22.4
There are several open source implementations for performing classification. You could use one of those and your implementation is always encouraged. Apply the spam-filtering algorithm onto the set of features you have extracted (bigrams, unigrams and trigrams) and use the test set of your data set (you could do 80:20 split on your data i.e. 80% for training (or learning) and 20% for testing). The expected accuracy is high (> 80%) and for anything less, the selected features need to be evaluated and new features might need to be introduced.

22.7
The simplest approach is to look for a string of capitalized words, followed by “Inc” or “Co.” or “Ltd.” or similar markers. A more complex approach is to get a list of company names (e.g. from an online stock service), look for those names as exact matches, and also extract patterns from them. Reporting recall and precision requires a clearly defined corpus, which should include at least 10 news articles with a number of company names mentioned in each.

One example code:

```
[0-9]*\s(((A-Z(&)*+[a-z.\(\)\;]\*\s)+)?=Company|company|Inc\.|Inc|Co\.|Corporation|Limited|Ltd\.)
```

The good regular expression has both accuracy and recall around or above 50%. But a little lower than that is fine.