2007/2008 that were further researched and presented as a design concept in 2009. Core concepts were to construct a flexible and semi-adaptive IT support system which allowed the supervising process to be individualised (Hansson et al., 2009). The point of departure for this concept was twofold; the economic factors of the increasing number of theses to be handled by fewer supervisory resources and to aid the students in making the transformation from reproducing to producing knowledge. Key aspects of this concept are the dialogue in the supervision process, utilising digital multimedia content and library resources. The concept includes a novel use of multimedia that is interactive and nonlinear. An originality check was envisioned to be part of all submissions in all phases of the writing process, and is currently called automatic check of form (see Figure 1).

Phase and sequence	Audiovisual Instructions /teaching material	Thesis text Version 1	Automatic check of form	Peer-review	Peer Community Sharing, Tips, links, resources	Thesis text Version 2	Supervisor
--------------------	---	-----------------------	-------------------------	-------------	--	-----------------------	------------

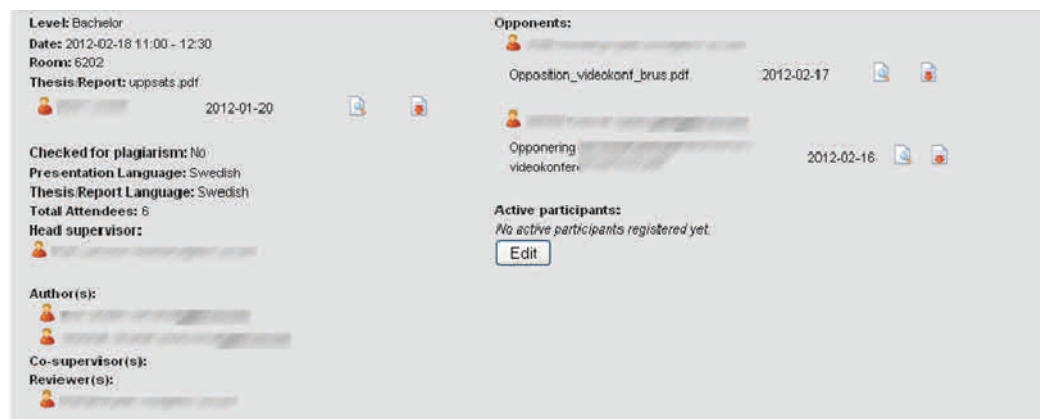
**Figure 1.** First concept of originality verification (Hansson et al., 2009)

At this stage, the anti-plagiarism consisted of two different strategies: prevention by the use of peer review and detection by automatic check for originality using specialised software (*form or structure style and patterns, i.e. text and picture matching and similar tools*).

So far the work had been focused on researching the concept as such, as well as gradually improving and refining the concept as more data was collected from workshops and presentations. Gradually the focus was shifted to prepare for producing a demonstration of the concept in the form of a skeleton system that focused on the most cumbersome parts of the supervision process.

For the first pilot run the skeleton system consisted of tools to bring students and supervisors together without the need for the students to roam the corridors, which means 400 plus students knocking on some 70 doors trying to get a supervisor. Each supervisor and student saves time by using the new IT-matching module in SciPro. Somewhere between 5 and 40 hours of administration are saved and instead a thesis project can just start. For more detailed descriptions of the new start phase see

(Hansson, Moberg, & Peiris, 2012). The pilot system also included a first version of the Content Management System (CMS) to reach and structure media and library resources and a special module was created to manage the final seminars. The final seminar module reduced administrative work for the supervisor and students presenting their thesis or attending a seminar. The final seminar module made it easier for students to find seminars to attend, to access the final thesis text and to register as an opponent (a mandatory assignment in the thesis course). Figure 2 shows a screenshot with information related to the final seminars in the SciPro system. The supervisor creates a seminar (time, date, room etc.) in SciPro for the degree project when the thesis manuscript is assessed as being of an appropriate standard. When a seminar is created by the supervisor, the SciPro system automatically lists the author/s, supervisor and reviewer involved in the project. Then the author uploads the final thesis manuscript to the seminar module, making it available to all. The opponent and other participants register for the seminar and submit their opposition reports (templates are provided to structure these reports according to the grading criteria, which are also provided) via the system.



**Figure 2.** Screenshot: Example of information provided by the final seminar module in SciPro

### *The second pilot run*

For the second pilot run in the academic year of 2011/2012, the functions for peer-review and automatic check for originality using *Turnitin* were added. Both of these functions were tested in smaller groups during the autumn of 2011. Also the checklists were expanded so that the author could add a visible signal and comments for each point in the checklist (see Figure 3); this prevents plagiarism in two ways – by educating the author and increasing awareness of what constitutes plagiarism.

Question	(Supervisor)	(Author)	(Author)
Are the background, aim, research questions/problem statement, methods and theory in alignment?			
Is the background relevant and sufficient with references to previous studies?			
Are selected methods and theories relevant and justified?			
Is the motivation of selected data collection method(s) sufficient?			
Are alternative data collection methods discussed and excluded on logical grounds?			

**Figure 3.** Example: Status of checklist items in SciPro. Used for self-evaluation by students (author) and confirmation/disagreement by a supervisor and/or peers in peer reviews.

The checklists can be commented on by the author/s and supervisors (see Figure 4), in addition to indicating green, yellow or red as the status for each question in order to more quickly assess the status of a large number of predefined aspects in a thesis manuscript. Comments on points that relate to preventing plagiarism serve to promote continuous education of the student on how to handle use of others' work in the thesis. Also the checklists, which are adaptable (supervisor can add more questions/ to do tasks), aim at educating the student and serve as a tool for reflective self-assessment. Checklist items can include both internal and external links to additional relevant information in the SciPro system or anywhere on the internet (Figure 5). Links from checklists to ethical codes of conduct in research, anti-plagiarism policies etc. are provided.

Is the background relevant and sufficient with references to previous studies?

Select

Ok Minor corrections Not done Not applicable

Comment

References are missing in most parts!

Save

**Figure 4.** Comment on checklist items

**Question**

Write according to the following subheadings

Background and state of art. Check relevant journals: <https://scipro. /resource/journals>. Search in the library: <https://scipro. /resource /how-search-library-database> and use suitable search engines: <https://scipro. /resource/search-engine>. Define key concepts, some may be found here: <https://scipro. /wordlist-english>

Aim and research questions/ Problem statement

**Figure 5.** Examples of links in checklist items

#### *Automatic check for originality – detection*

After the evaluation of available anti-plagiarism systems for originality check that were suitable to be integrated in the SciPro system was completed (Appelgren Heyman & Olofsson, 2011), it was decided that *Turnitin* was the most appropriate to use for seamless integration. Apart from accuracy, a deciding factor was speed in processing, measured in minutes instead of hours. The quality of the originality check as well as usability and readability of the generated results report were also strong factors. Implementation of the automatic check software was done so that when a student submitted his/her final thesis manuscript via the final seminar module in SciPro, the manuscript was automatically sent to *Turnitin*, which generated the originality report. The originality report created by *Turnitin* was then sent back to the SciPro-supervision system and was available for the author, supervisor and examiner for assessment.



As discussed above there are two main strategies for detecting plagiarism. The supervisor either looks for suspicious parts in the manuscript handed in by the student or uses some form of software. Choosing dedicated software as a method of detection has the advantage that it is faster, more systematic and it also offers consistency and transparency of the process. With the use of software for detection all students are subjected to the same process, reducing the risk of unfairness and bias. The transparency of the process supports the preventive actions to curb plagiarism. Deterring plagiarism is a management issue, and one of several other issues in handling large numbers of students and theses. There are several steps in managing plagiarism, such as the use of consistent processes and keeping track of originality reports (Zobel & Hamilton, 2002). Selecting a specialised tool for detection is also motivated by the transparency and fair treatment of each and every student.

#### *Preventing plagiarism – education*

There is a continuing effort to educate both students/authors and supervisors/reviewers about plagiarism and honesty in presenting one's own and others' academic work in a fair way. Presently this mostly takes place by inclusion of study materials in the SciPro system on the different aspects of writing theses. It is also part of the introduction on policy, procedures and consequences for new students. As described above the use of checklists is part of the education of students self-awareness. Checklists are also used in peer and supervisors' feedback in order to cover all relevant aspects systematically. Workshops on different aspects of supervision and using the system are conducted for the education of supervisors. Online support, complemented with a SciPro helpdesk providing weekly drop-in opportunities to address issues (for both students and supervisors) is also an important factor in the continual training. According to McLafferty and Foust (2004) there are two other aspects of plagiarism prevention besides education and training. These include the position the instructor has towards the students and the general atmosphere in their relationship and assignments that reduce dishonesty (McLafferty & Foust, 2004).

The need for training of supervisors is indicated by a survey of the originality reports filed in the administrative system. For the year 2010 (before it was mandatory to upload the originality report) a total of 169 theses were completed, but only 47 plagiarism reports were uploaded in the administrative system. Several anti-plagiarism systems, such as *Turnitin*, *Genuine text* and *Urkund*, were used in an unstructured and personal way by supervisors. During 2011 (when the plagiarism analysis was mandatory to upload into the central digital archive) there were a total of 157 theses completed and 119 originality reports uploaded. As there was no set procedure for conducting the originality check and no system support, the actual reports were not systematically available for evaluation. For the year 2012 the SciPro system was expanded to include the automatic generation of originality reports from *Turnitin*. At the time of sampling for this study there had been 22 theses completed and 20 had originality reports uploaded. See Table 1 for a summary of the proportion of theses with originality reports during these years. It is clear from this summary that the proportion of theses that are checked for originality has increased with the introduction of the SciPro system and the change in policy on the use of plagiarism tools. The reduced number of theses completed is explained by the introduction of a new process for grading completed work and a shorter sampling period of five months for 2012.

Table 1:  
*Proportion of theses with an originality report*

<b>Year</b>	<b>Percentage of originality reports</b>	<b>Assessment</b>
2010	28%	<i>Quality of report unknown</i>
2011	76%	<i>Quality of report unknown</i>
2012	91%	<i>25% probably false match 10% needs investigation</i>

Out of the 22 theses that were completed at the time of the study in 2012 all related originality reports had been accessed and categorised. The results were as follows:

- Two theses had no originality report. One thesis was missing an originality report and another thesis had an unrelated file uploaded instead of the originality report.
- One originality report had a supervisor comment informing about technical problems in retrieving the report from the anti-plagiarism system and when resubmitting the manuscript it got 100% text match, since the system matched with the earlier submitted version of the thesis, and is therefore a false match, not plagiarism.
- One originality report indicated 100% text match, but no explanation was provided by the supervisor regarding whether this was plagiarism or a false match.
- The originality reports of three theses indicated more than 90% text matches with other sources, or an earlier version of the manuscript in the same database, but no explanation or analysis was provided by the supervisor.
- Two theses had originality reports which indicated more than 10% match and lacked supervisor comments, analysis, or explanation.
- Thirteen theses had originality reports with less than 10% text match to other sources.

Another observation is that two of the originality reports were uploaded in the wrong file format and thus were not directly readable in the system.

Further study of the originality reports with a similarity index of 10% or more text matching is of interest as these are in the range where we would suspect possible plagiarism. Since the software detects text matching between the thesis manuscript and sources found elsewhere (internet, databases, etc.), a match of the same text does not necessarily mean that it is plagiarism; quotes, bibliographic references, appendices with source reference, etc. are perfectly acceptable. In fact an originality report with indication of no text match at all (0%) is very strange since it means that the thesis uses references not found anywhere else, which is highly unlikely. By investigating the reports with a similarity index of 10% or more, lessons can be learned on how to train supervisors in the process of verification. By just looking at the saved transcript of the originality report it became evident that this is not sufficient to verify or discard the possible plagiarism. Some of the sources are not possible to find without the link in the online materials, and as it turns out even with the link it is hard to find the original if it is a student report.

An attempt to perform verification on these two theses revealed that it is necessary to have access to the alleged copied original text source in order to verify true or false plagiarism. Even though the system provides a link to the text matched source it is not always possible to access the text as the link may require login or payment. The first

student thesis had a similarity index of 12% (more than a third of this was from student reports). Many of these used common expressions, but one stood out and needed to be investigated further as the originality report indicated that the main part of the purpose of the study was very close to another student report.

The second thesis had a similarity index of 31% out of which half was from other students' papers, and mostly from papers submitted to the same department. The verification indicated a potential case of plagiarism as there were several long paragraphs that were verbatim matches. It was also noted that the thesis contained more direct quotes than what is normally found in a bachelor's thesis. The author apparently used cut-and-paste.

The quality of verification is influenced by educating the supervisor/instructor and part of the information is aimed at supervisors. Transparency and fairness in the process are also factors that can have a positive effect on the supervisor–student relationship in a constructive way regarding the prevention of plagiarism.

### **Conclusions and recommendations**

There have been substantial improvements in policy development both on the international and national level, driven by an increased awareness of the plagiarism problem. Likewise, policy improvements on anti-plagiarism issues at the university and department level are now much more elaborate than in previous years.

We have at our department (one of more than 70 departments at the University) successfully integrated anti-plagiarism software with the SciPro thesis supervision support system. This has substantially reduced the administrative tasks for the supervisor, thus saving time. As the originality report is mandatory before the final seminar, all theses at the department are quality controlled for plagiarism systematically, and the results are transparent for all stakeholders.

By means of recurrent educational activities for students and supervisors in the form of kick-offs, workshops, seminars, newsletters etc., and online information such as FAQs, study guides and handbooks, we have raised awareness of plagiarism issues in the department. These actions are mainly designed to prevent plagiarism, and to inform about legal and other consequences.

For future work there are many ideas from the original concept to address and continue to work on and these include: 1) automatic anti-plagiarism check of thesis work from first until final thesis draft (recurrent controls); 2) comparisons of consistency and changes of personal language expressions between different versions of a student's manuscript; 3) pattern recognition of images, illustrations and tables; and 4) automatic and recurrent checks as part of the peer-review process that are visible for both author and reviewer. These efforts will greatly enhance understanding of plagiarism and prevent further plagiarism by education.

It would enhance a student's education if a system for checking selected submissions for originality during the thesis work could be integrated. In the borderline cases of plagiarism, it would then become visible for the students at an early stage and be voluntarily corrected before being handed in to the supervisor.

A system that can analyse text for consistency and changes in personal writing style in the work itself and between different versions of the manuscript will indicate potential plagiarism. There are several so-called *Paper Mills* available on the web for easy purchase of everything from a simple lab report to a full dissertation, and even PowerPoint™ presentations can be obtained in this way. By making students aware

that this check will be done will assist in a more clear definition of what constitutes plagiarism (Vernon et al., 2001).

From the study by Yeo mentioned above there is a need for detection of copied illustrations. This would be a valuable tool in the strategy to reduce plagiarism. Students need to learn to work with these important parts of a thesis with the same honesty as text. The study revealed that as many as one third did not consider copying images, tables etc. to be plagiarism (Yeo, 2007).

## References

- Aghaee, N., & Hansson, H. (2013). Peer Portal: Quality enhancement in thesis writing using self-managed peer review on a mass scale. *The International Review of Research in Open and Distance Learning*, 14(1), 186–203.
- Alam, L. S. (2004). *Is plagiarism more prevalent in some forms of assessment than others*. Paper presented at the 21st ASCILITE Conference, Perth, Western Australia.
- Appelgren Heyman, F., & Olofsson, M. (2011). *Utvärdering av antiplagiatsystem för högre utbildning (eng; Evaluation of anti-plagiarism software)*. (Bachelor), Stockholm University, Stockholm.
- Austin, M. J., & Brown, L. D. (1999). Internet plagiarism: Developing strategies to curb student academic dishonesty. *Internet and Higher Education*, 2, 21–34.
- Barrett, R., Malcolm, J., & Lyon, C. (2003). *Are we ready for large scale use of plagiarism detection tools*. Paper presented at the 4th Annual LTSN-ICS Conference, NUI Galway.
- Bull, J., Collins, C., Coughlin, E., Sharp, D., & Square, P. (2001). Technical review of plagiarism detection software report. *University of Luton. JISC publication*.
- Cameron, C., Zhao, H., & McHugh, M. K. (2012). Perspective: Publication ethics and the emerging scientific workforce: Understanding “plagiarism” in a global context. *Academic Medicine*, 87(1), 51.
- Chuda, D., Navrat, P., Kovacova, B., & Humay, P. (2012). The issue of (software) plagiarism: A student view. *Education, IEEE Transactions on Education*, 99, 1–1.
- Cohen, L., Manion, L., Morrison, K., & Morrison, K. R. B. (2007). *Research methods in education*. Abingdon UK: Psychology Press.
- Creswell, J. W. (2009). *Research design: Qualitative, quantitative, and mixed methods approaches*. London: Sage Publications, Inc.
- Culwin, F., & Lancaster, T. (2000). A review of electronic services for plagiarism detection in student submissions. *ITICSE'00*.
- Culwin, F., & Lancaster, T. (2001). Plagiarism, prevention, deterrence & detection. *Institute for Learning and Teaching in Higher Education, South Bank University, U.K.*
- Dee, T. S., & Jacob, B. A. (2012). Rational ignorance in education a field experiment in student plagiarism. *Journal of Human Resources*, 47(2), 397–434.
- DeVoss, D., & Rosati, A. C. (2002). “It wasn’t me, was it?” Plagiarism and the web. *Computers and Composition*, 19(2), 191–203.
- Drenth, P. J. D. (2009). A European code of conduct for research integrity: European Science Foundation.
- Ford, P., & Hughes, C. (2012). Academic integrity and plagiarism: Perceptions and experience of staff and students in a school of dentistry – A situational analysis of staff and student perspectives. *European Journal of Dental Education*.
- Gourlay, L., & Deane, J. (2012). Loss, responsibility, blame? Staff discourses of student plagiarism. *Innovations in Education and Teaching International*, 49(1), 19–29.
- Hallberg, D., Hansson, H., Moberg, J., & Hewagamage, H. (2011). *SciPro from a mobile perspective: Technology enhanced supervision of thesis work in emerging regions*. Paper presented at the Aitec East Africa ICT summit at Oshwal Centre, Nairobi, Kenya.

- Hansson, H. (2012). *4-excellence: IT system for theses. Going global: Internationalising higher education*. Paper presented at the British Council conference, London.
- Hansson, H., Collin, J., Larsson, K., & Wettergren, G. (2010). *SCI-PRO: Improving universities core activity with ICT supporting the scientific thesis writing process*. Paper presented at the Sixth EDEN Research Workshop Budapest.
- Hansson, H., Larsson, K., & Wettergren, G. (2009). *Open and flexible ICT: Support for student thesis production – design concept for the future*. Paper presented at the Cambridge International Conference on Open and Distance Learning, Cambridge, UK.
- Hansson, H., & Moberg, J. (2011). *Quality processes in technology enhanced thesis work*. Paper presented at the 24th ICDE World Conference on Open and Distance Learning, Bali, Indonesia.
- Hansson, H., Moberg, J., & Peiris, R. (2012). *SciPro idea bank: Matchmaking ideas, people and organizations to facilitate innovative theses*. Paper presented at the Advances in ICT for Emerging Regions (ICTer) 2012 International Conference, Colombo.
- Howard, R. M. (1995). Plagiarisms, authorships, and the academic death penalty. *College English*, 57(7), 788–806.
- Jones, K. O. (2008). *Practical issues for academics using the Turnitin plagiarism detection software*. Paper presented at the Conference on Computer Systems and Technologies and Workshop for PhD Students in Computing Gabrovo, Bulgaria.
- Larsson, K., & Hansson, H. (2011). *The Challenge for supervision: Mass individualisation of the thesis writing process with less resources*. Paper presented at the Online Educa Berlin 2011, 17th International Conference on Technology Supported Learning & Training, Berlin, Germany.
- Lukashenko, R., Graudina, V., & Grundspenkis, J. (2007). *Computer-based plagiarism detection methods and tools: An overview*. Paper presented at the Proceedings of the 2007 International Conference on Computer systems and Technologies, Bulgaria.
- McCullough, M., & Holmberg, M. (2005). Using the Google search engine to detect word-for-word plagiarism in master's theses: A preliminary study. *College Student Journal*, 39(3), 435.
- McLafferty, C. L., & Foust, K. M. (2004). Electronic plagiarism as a college instructor's nightmare – Prevention and detection. *The Journal of Education for Business*, 79(3), 186–190.
- NationalAgencyforHigherEducation. (2012). *Disciplinärenden 2011 vid universitet och högskolor (eng. Disciplinary cases at universities 2011) Swedish National Agency for Higher Education*. Stockholm, Sweden.
- Pecorari, D. (2003). Good and original: Plagiarism and patchwriting in academic second-language writing. *Journal of Second Language Writing*, 12(4), 317–345.
- Pecorari, D. (2008). Plagiarism, patchwriting and source use: Best practice in the composition classroom. In P. Friedrich (Ed.), *Teaching academic writing* (pp. 222–241). London: Continuum.
- Pincus, H. S., & Schmelkin, L. P. (2003). Faculty perceptions of academic dishonesty: A multidimensional scaling analysis. *Journal of Higher Education*, 196–209.
- Savage, S. (2004). *Staff and student responses to a trial of Turnitin plagiarism detection software*. Paper presented at the Australian Universities Quality Forum 2004, Adelaide, Australia.
- Stetter, M. E. (2012). Teaching students about plagiarism using a web-based module. *Teaching students about plagiarism using a web-based module*, 37(5), 675–693.
- Swedish Research Council. (2011). *Misconduct in research*. Retrieved March 26, 2011, from <http://www.codex.vr.se/en/etik6.shtml>

- Vernon, R. F., Bigna, S., & Smith, M. L. (2001). Plagiarism and the web. *Journal of Social Work Education*, 37(1), 193–196.
- Yeo, S. (2007). First year university science and engineering students' understanding of plagiarism. *High Education Research & Development*, 26(2), 199–216.
- Zafron, M. L. (2012). Good intentions: Providing students with skills to avoid accidental plagiarism. *Medical Reference Services Quarterly*, 31(2), 225–229.
- Zobel, J., & Hamilton, M. (2002). Managing student plagiarism in large academic departments. *Australian Universities Review*, 45(1), 23–30.

### **Acknowledgement**

A previous version of this paper was included in the *International Integrity and Plagiarism Conference Proceedings*, Newcastle-upon-Tyne, UK, 16-18 July 2012.

### **About the authors**

**Ken Larsson** is a lecturer in Computer and Systems Sciences at Stockholm University and teaches courses in user-centred systems design and creativity. He also works as a systems designer with the study administrative system at the department. He has been involved in local, national and international projects in the field of Technology Enhanced Learning for some 15 years, and his main interest is in user experience and management.

**Henrik Hansson** is an Associate Professor at the Department of Computer and Systems Sciences, Stockholm University. He is the coordinator of the research group “ICT for Development” and also a member of another research group “Technology enhanced learning”. Currently Henrik is involved in two EU-funded R&D projects related to IT and education, *MyUniversity* and *We.Learn.it*, and two international Sida (Swedish International Development Cooperation Agency) funded projects about ICT and pedagogical development involving 12 countries.