

30th Annual
Rowan University Programming Contest
hosted by the
Computer Science Department
Saturday, 23 April 2016



30th Annual
Rowan University Programming Contest
hosted by the
Computer Science Department
Saturday, 23 April 2016

Automatic Spelling Correction



Common Errors

Your program will need to check for these common errors:

Type	Correct Word	Example Error
Omission	happily	hapily
Insertion	billion	billion
Transposition	telephone	telehpone
Substitution	cadmium	cafmium
Capital	Christopher	christopher

Some Changes Are Not Errors

In standard English, a word at the beginning of a sentence gets a capital letter in the initial position.

If **tomorrow** is on the word list, **Tomorrow** is considered correct.

Sometimes, people use all-uppercase for emphasis (a text version of SHOUTING).

If **never** is on the word list, **NEVER** is considered correct.

Some Things Are Not Checked

People sometimes type things such as **WorldWar2**, or suggest password ideas such as **passw0rd**.

A word not on the word list, but which has non-alphabetic characters, does not need to be checked.

Sometimes people use acronyms, such as **CPU** or **UNICEF**.

A word not on the word list, but which is all-uppercase, does not need to be checked.

Check Some Things Only Once - Omissions

The word **stapler**, if a letter is omitted, might be **staper**.

If four letters are omitted, it might be **sar**.

But seeing the word **sar** and suggesting **stapler** would be ridiculous.

When checking for omissions, only check for single-letter omissions.

Check Some Things Only Once - Insertions

The word **lamp**, if a letter is added, might be **laamp**.

If five letters are added, it might be **clampring**.

But seeing the word **clampring** and suggesting **lamp** would be ridiculous.

When checking for insertions, only check for single-letter insertions.

Check Some Things Only Once - Transpositions

The word **steal**, if two letter are transposed, might be **setal**.

If two sets of letters are transposed, it might be **stale**.

But seeing the word **stale** and suggesting **steal** would be ridiculous.

When checking for transpositions, only check for a single adjacent pair of letters.

Check Some Things Only Once - Substitutions

The word **accepted**, if a letter is substituted, might be **accepred**.

If four letters are substituted, it might be **admitted**.

But seeing the word **admitted** and suggesting **accepted** would be ridiculous.

When checking for substitutions, only check for a single character.

Capital Letters Are Different

For upper-case letters, you do have to check for multiple errors.

If **McKinley** is correct, and you get **mckinley**, then that's two letters incorrectly rendered as lowercase.

You do have to suggest **McKinley** as a correction in that situation.

And while **Today** and **TODAY** are acceptable for **today**, things such as **toDAY** or **todaY** are not. You should suggest **today** as a correction in such a situation.

1. Uppercase letters switched to lowercase letters should be reported as mistakes.
2. Lowercase letters switched to uppercase are okay if it is only the first letter, or the entire word. Else, they should be reported.

Do Not Do Multiple Errors At Once

In theory, someone might see **havesatde**, and guess correctly it was supposed to be **harvested**.

It is possible to check for an omission, an insertion, and a transposition all in one word, but that's a level of complication beyond this contest.

You do not have to check for multiple errors at once.