Micromedia and Corporate Learning
Proceedings of the 3rd International Microlearning 2007 Conference

Editors:
Martin Lindner, Peter A. Bruck

Printed with support of
Vizerektorat für Forschung – Universität Innsbruck
Austrian Research Centers GmbH Research Studios Austria
# Table of Content

## Forewords

Peter A. Bruck  
Welcome and Introduction to “Micromedia and Corporate Learning” .......................... 1  
Martina A. Roth  
Welcome to Microlearning 2007 ................................................................. 4  
Tilmann Märk  
Preparing for the Digital Knowledge Society at the University of Innsbruck ............. 6

## Introductory Note

Martina A. Roth  
Global Economy & Microlearning:  
The Intel® Education Initiative for the 21st Century ........................................... 8  
David Smith  
Messiness, Education, Mobility ................................................................. 17  
Teemu Leinonen  
Building the Culture of (e-)Learning in Microcontent Environments .................... 24  
Ajit Jaokoar  
Mobile Web 2.0: Designing Open Gardens for Tomorrow’s  
Microcontent-based Education ................................................................. 34  
Stephanie Rieger  
Mobile Micro-learning ................................................................. 43  
Martin Lindner  
What Is Microlearning? ................................................................. 52
Papers

Norm Friesen
The Microlearning Agenda in the Age of Educational Media ....................... 63

Jana Herwig
The Dark Side of the Force: The Issue of Microplagiarism in Microlearning ........ 79

Andreas Schmidt
Microlearning and the Knowledge Maturing Process:

Junichi Azuma, Hermann Maurer
From Emoticon to Universal Symbolic Signs:
Can Written Language Survive in Cyberspace? ........................................ 106

Andreas Walbert, Mirko Ross
Small Portions of Knowledge —
Training for highly educated service engineers ........................................ 123

Linda Johanna Castañeda
On-line learning in on-site institutions: The challenge to
try new pedagogical models in higher education ..................................... 132

Markus Appel
Some ideas on learning through fictional (micro-) narratives ...................... 142

Thomas Sporer, Ulrich Fahrner, Ruben Schulze-Fröhlich, Tino Jahnke, Frank Vohle
Transforming traditional classroom lectures into
interactive digital media experiences ...................................................... 150

Taiga Brahm
The Development of Team Competencies through Social Software .............. 158
# Table of Content

## Best Practice Reports

Abdullah Lin  
Learning and Competency Development at Silterra Malaysia Sdn. Bhd. –  
A Practical Approach Towards Developing People in Semiconductor Wafer Fabrication ................................................................. 169  
Bachok M. Taib, Waidah Ismail  
Developing CGPA planner by using PDA for Islamic Science University of Malaysia (USIM) .......................................................... 179  
Azza H. Amer, Mahmoud M. El-Khouly, S.M. Khaled  
Web-based Graduate Diploma in Computer Sciences – A Case Study – ........ 187  
Tanja Jadin, Paul Maderthaner  
Laptops as a resource for microlearning scenarios?  
Findings from a survey on the implementation of laptop classrooms in Austrian upper secondary education ............................................. 197  
Christian Petter, Kathrin Helling  
365 Mini-Lessons to Learn English Module development for the project EIT – English for IT ................................................................. 211  
Yu-Hsiu Chou, Li-Wen Wang  
Web-based Instruction – Two Empirical Classes ........................................ 214  
Martin Ebner, Walther Nagler, Anna Saranti  
TU Graz goes Podcast ........................................................................... 221  
Pedro P. Sanchez-Villalon, Manuel Ortega, Asuncion Sanchez-Villalon  
AIOLE and AWLA application to e-Learning 2.0:  
an ePortfolio with an eDossier ............................................................... 234  

## Short Biographies (Key Speakers) .................................................. 243
Abstract: Based on the analysis of authentic examples of plagiarism in student assignments, this article proposes the term 'microplagiarism' to describe a new kind of plagiarism which uses relatively short sections of arbitrary sources and combines them to form a bigger, seemingly unified text. The author examines to which extent learning through plagiarizing may be an effective strategy in some areas of language learning and discusses the difficulties in separating micro-learning from microplagiarism. The current gap between the digital and the academic sphere and their methods of circulating and perpetuating knowledge is identified as a cause for the increase of plagiarism. The Semantic Web is discussed as a possible solution to bridge the gap between the two spheres and to enable a better cooperation between humans and computers in the academic world.

1 New Plagiarisms

The idea to propose a paper about microplagiarism grew out of the confrontation with a situation in which more than 25% of learners in a language course at university level had attempted mild to severe plagiarism in their final written assignment. A cursory examination showed that the quality of plagiarism found in these assignments had little to do with what I had believed to be typical characteristics of academic plagiarism. One of my, probably naïve, prior assumptions had been that the main intention of academic plagiarism was to conceal the plagiarist’s lack of his or her own research and original thought by copying someone else’s work and thus claiming credit for it. In the examples given, the passages that were copied were often simply too short, consisting of no more than one or two sentences, in order to be considered a thought in their own right. Furthermore, the plagiarized segments turned out to be deeply generic: the function they fulfilled in the text could have been fulfilled by many similar text segments that can be found on the World Wide Web, which was indeed the place where the students had found their sources. Criteria for assessing this new type of plagiarism were required urgently, along with a strategy to dissuade these students, who were in their second year of study at the time, from any further attempts at plagiarism. In this paper, I would like to
propose the term “microplagiarism” to refer to a new kind of plagiarism that uses snippets and clips from various arbitrary sources on the web to form a new text. In the example given, students used these snippets to draft a text in a foreign language which, for reasons that need to be discussed in more detail, may have contributed to them not perceiving of their copying as plagiarism.

2 Analysis of the Plagiarized Texts

The task given to students was to write a comparative film review in which they discuss and compare two English language feature films. This assignment was their final written assignment in a CLIL (Content and Language Integrated Learning) course entitled “Introduction to Film Studies”. Already in the definition of the task, my intention had been to build in an anti-plagiarism mechanism, in this case consisting of the requirement to compare two films. Most film reviews discuss only one film at a time. My assumption therefore was that a film review meeting the requirements of this task would not be readily available on the World Wide Web. Regarding the structure and content of the film review, students were asked to ensure that their film review contained the following elements, based on a proposal by David Bordwell, author of several seminal books on film studies and film history:

- **Introduction**: introduction of films, beginning with an attention-grabbing opening
- **Brief plot synopsis**: introduction of the key conflict without giving away the story ending
- **Discussion**: balanced discussion and comparison of the films mentioning specific elements, describing these quickly and vividly
- **Conclusion**: summary of discussion and judgment of quality, concluding with a striking sentence

---

As Figure 1 suggests, the biggest part in a balanced film review based on Bordwell’s formula would be consumed by the discussion (paragraphs 3, 6), followed by plot synopsis (4, 5), conclusion (7) and introduction (1, 2). Each of these elements fulfils a particular function within the film review and each requires a specific language and register.

What is crucial for this discussion is that these components also vary in the degree in which they, theoretically, could be plagiarized without perceptibly disrupting the flow of reading. An attention-grabbing opening, for instance, could be copied from another film review covering one of the discussed films. Brief plot summaries are available on many pages on the web and would naturally be set off in terms of language and register from their surrounding parts. The discussion, by contrast, is more difficult to plagiarize, as it

---

should ideally offer a comparative discussion of specific elements of both films (e.g. of cinematography, storytelling, set design, lighting or acting). The same applies to the conclusion which should correspond to and take up ideas from the previous discussion. A text that would meet exactly these requirements would be difficult to find.

Six of the plagiarized film reviews were examined in more detail. Three of these reviews were written by individual students, the other three by teams of two students. Among those who worked in a team, there was one team in which the one student was not aware that the contributions made by his or her partner were plagiarized. In the other two teams, the plagiarizing seemed to have occurred with their mutual consent.

2.1 Criteria for the identification of plagiarism

In order to identify plagiarism, both random samples and ‘suspicious looking’ samples of text were extracted manually. In the next step, a search query was performed using Google.com as a search engine. ‘Suspicious’ in this context means ‘as judged by the instructor of the course’ and based on her past experiences with a similar course and with the quality of writing that students at this level are normally able to produce.

A first rule of thumb for the ‘human-based’ identification of plagiarism is: ‘The more idiomatic, the more suspicious’. A phrase like the following, occurring in a student assignment, is likely to be a candidate for closer examination:

Wallace (voiced by Peter Sallis) is a cheese-chomping tinkerer of inventions of questionable merit, and Gromit, his silently stoic canine protector who is more nanny and housekeeper than pet.³

Another useful indicator is an apparent contrast in grammatical accuracy between successive phrases or sentences as in the following example, where the degree of accuracy decreases as the paragraph proceeds.

³ Original: “Graduating from their award-winning comic shorts to make their feature-film debut, Wallace (voiced by Peter Sallis), a cheese-chomping tinkerer of inventions of questionable merit, and Gromit, his silently stoic canine protector who is more nanny and housekeeper than pet, find their neighbors gripped with veggie-mania as the annual Giant Vegetable Competition nears.”
When Jack meets the ghost of Mr. Grady, the former custodian of the hotel who murdered his wife and his two daughters, things begin to get really nasty. Comparing to “Ringu” the scenes are resolved by more simple shots and camera panning then tracking shots to keep the attraction on the highest. The name of the movie “Ringu” is given by the ring of the telephone call and like “The Shining” is “Ringu” a movie version of a novel, both fascinating how their way of style is converted. Very exciting to watch both movies whether they have such a different story, but you will see where the differences are.4

4 Original: “When Jack meets the ghost of Mr. Grady, the former custodian of the hotel who murdered his wife and his two daughters, things begin to get really nasty.”
2.2 Plagiarized components

The analysis of the assignments showed that, as predicted, plot summaries were the most likely to be plagiarized, followed by introductions (figure 2): All six synopses and two introductions contained plagiarized passages. Discussions and conclusions were mostly original pieces of writing. Also, the amount of plagiarized writing was considerably higher in plot summaries and introductions than in conclusions or discussions.
What could also be observed – in relation to other, plagiarism-free reviews – was a certain tendency of the plot summaries to dominate the film reviews: While a plot synopsis would normally take up a quarter or at most a third of a film review, it consumed a third to a half of some of the plagiarized reviews (Figure 2, reviews B and F). At the same time the discussion, which ought to be the core of the review, was considerably reduced in these assignments. One possible explanation for this phenomenon could be that the students in question were hesitant to write a text of their own that – in comparison to the copied passages – could be faulty with errors or lack elegance and idiomatic expressions. But there were also other examples (in particular reviews C and E) where the writers offered a balanced discussion of the films, but nevertheless resorted to plagiarism when it came to writing the plot summaries.

The rule one can extract from these results is that a text component is more likely to be plagiarized the more generic it is, the more similar text segments fulfilling a similar function can be found somewhere on the World Wide Web. “Why should I write it again if it has already been written somewhere else?” was a response along the same vein which I received from a student in another group when we discussed the rising issue of plagiarism. Two things are noteworthy here: Firstly that it did not occur to the student that one could also reference a web resource instead of simply copying it. Secondly, it is very telling that he said “somewhere else” and not “by someone else”, as though the role of the author were irrelevant in this respect. I am going to get back to this question of authorship and the transformation of its function in the era of web-based, computer-mediated communication in paragraph 4.

2.3 Plagiarized sources

The sources the students had used in their reviews were then examined more closely. The first interesting result was that all students had used more than one source (Figure 3). Some of them (D and in particular B) had made a notable effort of combining the different sources to form a single paragraph, or even a single sentence. Another student (F) had initially tried to rephrase and adapt the original source as much as possible, in particular by using synonyms or by changing the word order and often at the expense of grammatical accuracy. As the text proceeded, the student decided to use a different strategy and to imitate only every fourth or fifth phrase of the original text, thus leaving out significant plot points, but still ending up with a synopsis too detailed for the task. The discussion, as a consequence, turned out relatively short and superficial. Another team (C) had produced a particularly awkward mistake by not only copying the plot synopsis from Wikipedia, but also a related image caption:
[...] Wetmore ends up as an inmate at the very asylum he was to have managed. John Coffey being escorted to his execution by Edgecombe and Brutus Howell. Due to the emotional performance he showed in the film, Michael Clarke Duncan was nominated to the Academy Award's Best Performance by an Actor in a Supporting Role. Notwithstanding Coffey's incredible abilities and the wrongness of his conviction, he ends up being executed [...].

The most popular source used by students was Wikipedia with five different copied sources in total, followed by several film fanzines and blogs (four sources), commercial music and entertainment platforms, online editions of print magazines and newspapers (three sources each), the Internet Movie Database (two sources), a DVD shop and Amnesty International (one source each). Considering the arbitrariness of the texts, one can assume that students used a search engine like Google, rather than directories or databases, to retrieve their sources.

Judging from this arbitrary selection and the way in which the sources were combined, one can furthermore assume that students were not too worried about their writing being eclectic or incoherent, at least not about it being incoherent as a result of the arbitrariness of these sources. What could not be identified – and what would have corresponded more with my ‘conventional’ notion of academic plagiarism – was that students used an undisclosed source to inform their argument or approach at large. Instead of relying on these sources as a substitute for their own research, most students...
seemed to use them as a repository for prefabricated stretches of content and language that would help them achieve their goal of writing a film review.

As I have suggested earlier, students probably did not perceive of their actions as plagiarism. We must also note here that the writing of a plot synopsis is an area in film studies where there is little room for interpretation or original thought, as the content of a plot summary is already restricted by the number of events explicitly presented by the film. In that sense, a plot synopsis is not an autonomous, original piece of writing, but one of the film’s paratexts, a “zone between text and off-text” and as such “at the service of a better reception for the text and a more pertinent reading of it.” The task of writing their own plot synopsis did apparently not make sense to the students, in particular not in the light of the abundance of similar texts which have already been written and which are available on the World Wide Web. Looking at the plot summary as a paratext, or service text, might also explain why nobody decided to make a reference to the source they had used – who would, for instance, make a reference to a table of contents or the imprint of a book, to draw an analogy to other paratexts? At this point I would like to devote some attention to the thought that copying, in the sense of imitating, can also be a useful strategy for learning, in particular for language learning. The difficulty, however, is to draw the line between microlearning and microplagiarism.

3 Microlearning vs. Microplagiarism

All of us who have learned a foreign language at one point in our lives have to a certain extent used imitation as a strategy for language learning. When we are uncertain about the use of a particular word, we will look it up in a dictionary and adjust our use of it to the recommended use of the dictionary. And while most of us would probably agree that this is and always was a form of microlearning avant la lettre, predating the digital age, we would certainly refrain from calling this strategy plagiarism.

A language consists of these two main components: the “whole body of words” (lexis or vocabulary) and the “methods of combination of words” (grammar). Oral or written texts

---

6 I am applying the formalist definition of plot – everything that is explicitly presented on the screen (including credits and musical score) – in opposition to the definition of story – the sum total of explicitly presented events (excluding credits and music) and inferred events presumed by the viewer. Cf. David Bordwell (2003), Film Art: An Introduction, Seventh edition, McGraw-Hill, pp. 70-72.


are considered a manifestation of these words and rules, but the words and rules themselves are not owned by anyone. For instance, nobody can claim the copyright for the English language. But at which level of combinations can one draw the line between words and rules and concrete, copyrightable manifestations?

In his definition of the Lexical Approach,\(^9\) Michael Lewis proposes an extended concept of lexis, maintaining that “language consists of grammaticalized lexis, not lexicalized grammar.”\(^10\) The Lexical Approach shifts the attention away from grammar and to the importance of set phrases in language learning, such as ‘if I were you’, ‘out of my mind’, ‘strong accent’, or ‘brings good luck’, referred to by different and overlapping terms including “prefabricated phrases”, “lexical phrases”, “formulaic language”, “frozen and semi-frozen phrases”, “lexical chunks” and “collocations”.\(^11\)

Going back to our plagiarized film reviews, the question arises whether it wouldn’t be justifiable to look at them differently and to consider these reviews as awkward attempt to harvest the chunks and collocations found in other plot synopses. The following juxtaposition of plagiarized passage and original source shows that the writers (review C) made at least an attempt at editing the source:

\begin{quote}
Sam Lowry (played by Jonathan Pryce), is a low-level technician who regularly tries to break out of his life in this miserable world by daydreaming and fantasizing. When a bug gets in the system, an innocent man is killed and Sam reexamines what he wants out of life. He decides to fight the totalitarian system in his search for freedom and the woman he loves.

“Twelve Monkeys” is also a future vision by Gilliam. The story opens in the wintry wasteland of the year 2035, where a virus has killed 99% of the human population, forcing the survivors to live beneath the earth’s surface.
\end{quote}

The original sources:

\(^9\) Michael Lewis (1993), The Lexical Approach, Hove Language Teaching Publications.
\(^10\) Lewis 1993:9.
BRAZIL is Terry Gilliam’s masterpiece. Cowritten by Gilliam, playwright Tom Stoppard, and Charles McKeown, the cult-favorite film is set in a futuristic society laden with red tape and bureaucracy. When a bug (literally) gets in the system, an innocent man is killed, leading mild-mannered Sam Lowry (an excellent Jonathan Pryce) to reexamine what he wants out of life. He decides to fight the totalitarian system in his search for freedom--and the woman he loves.\textsuperscript{12}

\textit{12} Monkeys combines intricate, intelligent storytelling with the uniquely imaginative vision of director Terry Gilliam. The story opens in the wintry wasteland of the year 2035, where a virulent plague has forced humans to live in a squalid, oppressively regimented underground.\textsuperscript{13}

The aim of proposing this perspective is not to excuse plagiarism and certainly not to inappropriately celebrate it as a successful example of microlearning. But it points us to the difficulty of distinguishing microlearning from microplagiarism and to the arbitrariness of our decision-making, depending on which perspective we decide to adopt. The ability to use foreign language texts as a repository for pre-fabricated phrases is a vital skill, allowing foreign language students to learn independently. The web in particular lends itself to this purpose because it is highly searchable and provides texts for virtually all areas of human interest. It can thus be considered the largest ‘corpus’, i.e. collection of samples of a language, currently available.

\textbf{3.1 The moral dimension}

Using the web for this purpose, however, immediately raises moral questions that need to be addressed in the same context. I would like to propose that, although the types of content used for microlearning and for microplagiarism have many traits in common, it is their diverging intent that changes the situation drastically. A decisive criterion for all content intended for microlearning is that single items of microcontent must be addressable: It must be possible to “find the item back”.\textsuperscript{14} Microplagiarism, however, seeks to evade addressability. It does not want to be identifiable or draw the reader’s attention to it as a separate form or unit. It pretends to be part of a bigger unit, unified by a common author, and it passes as such for as long as it is not conspicuous, for as long as the original source and the original authors cannot be retrieved.

\textsuperscript{12} \url{http://www.cduniverse.com/productinfo.asp?pid=1282315&style=movie&cart=457851446&BAB=E} (accessed Jan 31, 2007)
\textsuperscript{13} \url{http://www.imdb.com/title/tt0114746/amazon} (accessed Jan 31, 2007)
\textsuperscript{14} Arnaud Leene (2006), ‘Microcontent is everywhere’, in: Theo Hug, Martin Lindner, Peter A. Bruck (eds.), Micromedia & e-Learning 2.0: Gaining the Big Picture, Innsbruck University Press, p. 25.
In order not to be exposed, the plagiarist needs to cover the same tracks that he or she followed when searching for suitable sources. The only way of doing that is by editing, rephrasing and adjusting the original source to such an extent that the newly created content can no longer be used as a resource locator. For a language teacher, this effort would be sufficient to acknowledge the product as original and the learning process as successful. Regarding our film reviews, of course, we cannot speak of a successful learning process in that sense. It seems ironic that the plagiarism was exposed using the same instrument that the students most likely used in order to retrieve their sources, a crawler-based search engine. The method of exposure itself, which in most cases pointed to only one possible source on the web, shows that the plagiarized phrases were clearly not examples of pre-fabricated phrases and not shared communally by the users of the language. These were cases of intellectual property theft. But are students aware of that?

Awareness of the practical consequences – ‘If you plagiarize, you will fail’ – is easily instilled in students, yet many of them seem to be ignorant of the moral dimension. “How do you always find out so quickly?” complained a student from another course who, together with his ‘co-writer’, had failed after handing in an assignment that qualified as straight-forward ‘macro-plagiarism’: They had purchased a student paper for € 4.99 from a website that advertised itself as a ‘scholarly publishing house’ and had made just a few superficial changes before submitting it as their own. He admitted that they had even ‘tested’ it with Google beforehand to see whether it could be identified as plagiarism. This example shows that some students have already adapted the formula ‘If you plagiarize, you fail’ into ‘If they find out that you plagiarize, you fail’. Other students seem to assume that anti-plagiarism regulations are just an eccentric invention of a particular university or lecturer, but not a general rule of academic writing. This becomes apparent in the following email from a student who had microplagiarized a jointly written film review without telling her team partner:

As you have noticed, a part of the review was copied from the internet. Actually my part (Wallace & Gromit) was copied. I worked together with Claudia (name edited; J.H.) and unfortunately [sic!] a misunderstanding occurred. Claudia wrote the review for Corpse Bride and I wrote the review for Wallace & Gromit. I made the mistake and did not tell her that I had some parts from the internet, because I thought when we put the two reviews together it would be our work.

---

and not plagiarism. I want to apologise for that. I'm new at the FH and didn’t know all the rules. But I am responsible for that situation.

The second example powerfully illustrates the naïveté and unawareness of some micro-plagiarists towards the ethical implications of their actions. It emphasizes that we as educators, in order to allow students to get the most out of microlearning, must also help them to anticipate the pitfalls of microplagiarism and show them how to deal with these.

4 The Transformation of Knowledge

So far we have not explicitly discussed the question of referencing, although this would probably be a convenient way of cutting a long story short: Why did the students in question not simply reference their sources properly? And is this truly their failure or the failure of the lecturer who did not teach them how to write a proper citation?

In my view, this is not the issue actually at stake here. Using and responding to circulating ideas within an academic discipline is a prerequisite for a functioning academic discourse, and giving the original author credit is not only a matter of respect or decency, but also an absolute necessity to ensure the continuation of the discipline in question. But outside of the academic disciplines, citations and referencing are rarely demanded from a writer. Journalists do not always reveal or specify their sources, whether this is to protect their sources or to not disrupt the flow of reading, and “cut-and-paste journalism” has become a plague to the craft of writing.16

Instead, I would like to argue that the main cause and issue is to be seen in the transformation of knowledge and subsequently of content in the age of computer and web-based communication and publishing. This transformation includes a different perception of intellectual property and, as a result, also a different perception of plagiarism and the dimension of the transgression it represents in the academic arena.

4.1 The Natural Law of the Digital World

The question of authorship and of the transformation that it has undergone in recent years is crucial to the explanation of the emergence of microplagiarism. Authorship in

---

contemporary Western civilization, as Michel Foucault has shown, is above all a discourse of appropriation: Texts are the objects that are governed by this discourse whose system of ownership and copyright rules was established in the later eighteenth and early nineteenth century. These copyright rules, however, have been dramatically challenged by the digitization of their objects.

To begin with, by virtue of the system of online distribution, any reader of a document published on the world wide web is, pragmatically speaking, automatically the owner of the document: In order to view an online file, a copy has to be created which is then stored on the user’s computer, mostly temporarily, but possibly permanently. Although the Internet and World Wide Web are often compared to libraries or archives, this situation is fundamentally different from the conventional library set-up where the user is not automatically entitled to a copy of the book or medium he or she takes out on loan.

In response to this, various ‘digital rights management solutions’ – effectively copy prevention solutions – have been developed which seek to control the user’s access to a digital file – for instance by restricting the user’s rights to viewing, while preventing editing or printing, or by restricting the number of times that a file can be accessed. None of these solutions, however, seem to be effective in a wired, digital world: Once a file has been hacked, de-restricted versions will come into circulation, with the hacker’s skill encapsulated in code and handed down to all subsequent users. “This is a natural law of the digital world, and makes copying on the Internet different from copying Rolex watches or Louis Vuitton luggage,” internet security expert Bruce Schneier argues. “If you have a digital file – text, music, video, or whatever – you can make as many copies of that file as you want, do whatever you want with the copies.”

One of the things you can do with digital files is to copy them, paste them, tear them apart or edit them; “Find it Rip it Mix it Share it. Come and get it,” is a typical rallying cry among the many parties that warmly welcome this phenomenon as ‘digital creativity’.

---

20 Slogan of the Creative Archive License Group, a joint effort of the BBC, the British Film Institute, Channel 4 and the Open University. http://creativearchive.bbc.co.uk/ (accessed Jan 28, 2007)
It is a technological and pragmatic rationality that supersedes conventional copyright regulations – if copying is possible at a few strokes of the cursor and clicks of the button, how can it be illegal or immoral? And isn’t it schizophrenic that the same phenomenon is embraced if it is part of a prestigious project, but punished if it occurs in a student assignment? It is ultimately a dilemma of the contradictory rules of the material and the digital world that we are facing – whether we are looking at piracy or plagiarism. The Semantic Web, even if still mainly a vision, might be able to point a way out of this dilemma – even if it might not be able to address the issue of autonomous learning.

4.2 The Semantic Web

The Semantic Web, according to the original definition by Tim Berners-Lee, James Hendler and Ora Lassila, is not “a separate web, but an extension of the given one, on which information is given well-defined meaning, better enabling computers and people to work in cooperation.” The Semantic Web aims to enrich the documents on the World Wide Web – i.e. information that only humans are able to make sense of – with data, i.e. with information that machines are able to read, to process and to generate meaning from. Ivan Herman describes the practical obstacles currently presented to this form of opening up “knowledge and working of humankind to meaningful analysis by software agents,” by the impossibility of universal data exchange in his introduction to the Semantic Web on the W3C (World Wide Web Consortium) webpage:

The Semantic Web is a web of data. There is lots of data we all use every day, and its not part of the web. I can see my bank statements on the web, and my photographs, and I can see my appointments in a calendar. But can I see my photos in a calendar to see what I was doing when I took them? Can I see bank statement lines in a calendar? Why not? Because we don’t have a web of data. Because data is controlled by applications, and each application keeps it to itself.

How can such a web of data be achieved? How can computers, independent from applications, be taught to retrieve and combine data that are about the same things, in particular the same real world objects? In order to realize the Semantic web, a model

---

that regulates the representation of knowledge on the web is needed. The model recommended by the W3C is called RDF (Resource Description Network).\textsuperscript{24}

RDF is a method for modeling information in two dimensions: On the one hand, it provides a global identification system using so-called URIs: “Associating a URI with a resource means that anyone can link to it, refer to it, or retrieve a representation of it.”\textsuperscript{25} RDF/XML on the other hand allows the definition of common conceptualizations of knowledge within a specific domain, so-called ontologies. An ontology is “a data model that represents a set of concepts within a domain and the relationships between those concepts.”\textsuperscript{26} One example of an ontology is Dublin Core, a model that conceptualizes information related to documents and publishing and that can be used to represent bibliographic information. It was first defined in 1995 and currently includes 22 metadata elements (e.g. Title, Creator, Subject, Description), 33 element refinements (e.g. isPartOf, isReferencedBy) and 12 domain-specific vocabulary terms (e.g. Image, Text, Collection), which are documented in the Dublin Core Metadata Registry.\textsuperscript{27} Using Dublin Core, it is possible to describe documents in a way that can autonomously be used, reused and integrated by computers. A machine-readable, Dublin Core-based ‘Title’ description of the Registry’s own website would look like this in code:

```xml
<rdf:RDF>
  <rdf:Description rdf:about="http://dublincore.org/dcregistry/">
    <dc:title xml:lang="en-US">The Dublin Core Metadata Registry</dc:title>
  </rdf:Description>
</rdf:RDF>
```

Not all knowledge domains, of course, can be conceptualized with similar ease. Dublin Core is an example of a so-called shallow ontology that “comprise[s] relatively few

\textsuperscript{24} Further information for the interested reader to follow up at her own pace: RDF (http://en.wikipedia.org/wiki/Resource_Description_Framework) is based on two technologies similar to HTML (a mark-up language that describes the format of websites) and the global address system that is used to generate unique web addresses (URLs), but two much more powerful ones: XML or eXtensible Mark-up Language (http://www.w3.org/XML/) and URIs or Uniform Resource Identifiers, (http://www.w3.org/Addressing/#background, all accessed Apr 3, 2007).


\textsuperscript{26} http://en.wikipedia.org/wiki/Ontology_%28computer_science%29 (Apr 3, 2007)

\textsuperscript{27} http://dublincore.org/dcregistry/ (accessed Apr 3, 2007)
unchanging terms that organize very large amounts of data.” 28 Deep ontologies, by contrast, are “often those encountered in science and engineering, where considerable efforts go into building and developing the conceptualization.” 29 Any developed ontology will continue to require maintenance and revision by a devoted, small community of experts in the application domain. But thanks to the aforementioned capability of code to encapsulate skill and knowledge, a new concept can be passed on, from experts to software agents, and from agent to agent, and eventually the ordinary users will be able to reap the benefits of the Semantic Web. In that sense, it will not merely be “the tool for conducting individual tasks”, but “if properly designed, the Semantic Web can assist the evolution of human knowledge as a whole.” 30 The question this article is concerned with, however, is how the Semantic Web can assist learners and educators to make plagiarism redundant in the near future.

5 Conclusion: The Future Redundancy of Plagiarism

I have argued above that the emergence of microplagiarism and of the increase of plagiarism in student assignments in general have to be assigned to the contradictions between the rules of the academic and the digital world. If the promise of the Semantic Web comes true, as was suggested in the original article from 2001, it will “better [enable] computers and people to work in cooperation.” 31 The benefits this holds for the world of academia are indeed alluring: Individuals doing research online could harvest all metadata associated with a document using Semantic Web browsers. This information – and for our purpose: in particular bibliographic information, such as provided by Dublin Core – could be reused to generate references and reference lists. Anybody who copies from a Semantic Web compliant website would not merely copy words, but automatically import all the information and meaning that she or he – in order to avoid plagiarism – was required to extract manually in the past. Writing essays and articles that meet the requirements of different style guides would be an easy task for everyone. If this scenario were to come true, the mere possibility of plagiarism would cease to exist.

Of course this would change the customs and traditions of the academic world fundamentally and would not be embraced by everyone. Certainly for a transitional period, and probably much longer, some educators might insist that students continue to

write the reference lists manually, just as there are some today that insist that students refrain from using calculators. The propensity of media technology to serve as an extension to human cognitive abilities has always been greeted with skepticism. One can find a great deal of the criticism that some students’ use of web resources attracts, preempted in Plato’s *Phaedrus’* conclusion about writing, one of the first media technologies to fundamentally transform human knowledge and discourse:

*Once any account has been written down, you find it all over the place, hobnobbing with completely inappropriate people no less than with those who understand it, and completely failing to know who it should and shouldn’t talk to.*

The opportunities for the circulation of content and for informal learning have never been greater than now, in the web-supported knowledge society. Microplagiarism draws our attention to the ‘Dark Side of the Force’ of microlearning, to the flip side of a socio-technological constellation in which the pragmatic opportunities for copying, using and editing someone else’s work are greater than ever. The problem of microplagiarism will only cease to exist, if we manage to close the gap between the technological and the academic sphere and their diverging rules for the circulation and continuation of information. If ‘copy and paste’ technically meant to not only copy alphanumeric characters, but also the semantic relations of a particular item of microcontent, then we would know that the gap has been closed. The Semantic Web, the first stirrings of which are already visible in the development of ontologies and Semantic Browser extensions like Piggy-Bank,

*33* will be able to solve this problem and bridge the gap between human knowledge and computer data. Admittedly, the Semantic Web solution does not address the problem of learning and effective learning strategies, but it can help to raise student awareness of plagiarism and of ways to avoid it.

For the time being, the best strategy for dealing with the situation seems to be in setting students tasks that rule out or reduce the possibility of plagiarism in general.

---

33 http://simile.mit.edu/wiki/Piggy_Bank (accessed Apr 4, 2007)
34 I would like to thank my dear colleague Dr. Colin Gregory-Moores for assisting me with last-minute revisions and sharing his ideas regarding the Semantic Web.
6 References


