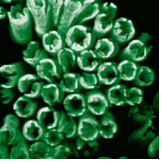

6 IS GOOGLE SUITABLE FOR DETECTING PLAGIARISM?

PLAGIARISM IS GENERALLY PERCEIVED TO BE ON the increase in UK HE institutions because of the ease with which electronic information can now be widely accessed, copied and shared. There is a resultant increased need for HE institutions to have transparent procedures in place that are effective at deterring and detecting plagiarism in order to maintain the quality of their educational provision and output. Many HE staff in the UK routinely use Google (*http://www.google.com*) to detect plagiarism because it is readily available and often quite accurate in returning the addresses of plagiarised websites. The evidence then presented is deemed suitable to show that a plagiarism offence has been committed. But is using a search engine such as Google a suitable way to detect plagiarism offences and are there better alternatives?

In this article, I briefly compare Google with two other software products and services that can be used to detect plagiarism in the hope that such a comparison will improve current methods of avoiding and detecting plagiarism among HE students. Google is used here as an example of a search engine that is widely used by academics. The application of it for detecting plagiarism applies equally to most of the other popular search engines that are available.

WHAT DOES GOOGLE DO?

Well, depending roughly on how the key words are placed in the search box, Google simply matches a (limited) string of words with those in web pages. Depending on how key words are entered into the search box, the returns are ranked according to whether the whole string, or words present in it, are found in a web page and how many other pages link to it. This method often produces quick positive results, with sections of text in documents being matched to that in websites. But Google does not provide all the answers:

- what is the nature and extent of text matching that occurs in a document? Google only matches strings of words entered in the search box with those found on web pages held in its database. They do not determine the proportion of an assignment that matches pages on the web nor how many sites were used in the whole document. Of course, we could look further in the document for signs of other plagiarised pieces, and key in phrases from it, but that is time consuming and not suitable for routine screening of work for detecting plagiarism;
- is the internet search performed by Google really comprehensive? Although search engines usually search websites quite effectively, they do not search access controlled resources, such as databases with quality resources, or "cheat sites" where students can buy essays;
- is the use of Google for detecting plagiarism consistent and fair? I have heard some staff claim that they "can spot a plagiarised piece of work a mile away and in any case they are often submitted together on the same pile". But is this statement always true when scores of essays are being marked? It may be possible to unconsciously remember sections from a few assignments but is it fair to penalize those that are unlucky enough to be caught almost by accident? Should we not routinely screen all submissions? and
- does it detect collusion? Collusion is a form of plagiarism that occurs when students copy from each other's work and then submit it as their own. The material may not have been plagiarised from the web, so search engines cannot detect it.

A more detailed analysis of some of these points can be found on the Turnitin website (*http://www.turnitin.com/static/products_services/search_engines.html*).

WHAT ARE THE ALTERNATIVES?

Table 1 compares the features available in two alternative software products that are currently freely available to UK HE institutions, namely the JISC Plagiarism Detection Service (JISCPDS) (*http://www.submit.ac.uk*) and CopyCatch Lite from CFL Software (*http://www.copycatchgold.com/copycatchesreview.htm*). JISCPDS, unlike CopyCatch and Google, is a service based on Turnitin software and run, via JISC, by US company, iParadigms. All submissions from UK institutions are held on a database in Reading, UK. Access to the service is controlled through account creation and all submissions require student permission for their work to be submitted. Contractual safeguards have been put in place to ensure iParadigms complies with EU Data protection laws.

JISCPDS is based on Turnitin software (*http://www.turnitin.com*) and primarily used for detecting internet plagiarism, but can be used for limited detection of collusion in a batch of documents via repeated comparison of pairs of documents in a batch. The free CopyCatch 'Lite', on the other hand, is only able to be used for making pair-wise comparisons of text in documents, although CopyCatch 'Gold', can make multiple comparisons. Use of both JISCPDS and CopyCatch enables very comprehensive screening for both internet plagiarism and collusion.

Both these products are easy to use, screen whole documents in seconds (or minutes) and allow bulk upload of files. This enables rapid and accurate screening of large numbers of documents and ensures that the whole process is much less random than the use of Google. Both software products can handle a wide variety of commonly used file formats and provide very comprehensive comparison reports. They can also be made available to students for