- Online News Articles
- USENET Postings
- Academic Paper Extract
- Text From a Piece of Fiction

Where the source (such as the news article) contains headings or extra pieces of text (such as extracts), only the main body of the text will be used.
The data to be hidden will be randomly generated bitstream of around 2000-3000 bits. This data will be used for all pieces of text so a direct comparison can be made. The tests will be repeated with a second set of random data as the bits in the bitstream can make a different to the output, in particular the bitrate.
All the sample text and data files along with their outputs for the two bitstreams, can be found on the CD.

5.3 Statistical Evaluation Criteria

The text will be tested in three ways:

- BitRate - For each piece of text the bitrate per 10/100 words will be calculated.
- Quality Test - For each word in the text the quality will be calculated before the data is hidden and afterwards to discover the impact on the quality of the text. The quality will be calculated using the bigram data in the same way that the quality is tested during obfuscation in the actual algorithm. The quality will also be given as an average per word in the text. Two values will be provided, total score for the text and the average per word.
- Correctness - This will be calculated by calculating fie different values.
- User Survey To test the robustness of the algorithm to human inspection a short survey will be used containing sample output from the above tests.
  - False Positives - The number of words in which bits are found by deobfuscation which do not contain any hidden data.
  - False Negatives - The number of words in which data is hidden but not recovered by deobfuscation.
  - True Positives - The number of words in which data is hidden and found.
  - True Negatives - The number of words in which data is hidden and no data is found.
  - Error Count - This will be a count of any situation where bits are hidden and bits are found, but the recovered bits are not the bits that were originally hidden.