

quality is found to be above the threshold (set to the same as for obfuscation), then the algorithm proceeds to attempt deobfuscation. Only words which have passed this quality test can contain data, so if it fails you know that no data is hidden in the word.

To deobfuscate a word, the synonym list for the word is found as previously described. This should return as close to the synonym list of the original word as possible. The algorithm then retrieves the position in this list of the word to be deobfuscated, and this position is taken to be the hidden data (with the first and second positions taken to be 0 and 1 respectively, and the third position 1,0). If the word is not within the first three positions then it is ignored and it is assumed no bits are hidden (as it would fail the deobfuscation check when

```

Data: Synset, word
Result: Hidden bit(s)
for  $i \leftarrow 0$  to synsetsize do
    if synset[ $i$ ] = word then
        | return  $i$ ;
    end
end
return null;

```

Algorithm 3: Deobfuscation Algorithm Pseudo-code

3.1.5 Quality Checks

The quality test uses the bigram data from the American Nation Corpus [28]. The tests are performed on the word that is being processed by locating the frequency in the bigram data by searching for the bigram previousword-current. If the word is the first in a sentence, you cannot test with the word before, so in this case the search is for the bigram current-nextword. If the bigram is not in this list it has a frequency less than 4, so the returned quality is zero, else the returned quality is the frequency.

3.2 StegChat

The StegChat application will be a prototype demonstration application for the described synonym-based text steganography algorithm. The application will be a web based application which will facilitate the continuous “chat” between two users. When supplied with data, the algorithm should attempt to hide as much data as possible without compromising the quality of the cover-text messages. The user who receives the messages should be able to easily retrieve the hidden data, preferably automatically.

The application will take the form of a web application, so the entire application will be usable within the users browser. The site will be built using Java servlets and jsp pages as much as possible for two reasons; firstly Java is my primary language and secondly by keeping the algorithm code within Java it will be easy to produce a local version of the application if required.