Implementing spelling correction

There are two basic principles underlying most spelling correction algorithms.

1. Of various alternative correct spellings for a mis-spelled query, choose the ``nearest'' one. This demands that we have a notion of nearness or proximity between a pair of queries. We will develop these proximity measures in Section 3.3.3 .
2. When two correctly spelled queries are tied (or nearly tied), select the one that is more common. For instance, grunt and grant both seem equally plausible as corrections for grnt. Then, the algorithm should choose the more common of grunt and grant as the correction. The simplest notion of more common is to consider the number of occurrences of the term in the collection; thus if grunt occurs more often than grant, it would be the chosen correction. A different notion of more common is employed in many search engines, especially on the web. The idea is to use the correction that is most common among queries typed in by other users. The idea here is that if grunt is typed as a query more often than grant, then it is more likely that the user who typed grnt intended to type the query grunt.

Beginning in Section 3.3.3 we describe notions of proximity between queries, as well as their efficient computation. Spelling correction algorithms build on these computations of proximity; their functionality is then exposed to users in one of several ways:

1. On the query carot always retrieve documents containing carot as well as any ``spell-corrected'' version of carot, including carrot and tarot.
2. As in (1) above, but only when the query term carot is not in the dictionary.
3. As in (1) above, but only when the original query returned fewer than a preset number of documents (say fewer than five documents).
4. When the original query returns fewer than a preset number of documents, the search interface presents a *spelling suggestion* to the end user: this suggestion consists of the spell-corrected query term(s). Thus, the search engine might respond to the user: ``Did you mean carrot?''
