1. Самым банальным вариантом решения будет банальный поиск по словарю и в случае неудачи перебор возможных изменений с последующим повторнып поиском по словарю. Однако этот вариант очень долго работает что неприемлимо.

Поэтому лучше использовать другой метод.

2.Для работы необходим датасет, в случае поисковика на сайте можно использовать пользовательские коментарии,отзывы и посисковые запросы.

Для исправления ошибки в слове необходимо сначала оценить вероятность использования правильного слова а затем вероятность совершения в этом слове конкретной опечатки, для этого нам лучше использовать две раздельные модели: модель ошибки и модель языка.

Для создания модели языка из датасета необходимо убрать все лишние слова (слишом длинные >20 символов) и слишком короткие (меньше 3) кроме некоторых известных исключений. После чего текст нужно побить на комбинации фиксированной длинны к примеру 3 для того, чтобы оценивать вероятность слова можно было относительно контекста. Далее находим встречаемость слова, также триграмм и двуграмм в нашем корпусе. Теперь можно посчитать вероятность написания пользователем какой-либо конкретной фразы. Допустим такую фразу (w1,w2,w3) вероятность всей фразы будет равна произведению p(w1)\*p(w1,w2)/p(w1)\*p(w1,w2,w3)/p(w1,w2).

Для модели ошибки можно воспользоваться старым методом [2]. Чтобы узнать вероятность опечатки в каком-либо конкретном слове необходимо разбить правильное и опечатанное слова на одинаковое количество кусков всеми способами, после чего можно сравнить произведение вероятностей замены одного элемента на другой для каждого способа и выбираем максимум. Выбрав таким способом некоторое количество наиболее вероятных опечаток мы можем использовать имеющуюся модель языка для вычисления самого вероятного правильного слова.

[2] E.Brill, R. C. Moore. [*An Improved Error Model for Noisy Channel Spelling Correction*](http://www.aclweb.org/anthology/P00-1037). Proceedings of the 38th Annual Meeting on Association for Computational Linguistics, 2000.