detect exact or partial copies of documents in the library in TeX, DVI, and Troff formats. The system looks for documents with significant overlap as well as exact copies.

These documents are first converted into ASCII format. They are then divided into sentences called units, and these sentences are further grouped together to form a series of sentences called chunks. These sentences are stored in a registration server that is simply a large hash table using a standard hashing algorithm. These chunks are compared with the other documents to check whether there is overlapping. If the documents share a preset number of sentences, then a violation is flagged. Figure 9-4 shows the COPS architecture.