ural language processing, including spelling correction task. (Brill and Moore, 2000) presented an improved error model over the one proposed by (Kernighan et al., 1990) by allowing generic string-to-string edit operations, which helps with modeling major cognitive errors such as the confusion between *le* and *al*. Via explicit modeling of phonetic information of English words, (Toutanova and Moore, 2002) further investigated this issue. Both of them require misspelled/correct word pairs for training, and the latter also needs a pronunciation lexicon, but recently (Ahmad and Kondrak, 2005) demonstrated that it is also possible to learn such models automatically from query logs with the EM algorithm, which is similar to work of (Martin, 2004), learning from a very large corpus of raw text for removing non-word spelling errors in large corpus. All the work for non-word spelling correction focused on the current word itself without taking into account contextual information.

Real-word spelling correction is also referred to be context sensitive spelling correction (CSSC), which tries to detect incorrect usage of valid words in certain contexts. Using a pre-defined confusion set is a common strategy for this task, such as in the work of (Golding and Roth, 1996) and (Mangu and Brill, 1997). Opposite to non-word spelling correction, in this direction only contextual evidences were taken into account for modeling by assuming all spelling similarities are equal.

The complexity of query spelling correction task requires the combination of these types of evidence, as done in (Cucerzan and Brill, 2004; Li et al., 2006). One important contribution of our work is that we use web search results as extended contextual information beyond query strings by taking advantage of application specific knowledge. Although the information used in our methods can all be accessed in a search engine's web archive, such a strategy involves web-scale data processing which is a big engineering challenge, while our method is a light-weight solution to this issue.

3 Motivation

When a spelling correction model tries to make a decision whether to make a suggestion *c* to a query *q*, it generally needs to leverage two types of evidence: the similarity between *c* and *q*, and the validity plausibility of *c* and *q*. All the previous work estimated plausibility of a query based on the query string itself – typically it is represented as the string probability, which is further decomposed into production of consecutive n-gram probabilities. For example, both the work of (Cucerzan and Brill, 2004; Li et al., 2006) used n-gram statistical language models trained from search engine's query logs to estimate the query string probability.

In the following, we will show that the search results for a query can serve as a feedback mechanism to provide additional evidences to make better spelling correction decisions. The usefulness of web search results can be two-fold:

First, search results can be used to validate query terms, especially those not popular enough in query logs. One case is the validation for navigational queries (Broder, 2004). Navigational queries usually contain terms that are key parts of destination URLs, which may be out-of-vocabulary terms since there are millions of sites on the web. Because some of these navigational terms are very relatively rare in query logs, without knowledge of the special navigational property of a term, a query spelling correction model might confuse them with other low-frequency misspellings. But such information can be effectively obtained from the URLs of retrieved web pages. Inferring navigational queries through term-URL matching thus can help reduce the chance that the spelling correction model changes an uncommon web site name into popular search term, such as from *innovet* to *innovate*. Another example is that search results can be used in identifying acronyms or other abbreviations. We can observe some clear text patterns that relate abbreviations to their full spellings in the search results as shown in Figure 1. But such mappings cannot easily be obtained from query logs.

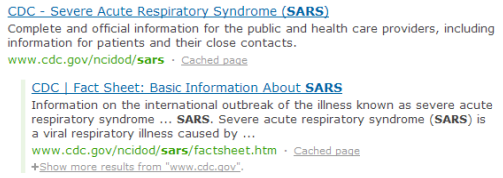

Figure 1. Sample search results for *SARS*

Second, search results can help verify correction candidates. The terms appearing in search results, both in the web page titles and snippets, provide additional evidences for users intention. For example, if a user searches for a misspelled query *vaccum cleaner* on a search engine, it is very likely that he will obtain some search results containing the correct term *vacuum* as shown in Figure 2. This