

Coping With the Copy-Paste-Syndrome

Narayanan Kulathuramaiyer

*Faculty of Computer Science and Information Technology,
Universiti Malaysia Sarawak, Malaysia*
nara@ift.unimas.my

Hermann Maurer

*Institute for Information Systems and Computer Media,
Graz University of Technology, Austria*
hmaurer@icm.edu

Abstract: The Copy-Paste Syndrome describes a situation whereby students at all levels are becoming more and more reliant on wide-range of easily-available digital content. This is a universal problem that has to be addressed effectively, especially with the revolutionary development of the Web. Weber (Weber, 2006) refers to it as the Google-Copy-Paste-Syndrome, which according to him will drastically affect the quality of scientific publications, leading to a degradation of the quality of life. The expansion of digital content together with an emerging participative social learning (and E-learning) ecosystem could result in even more devastating implications. As opportunities for the proliferation of such infringements becomes widespread, a holistic solution is required combining an institutional approach together with the application of viable technologies. This paper describes an E-learning ecosystem combined with a copy-paste detection suite to comprehensively address the emerging phenomenon.

Introduction

Web 2.0 (O'Reilly, 2006) describes the evolution of the Web which is fast becoming a platform for social networks. Its social engineering power is demonstrated by the many emerging applications such as MySpace, Wikipedia, Flickr and YouTube. The re-mixability of multiple services facilitates highly personalized experiences to suit individual needs.

Web 2.0 is changing the way learning communities build, organize, share and exchange knowledge. E-learning 2.0 (Downes, 2006) explores the social empowerment of learners by harnessing the social power of Web 2.0. Wikis and blogs have provided a powerful means for learners in expressing themselves and collaboratively creating knowledge. E-learning 2.0, is thus anticipated to be able to address the unfulfilled goals of Learner management Systems of the past (Pitner and Drasil, 2006).

In the development of E-learning 2.0 and future E-learning systems, there are social implications that will have to be addressed. As tools get easier to use, it also becomes easier for networked learners to commit violations such as plagiarism and IPR violation. It will also become much simpler to acquire information from the web community as opposed to meeting up with co-learners and experts in the real world (Alexander, 2006). Publishing online has also become much easier. The openness of the Web environment poses a number of challenges in monitoring and keeping track of the explorative expressions of students.

The term copy-paste is used in this paper to refer to an emerging practice of fast and easy publication by millions of people. The 'Google Copy Paste Syndrome' (GCPS), (Weber, 2006) describes a common activity of performing a fast, easy and usually "not diligently researched" copying of passages in text by people of all walks of life which includes scientists, journalists, academics and students. The GCPS has resulted in a proliferation of infringements such as plagiarism and IPR violations. Acquiring insights is performed by 'conveniently searching' the Web as opposed to a rigorous process of learning through scientific discovery. Information from Web sources such as Google and Wikipedia are often used without even considering the validity of the source. According to Weber, GCPS will lead to a degradation of scientific quality, and eventually affecting the quality of life.

Plagiarism is seen as dangerous form of copy-paste particularly in an academic environment, where it could affect both the credibility of institutions as well as the quality of its graduates. Plagiarism has been constantly on the rise, (see Plagiarism Stats, 2007) largely attributed to the ease of access to a large number of paper mills, (see Paper Mills, 2007 for a large list) and other sources of information such as search engines, web directories, Wikipedia, book reviews on online bookstores and scholarly publications. **Plagiarism and copy-paste syndrome**

(Midolo and Scott, 2003) highlights a sharp rise in CD burning and Internet file sharing. According to them, almost a third of music owned by people aged seventeen and under comes from illegally burnt CDs or Internet file sharing. Furthermore, more than half of those in this age group are not even aware that it is illegal to copy music without permission. (Midolo and Scott, 2003)

Examples of copy-paste performed in a classroom includes (Harris, 200) downloading a free research paper, buying a paper from a commercial paper mill, copying an article from the Web or an online, copying a paper from a local source or from friends and cutting and pasting to create a paper from multiple sources (including blogs and answer brokering sources). These infringements could be considered as plagiarism depending on to the extent of copying and the failure to appropriately cite sources. Forms of copying could range from verbatim copying of text, paraphrasing to even a translation of texts. Copying is however not restricted to textual sources, it may involve images and other multimedia documents as well.

Dealing with plagiarism (and copy-paste syndrome)

As described by (Kennedy, 2004), students are generally not aware of the full implication of the copy-paste syndrome. They do not value the importance of intellectual property or take pride in their ability to produce creative works. (Kennedy, 2004) There is a need to instill a moral and ethical values in students in relation to their education. Students will begin to understand the need to respect other people's copyright when they themselves are actively engaged in creating their own intellectual property (Midolo and Scott, 2003) Best practices in teaching and learning and academic integrity can be achieved if students are aware that their inputs have been valuable and taken into consideration by instructors. (Kennedy, 2004)

Students are often expected to read a policy in the handbook and thereafter comply with the non-plagiarizing culture. This approach is likely to be unsuccessful as the core problem lies in the fact that the students do not understand the concept of plagiarism and, most of all, do not know how to deal with it. (Kennedy, 2004) As pointed out by (Duff, et al, 2006) there is the lack of appreciation of the Western system of scholarship, especially among new students or foreign students. There is thus a need to teach the skills required to avoid plagiarism such as the ability to paraphrase, summarize and reference accurately. (Kennedy, 2004)

Another approach to avoid copy-paste syndrome is through the well structured and clearly articulated assessment tasks. Course designers will have to carefully design courses and course content to ensure that they do not encourage plagiarism. There are factors that may indirectly encourage plagiarism, e.g. the same questions are set each year, or if questions are not understood clearly or a clear criteria has not been specified. (Kennedy, 2004)

There are a number of approaches that can be undertaken in reducing the possibility of occurrence of plagiarism as suggested by works in (Harris, 2004). Instructors are encouraged to use of one or more sources not written within the past year. This approach is important as it could effectively invalidate results of paper mills. By using one or more specific articles, books or specific information specifically made available, students could be encouraged to formulate their own thoughts. Another effective technique is by requiring students to perform assignments as a series of process steps before the final completion of a project. Student learning could then be demonstrated and partially assessed at each stage. As a highly personalized student tracking and assessment administration may tend to overwhelm instructors, viable technologies to effectively address plagiarism and copy-paste-syndrome, need to be applied. A technological platform would be required to guide students in using material from various sources in a constructive way and to promote critical thinking.

Typical approach for Dealing with Plagiarism

A typical approach applied in educational institutions involves the employment of tools for plagiarism detection such as Turnitin or Mydropbox. An overview of a broad range of tools for fighting Plagiarism and IPR violation is presented in (Maurer,et al, 2006). A layered application of plagiarism detection has also been proposed by (Kulathuramaiyer, Maurer, 2007) to systematically perform elaborate mining on a smaller subset of documents. Table 1 summarizes the list of tools and techniques currently used for plagiarism detection. A suite of plagiarism detection tools is required to establish and substantiate a suspicion of plagiarism with as much evidence as possible. As these tools have been adequately dealt with in our previous publication, (Kulathuramaiyer, Maurer, 2007) this

paper will then focus more on a broader range of tools for fully addressing the copy-paste-syndrome.

Task	Tool
Manual Technique	Search Engines (Maurer et al. 2006)
Text-based Document Similarity Detection	Dedicated Software, Search and Web Databases (Maurer, Zaka, 2007)
Writing Style Detection	Stylometry software (Eissen, 2006)
Document Content Similarity	Semantic Analysis (Dreher and Williams, 2006) (Ong and Kulathuramaiyer, 2006) (Liu et al, 2006)
Denial of Plagiarism	Cloze Procedure (Standing and Gorassini, 1986)
Content Translation	Normalised Representation (Maurer, Zaka, 2006)
Multi-site Plagiarism	Distributed Plagiarism (Kulathuramaiyer, Maurer, 2007)

Table 1: Tools for Plagiarism Detection

Comprehensively Addressing Copy-paste Syndrome

In exploring a technological solution to comprehensively address copy-paste syndrome, the following questions need to be answered. Would it be possible to comprehensively address copy-paste syndrome, beyond the detection tools mentioned above? Can we not effectively employ technology in shaping teaching-learning to address the copy-paste issues presented earlier? We present an ecosystem consisting of a sophisticated blended learning environment developed at the Graz University of Technology with a powerful knowledge management platform, the Hyperwave Information Server as a basis for the development of comprehensive solution. These systems are augmented with a copy-paste detection and administration suite together with specifically prepared E-learning modules to address all the issues mentioned earlier. We will refer to the proposed suite of software and content as ICARE, aimed at the inculcating of effective copy-paste skills. ICARE, could either stand for **I**dentify-**C**orrelate-**A**ssimilate-**R**ationalize-**E**xpress or **I**nternalise, **C**onceptu**A**lise and **E**xp**R**Ess for short. Table 2 describes the functions to be implemented and corresponding technological solution to be employed in the ICARE ecosystem.

Function	Tool Needed
Education and Awareness	E-learning module on Effective Copy-Paste
Promoting Effective copy-paste	Controlled Environment for E-learning
Skills of managing Copy-paste	Self-paced and Just-In-Time learning with a copy-paste tracking suite
Procedural support for Ensuring Compliance	WBT-Master, Hyperwave Workflow System, knowledge maps
Well-defined Assessment	Pedagogy-based Blended Learning Environment
Plagiarism Detection Workbench	Refer to Table 1

Table 2: ICARE Ecosystem Components

ICARE Ecosystem : Comprehensively addressing Copy-Paste Syndrome

ICARE benefits from the sophisticated E-learning functionalities of WBT-Master and the knowledge and document management capabilities of Hyperwave Information System. Hyperwave is a distributed hypermedia system, which has demonstrated the ability to overcome limitations arising from the lack of structure of the WWW. E-learning systems built on top of Hyperwave such as eLS (Mödritscher et al, 2005) have been previously applied in the Classroom 2000 project in Northern Ireland. WBT-Master (WBT 2005) is a novel E-learning system which offers a wide range of support for personalized and adaptive learning, and learning management. The following features WBT-Master and Hyperwave, makes it absolutely suited for the ICARE E-learning ecosystem:

- The ability to define scenarios and to establish a controlled environment for student activity tracking enables explorative and collaborative activities of students to be tracked. (Helic et al, 2004a)
- Pedagogy driven learning - A teaching scenario or environment can be built where a tutor works with a group of learners in both synchronous and asynchronous mode, leading them to achieve a particular learning goal.
- Project-oriented learning - A controlled learning environment can be built to allow a group of learners works together on a project, e.g., a software engineering project (Helic et al, 2003)
- Adaptive discovery of personalized background reference knowledge. A reading room paradigm can be created enabling learners to chart out knowledge discovery process (Mödritscher et al, 2005)
- Annotations: Annotations allow the attachment of text segments, system or media objects or an URL to an object or material (Korica et al, 2005). It is possible to annotate any kind of material such as papers, parts of a digital library, other user contributions, etc.
- Active Documents: The idea of active documents presents a means of students learning in a collaborative question-answering environment. Active documents presents an innovative means to demonstrate student learning and at the same time, an effective way for an instructor to direct knowledge discovery.(Heinrich and Maurer, 2000)
- Visualisation as knowledge maps: The cluster of a document with documents containing similar concepts or ideas can be visualized via a knowledge map typical of knowledge management systems. A knowledge map with similar articles can be created and visualized. (Helic et al. 2004b)
- Workflow management and compliance checking capabilities. Learning can be visualized as a process flow of learning tasks. Non-compliance can then be automatically flagged by the system.

WBT-Master allows the copy-paste syndrome avoidance procedures to be integrated directly as part of the E-learning system. E-learning modules on effective copy-paste would be embedded to educate students on the rightful procedure of academic publishing (reading and writing). Apart from employing a plagiarism or copy-paste detection suite for summative conclusion of a breach of conduct, we propose the formative application of such tools for self-plagiarism checking and in cultivating constructive 'copy-paste skills'. For example, existing document similarity detection (as used in plagiarism detection tools) can be applied in conjunction with a learning scenario paradigm for facilitating students to master academic publishing. By consolidating the results from similarity search engines on local databases as well as the internet, a plagiarism detection tool can be applied to assist students to teaching them how and when to cite a referred publication.

The technological support to avoid the use of blatant copying by students can be achieved by imposing on the use of annotations, which avoids the need to duplicate, content. The use of annotations can be explored as a means of training students' use of the correct form of citations and referencing. By using annotations, a much simpler similarity checking system would suffice to overcome plagiarism to a large extent. Annotations and its sophisticated communicational and collaborative features plays an important role in the realization of a culture of Web-based reading and writing. As illustrated by the term 'ICARE'; copy-paste can be made more meaningful by students ability of absorbing concepts, consolidating and assimilating it before expressing with full understanding and conviction. A system should guide and allow students to become aware of the correct approach of reading, digesting and applying knowledge. At the same time, the platform should support creativity in their associational and expressive ability. The ecosystem proposed would allow an instructor to view and monitor the learning process of students, in observing and monitoring the rightful practice of 'copy-paste skills'. At the same time creative expressions of students can be pin-pointed, highlighted and recorded.

ICARE also provides a controlled environment in which the instructor is able to track the usage of reference materials by students. Such a controlled environment makes it much easier to curtail unethical practices and also promotes constructivist learning among students. Furthermore the user tracking and user activity-logging facilities can also be enforced on learners to read certain parts of a document before being allowed to annotate an article or ask questions about an article. (Helic et al, 2004a) Dynamic documents can then be established to remember the questions asked by students as well as to record the expert advice in answering them. Knowledge profiling can then be used to support the acquisition, structuring, and reuse of extracted expert knowledge. Student assessment can thus be enriched to a large extent by the multiple learning objects in ICARE.

An environment that closely monitors students' learning activities in the knowledge construction and collaboration can help the instructor to assess and guide students' ability to publish effectively. Process level support

can be achieved via the workflow management and compliance checking capabilities of systems (of Hyperwave). The system can be trained to recognize non-conforming patterns that will flag instructors.

Integrated visual tools assist in the management and display of information as a highly formalized, refined, process designed to apply to illustrate student learning and mastery of concepts. They can be used by the instructor to impart particular skills, to refine processes used for a defined task, or to organize information into a structured form. It can also be used by a student to express his understanding of concepts. Knowledge visualization tools can be applied as a form of assessment of students' incremental knowledge gain over a period of time. Learners can also be supported by means of personalized knowledge retrieval facilities. Such a tool may not only be effective in identifying potential infringements by students, it can also be used to aid students in mastery of useful skills.

We propose support mechanisms for the careful design and execution of assessments. The pedagogy driven learning together with the ability to define learning scenarios and rooms allow for highly personalized assessment design and curriculum development. As an example personalized reading lists can be created separately for different groups of readers. These lists can even be based on students' demonstrated learning of particular topics.

These functionalities can be further integrated with emerging Web2.0 services (Alexander, 2006) such as e-portfolios and recommendations systems to allow students to enhance their skills in valuing their published works. Incentive schemes can be tied to e-portfolios and student achievement, to emphasize student-learning achievements. Recommendation systems play an important role in the development of judgment of the goodness of contents to the task at hand. A combination of human and automated ranking of important topics, ideas, suggestions and contributions will be invaluable to provide valued, personalized contents.

Usage of ICARE ecosystem for E-Learning

As an example of the application of ICARE in a classroom, we propose the following illustration:

1. Students in a class are first asked to collaboratively construct a collective concept map for a domain of study.
2. Individual students are then required to form a personalized concept learning space
3. Subsequently, students are assigned selected reading material. An online copy of the reading material is placed in the reading room of each student (or a group of students)
4. Students are then required to identify key points by using the wizard in closed monitoring mode with all activities tracked by system. The highlighted text segments by students can be used to reflect their understanding. Both individual student learning and group learning can be highlighted.
5. The highlighted texts are then visualized for the instructor as annotations attached to the selected document. Statistical information is used to demonstrate student learning e.g, common mistakes made by student, misunderstanding of text, etc.
6. An instructor's comments can either be placed in personal spaces of students or public spaces for the attention of a whole class
7. Students are then requested to paraphrase the text selected, in the guided mode as well
8. A visualization of all student inputs is then made available for the instructor. Additional statistical information is presented to support student evaluation. By incorporating Hyperwave, non-compliance in student learning work-flows can also be visualized.
9. The next step involves a peer-learning mode, where student are requested to discuss the points selected by their peers in the brainstorming room. All points being discussed are referenced and the system links them together (for visualization). The instructor or facilitator then provides interactive feedback in the brainstorming room.
10. Students are then required to update their personal concept maps, with the knowledge gained in 9.
11. Statistics of popular concepts in knowledge-map, popularly selected key points, list of questions posed during brainstorming or during any other phase in the exercise is all presented to the classroom.
12. As the final task students then asked collaboratively construct a single concept map while continuing with discussions in the brainstorming rooms. All concepts in the knowledge map are uniquely identifiable as they are implemented using Knowledge cards. As such students are able to discuss the addition of particular concepts or place to links and types of links as well.

The illustration above indicates the power and potential of the proposed system, to serve as a futuristic E-Learning system that can be employed to address the various emerging concerns of E-Learning. The resulting blended learning environment discussed, provides a basis for futuristic E-Learning system. We are currently in the process of

extending and integrating the various tools at the Institute for Information Systems and Computer Media, Technological University Graz.

Conclusion

In this paper, we have adopted the stand that copy-paste need not be entirely considered a wrong doing. In order to address the concerns of Weber, students would then need to be educated and guided on the constructive use of copy-paste skills (as a learning mechanism). We have presented an academic ecosystem with technological support to comprehensively address the copy-paste-syndrome. By effectively addressing the copy-paste-syndrome many of the social problems that we are likely to face (arising from the degradation scientific quality and quality of life as described earlier) in future can be averted. Without the full institutional backing and commitment of academics however, a culture that withstands and compensates the prevalent copy-paste culture cannot be achieved.

We have proposed the use of a sophisticated blended learning system, together with carefully planned student assessments and the close monitoring of student learning to address the problem. Plagiarism and copy-paste syndrome avoidance mechanisms and procedures need to be integrated into the E-learning system and applied throughout the programme of study. E-learning modules integrated with a suite of copy-paste detection tools are required especially for the formative development of 'effective copy-paste skills'. A complete suite of copy-paste detection and avoidance tools will need to be established in all educational institutions.

References

Alexander, B. (2006). Web 2.0: A New Wave of Innovation for Teaching and Learning?, *EDUCAUSE Review*, Vol. 41, no. 2 (pp. 32–44).

Downes, S. (2006). E-learning 2.0, *ACM eLearn Magazine*.
<http://www.elearnmag.org/subpage.cfm?section=articles&article=29-1>

Dreher, H.V., Dreher, L.H., and McKaw, K. (1994). The Active Writing Project - small movements in the real world. *Proceedings of Asia Pacific Information Technology in Training and Education*, Brisbane, Australia.

Dreher, H., Krottmaier, H., Maurer, H. (2004). What we Expect from Digital Libraries, *Journal of Universal Computer Science*, Vol 10, no. 9 (pp. 1110-1122).

Dreher, H., Williams, R. (2006). Assisted Query Formulation Using Normalised Word Vector and Dynamic Ontological Filtering: Flexible Query Answering Systems. *Proceedings of 7th International Conference, FQAS 2006*, Milan, Italy, June 7-10 (pp. 282 –294).

Duff, A.H., Rogers, D.P., Harris, M.B. (2006). International engineering students - avoiding plagiarism through understanding the Western academic context of scholarship, *European Journal of Engineering Education*, Vol. 31, no. 6 (pp. 673-681).

Eissen, S., Stein, B. (2006). Intrinsic Plagiarism Detection. In: *Proceedings of the 28th European Conference on Information Retrieval; Lecture Notes in Computer Science*, vol. 3936, Springer Pub. Co. (pp. 565-569).

Harris, R., (2004) Anti-Plagiarism Strategies for Research Papers, November 17, 2004
<http://www.virtualsalt.com/antiplag.htm>

Heinrich, E., Maurer, H. (2000). Active Documents: Concept, Implementation and Applications. *Journal of Universal Computer Science* Vol. 6, no 12 (pp. 1197-1202)

Helic, D., Krottmaier, H., Maurer, H., and Scerbakov, N. (2003). Implementing Project-Based Learning in WBT Systems, *Proceedings of E-Learn 2003*, AACE, Charlottesville, USA (pp 2189-2196).

Helic, D., & Maurer, H., & Scerbakov, N. (2004a). Discussion Forums as Learning Resources in Web-Based Education, *Advanced Technology for Learning*, Vol. 1, no. 1 (pp. 8-15).

Helic, D., Maurer, H., Scerbakov, N. (2004b). Knowledge Transfer Processes in a Modern WBT System, In *Journal of Network and Computer Applications*, Vol. 27, no. 3 (pp. 163-190).

Kennedy, I. (2004). An assessment strategy to help forestall plagiarism problems. *Studies in Learning, Evaluation, Innovation and Development*: Vol. 1, no. 2 (pp. 1-8).
http://sleid.cqu.edu.au/viewissue.php?id=5#Refereed_Articles

Korica, P., Maurer, H., Scerbakov, N., (2005) Extending Annotations to Make them Truly Valuable Extending Annotations to Make them Truly Valuable, *Proceedings of E-Learn 2005*, AACE, Charlottesville (pp. 2149-2154).

Krottmaier, H., and Helic, D. (2002). More than Passive Reading: Interactive Features in Digital Libraries. *Proceedings of E-Learn 2002*, AACE, Charlottesville, USA (pp. 1734-1737).

Kulathuramaiyer, N., Maurer, H. (2007). Why is Fighting Plagiarism and IPR Violation Suddenly of Paramount Importance?. *Proceedings of International Conference on Knowledge Management*, World Scientific, Vienna, August 2007.

Liu, C., Chen, C., Han, J., and Yu, P.S. (2006). GPLAG: Detection of Software Plagiarism by Program Dependence Graph Analysis. *12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining 2006*, Philadelphia, USA (pp. 872-881).
<http://www.ews.uiuc.edu/~chaoliu/papers/kdd06liu.pdf>

Mödrtscher, F., García-Barríos, V.M., Maurer, H., (2005) The Use of a Dynamic Background Library within the Scope of adaptive e-Learning. *Proceedings of E-Learn 2005*, AACE, Charlottesville (pp. 3045-3052).

Maurer, H., Kappe, F., Zaka, B. (2006). Plagiarism- a Survey. *Journal of Universal Computer Science*, Vol.12, no. 8 (pp. 1050-1084).

Maurer, H., Zaka, B., (2007) Plagiarism - a Problem and How to Fight It. *Proceedings of Ed-Media 2007*, AACE, Vancouver (pp. 4451-4458).
http://www.iicm.tugraz.at/iicm_papers/plagiarism_ED-MEDIA.doc

Midolo, J. and Scott, S. (2003). Teach Them to Copy and Paste: Approaching Plagiarism in the Digital Age, *Curriculum Materials Information Services*, October 2003
<http://www.det.wa.edu.au/education/cmis/eval/curriculum/copyright/islandjourneys/documents/paper.pdf>

Ong, S.C., Kulathuramaiyer, N., Yeo, A.W. (2006). Automatic Discovery of Concepts from Text, *Proceedings of the IEEE/ACM/WIC Conference on Web Intelligence 2006* (pp.1046-1049).

O'Reilly, T. (2006). What Is Web 2.0 Design Patterns and Business Models for the Next Generation of Software. <http://www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html>, accessed April 17, 2006.

Paper Mills (2007). <http://www.coastal.edu/library/presentations/mills2.html>

Plagiarism.org: Statistics (2007). http://www.plagiarism.org/plagiarism_stats.html

Pitner, T., Drážil, P. (2006). An E-learning 2.0 Environment – Principles, Technology and Prototype, *Proceedings of I-Know 06*, Graz, 2006

Standing, L., Gorassini, D. (1986). An Evaluation of the Cloze Procedure as a Test for Plagiarism. *Teaching of Psychology*, Vol. 13, No. 3 (pp. 130-132).

Stanley, M. (2005). Web site, <http://www.morganstanley.com/institutional/techresearch/gsb112005.html>,
WBT Master White Paper (2006).
<http://www.coronet.icm.edu>

Weber, S. (2006). Das Google-Copy-Paste-Syndrom. *Wie Netzplagiate Ausbildung und Wissen gefährden*, Heise, Hannover.