Incoherent text: The text is not in the proper style and appears to be written by many people.

Punctuation: The punctuation marks used in one text are the same as in another text. It is not likely for two different authors to use the same punctuation marks while writing the text.

Dependence on certain words and phrases: Certain words and phrases are used by one author as well as by another author. Different authors tend to have different word preferences.

Amount of similarity between texts: Two texts written by two different authors share large amounts of similar text.

Long sequences of common text: Long sequences of common words or phrases are in the text.

Similarity in the order of text: Two texts have the same order of words and phrases.

Frequency of words: Two texts contain the same frequency of words.

Common spelling mistakes: An independent author makes the same spelling mistakes repetitively as another author.

Distribution of words: The distribution of word usage by an independent author appears in the same fashion throughout the document as another's work.

Syntactic structure of the text: Two texts written by different authors have similar syntactic structure. Different authors often use different syntactic rules.

Preference for the use of long or short sentences: If a sentence is long and shows no meaning in the text, it is possible that the author has combined sentences copied from another text.

Readability of written text: The same readability is found in the works of two different authors.

Inadequate references: References appear only in the text, but not in the bibliography.

Plagiarism Detection Tools

The following are the three categories of plagiarism detection tools:

1. Tools to detect plagiarism in text, such as Sumbit.ac.uk and CopyCatch, are helpful in checking plagiarism in works submitted in Microsoft Word, Corel WordPerfect, and text formats.

2. Tools to detect plagiarism in source code, such as JPlag and CodeMatch, help in finding similar source code from multiple sets.

3. Tools such as BOSS from Warwick University's computer science department assist in the process of data collection.

Tool: Turnitin

Turnitin is an online plagiarism detection tool primarily for educators and students. Turnitin detects plagiarism by comparing the submitted work to pages available on the Internet and in its database. Figure 9-2 shows a screenshot from Turnitin.

The following are the key features of Turnitin:

- Plagiarism prevention: It helps identify the plagiarized work of students and also acts as a deterrent, stopping plagiarism before it starts.

- Peer review: It helps students review each other's work.

- Grademark: This tool helps instructors, without much hassle, in assessing works submitted by students. Instructors can add comments to the submitted work without altering the formatting of the document.

- Gradebook: It is similar to a paper gradebook, where the instructor can manage assignments and grade students in a more organized manner.

- Digital portfolio: It is an online student record book, helping to track student records for academic purposes or for placements.

Tool: CopyCatch

CopyCatch supports various formats such as Rich Text Format (RTF), Microsoft Word documents, and text. After checking documents for plagiarism, this utility highlights the changes on the screen and saves them in RTF format. It includes Web search comparison, zip archive submission, and course/module filtering. CopyCatch is shown in Figure 9-3.

Tool: Copy Protection System (COPS)

The Copy Protection System (COPS) is an experimental working prototype of a copy detection system that can be used in a digital library. The COPS part of the project is to