SRS SPELL RATER

Spell Checker for Web Sites

A F DilshanFardil
rishieni@gmail.com
R M P D Rathnayake
T A I S Thembuwana

Table of Contents

1 Introduction

At the expense of pages with weak, thin or poor-quality content, Google has implemented numerous significant changes to its PageRank algorithm which reward excellent content on webpages with higher rankings in the search engine results pages (SERPs) since early in 2011.

The misspelling and typographical errors can be observed in content of the poor-quality webpages. If the written text in the web pages is not developed properly and is unable to convenience the task of conveying meaningful and useful information to readers, the web page has inherently little value. It makes the web page sophisticate as a text with no an expert content. Therefore, users will likely not stay long on the page. Bounce rate is a site value factor for ranking pages in search.

The pages with one or two typos cannot be avoided concerning and also not all junk pages are contained with only misspelling. However, the correlation between poor quality text content and the minimal value to users is high, and numerous, sloppy spelling errors are often indicative of low-value content. It should be ensured that it is well-written, clear in its objective, well-written, clear in its objective, and free from misspelling to optimize the value of your text content to readers and thus to search engines. The typos can be a reason to diminish the value of the content and the worst-case scenario is when the keywords of the text are misspelled, the very important message it tries to convey is not expressed.

1.1 Purpose

A tool that can spell check online is can be the suitable solution for optimizing the quality of the text content on the webpages.

Use the data resulting from this tool to improve the quality of the site's content pages. Readers, and the search engines, will appreciate the improvements because this spell check tool is able to support the development of the value of the text content.

Once this online spellchecker tool has completed the scan, it displays a tabular report listing each URL scanned and the number of possible misspellings found on each page. These potential errors can then be review and corrected as needed.

1.2 Scope of Project

A website's bounce rate is a valuable factor that will impact the ranking of that page within the SERPs. Overall, poor content is likely to increase the webpage's bounce rate; which is the act of a website visitor clicking the 'back' button and leaving a website, when they don't receive the information they were looking for.

This free utility crawls your site and shows you all your misspellings on hover over.

This tool does not check the following:

- Words that feature a capital letter

The tool is designed to scan text displayed on the page, not text in metadata tags, such as description, or alt tags, nor text in images, Flash or Silverlight content.

Running this tool on numerous pages takes time; the larger the number of pages to be scanned, and the more text per page to scan; the more time it takes to complete the report.

To help better leverage the value of the content the website offers to human readers, in turn how it is viewed by search engines, it is important to ensure that the content is:

- well-written
- clear in its objective
- uses proper grammar
- free of misspellings

To run this online spell check tool for the content on a website, type in or copy/paste the URL of the website's homepage in the provided text box. Select the number of 'Pages to be Crawled', from the ranges provided and input any misspellings to be ignored in the scan, then click 'Spell Rater'.

1.3 Definitions, Acronyms and Abbreviations

Term	Definition
SERP	**Search Engine Result Page** A search engine results page (SERP) is the page displayed by a search engine in response to a query by a searcher. The main component of the SERP is the listing of results that are returned by the search engine in response to a keyword query, although the page may also contain other results such as advertisements.
URL	**Uniform Resource Locator** commonly informally termed a web address (a term which is not defined identically) is a reference to a web resource that specifies its location on a computer network and a mechanism for retrieving it.

1.4 References

IEEE. *IEEE Std 830-1998 IEEE Recommended Practice for Software Requirements Specifications.* IEEE Computer Society, 1998.

1.5 Overview of Document

The next chapter, the Overall Description section, of this document gives an overview of the functionality of the product. It describes the informal requirements and is used to establish a context for the technical requirements specification in the next chapter.

The third chapter, Requirements Specification section, of this document is written primarily for the developers and describes in technical terms the details of the functionality of the product.

2 Overall Description

2.1 System Environment

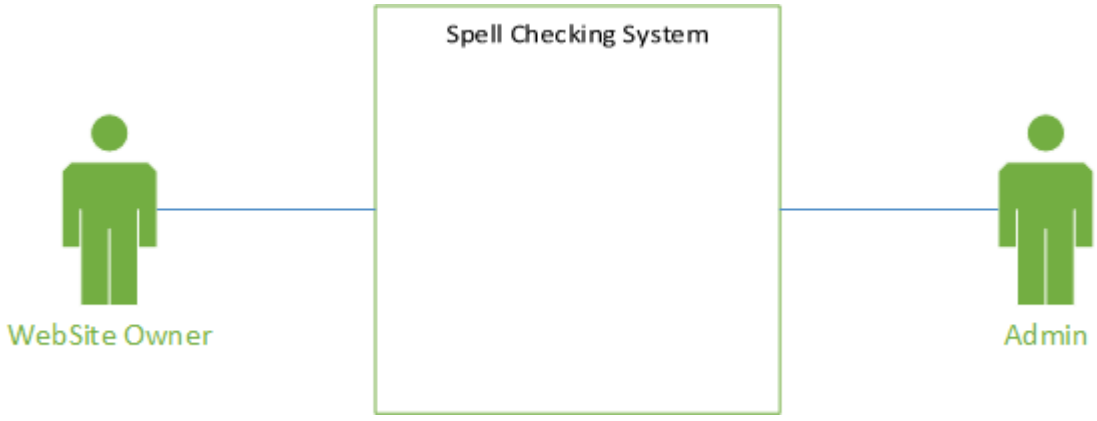


Figure 1 - System Environment

The Spell-Checking System has two active actors, named Website owner and admin. Website owner accesses from the front end (client side) through the internet and admin can accesses from the back end (Server side).

2.2 Functional Requirements Specification

This section outlines the use case for each of the actors separately. The main actor is Web site owner and he has one main use case which included and extended many other use cases. Admin has special administrative use cases.

2.2.1.1 Web site owner Use Cases

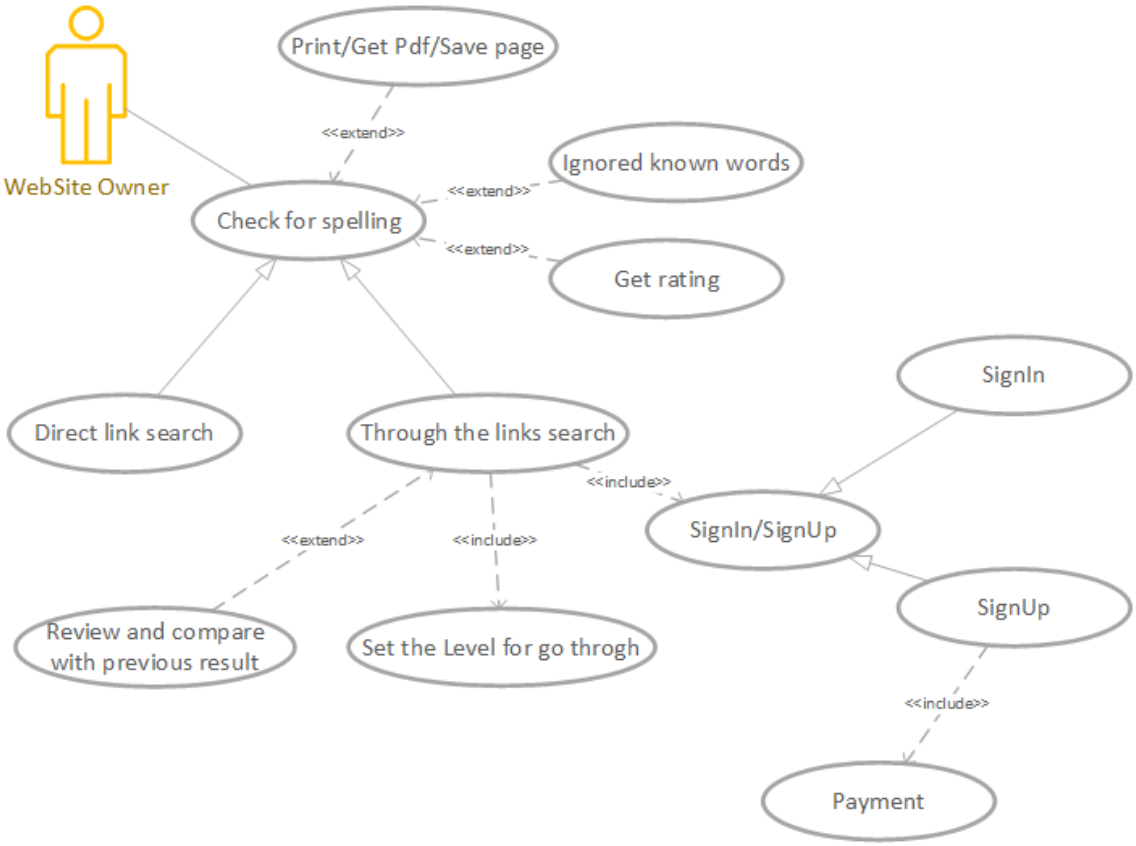
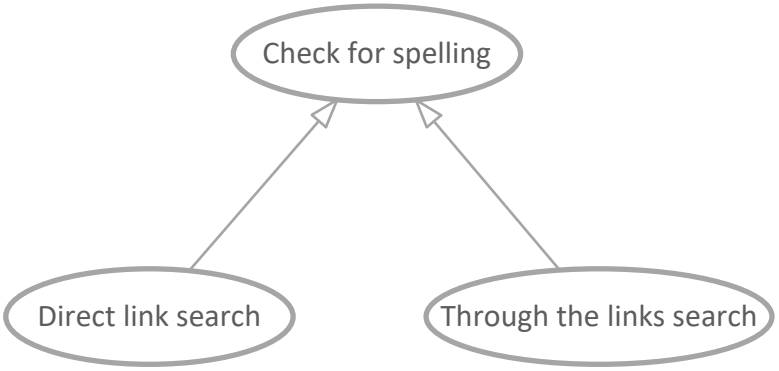


Figure 2 - Web site owner Use Case Diagram

2.2.1.2 Use case: Check for spelling

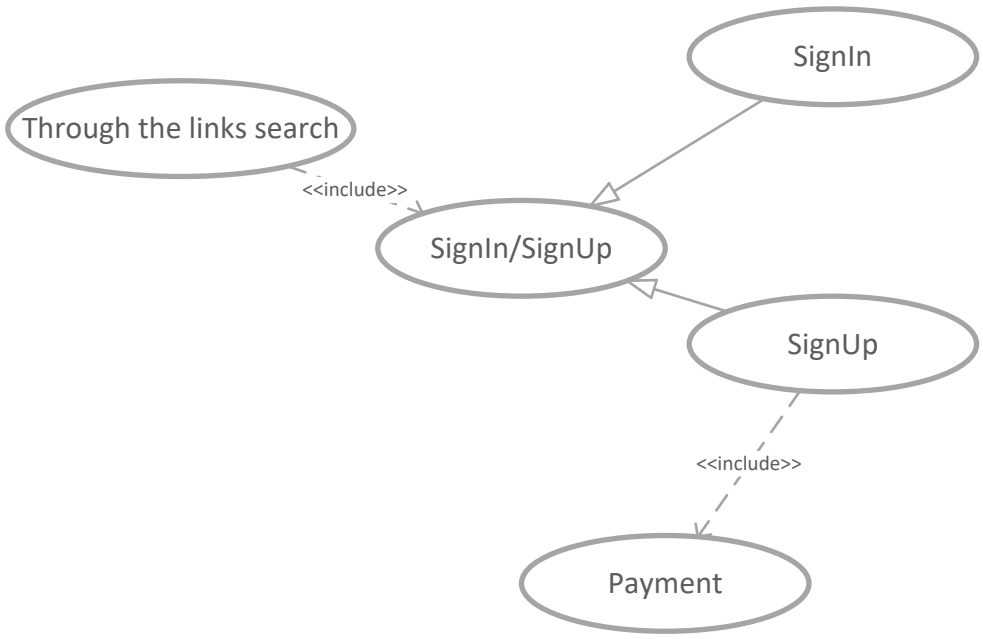
The website owner can access to the spell-checking web app through internet and there have two options for direct link searching and through the links searching. User can choose any option and each option has different features.



2.2.1.3 Use case: Direct link search

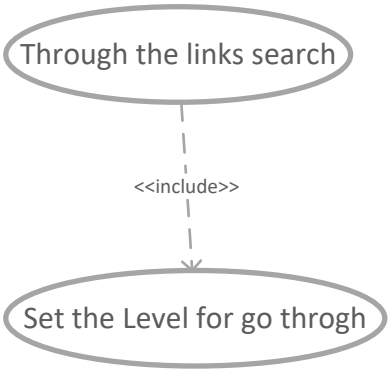
In main window, there is a text area for paste the link of searched web. User can be paste any web link on this area and by clicking “Check” button the whole page referred to the link can be checked. This use case is totally free but system checks only web page that referred to the pasted link and not interested for the linked web pages in this.

2.2.1.4 Use case: Through the links search



For use this option, use should be a registered member. If user has existing profile entering user name and password, he can sign in to their account. If user hasn’t existing profile user should sign up by giving basic personal details. For get an account also it has a payment and after fill the basic details and payment requirement user will get a permanent account.

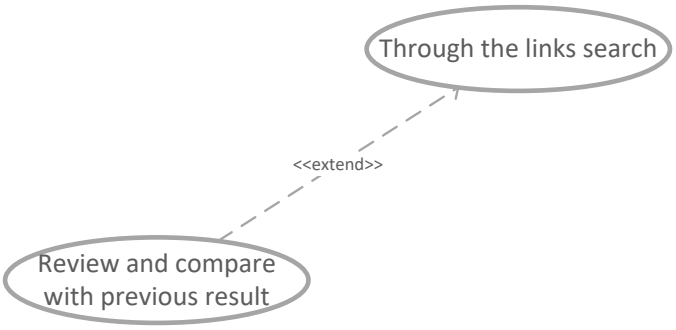
2.2.1.5 Use case: Set the level for go through



Special future of this option is when you check spelling for a link system will go through all the links on the main page and check their content also. User has an option to select levels that how many steps need to be checked going through the links. (More no of levels take more time to execute for the result)

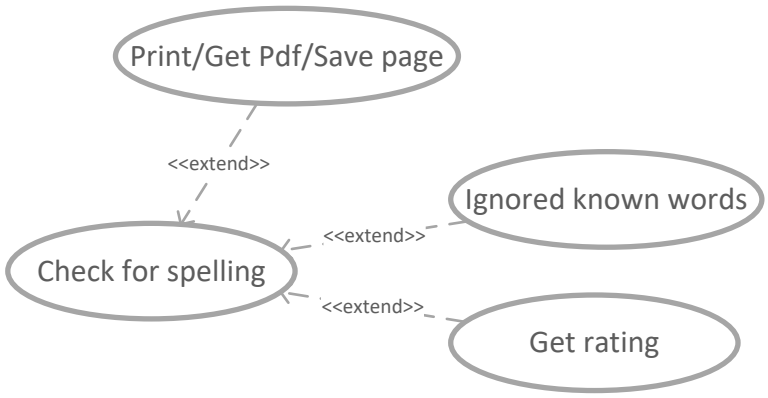
Then user can get a rate for their whole web page about the accuracy of content easily and by pasting one main link.

2.2.1.6 Use case: Set the level for go through



System stores the all past results of the user and their web pages. With this use case user, can compare their result with previous results and can get more details about the accuracy of the website in present state comparing previous states.

Check for spelling use case extend by three other use cases.



2.2.1.7 Use case: Print/Get Pdf/Save page

After user get their result as a document view user can get a print of whole document or can get PDF or can save the page for further reference.

2.2.1.8 Use case: Ignored known words

By this use case user can select known, but system have been noticed as incorrect words such as persons or places names, special technical words as ignored. Then they will count as correct words and not affected to the rating of the website.

2.2.1.9 Use case: Get rating

By this use case user can get rate for their web site or web page. System gives rating as a percentage according to the accuracy of the spelling in content. User can get a final overview about his website.

2.2.1.10 Administrator Use Cases

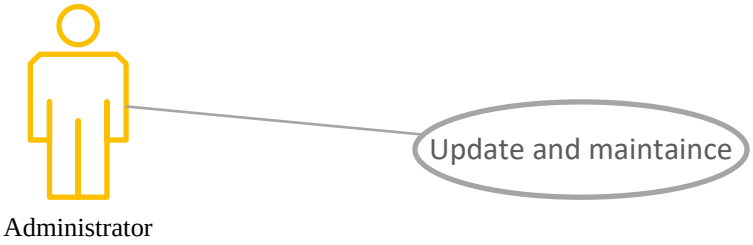


Figure 2 – Administrator Use Cases

Administrator can log from the server side and can fix the bugs, update the system and maintain the system.

3 Specific Requirements

This application is planning to implemnt on different satges.

Our first step is to get all the spelling mistakes on the target URL with all the Field names, paragraph, comments and ect.Then the user finaly sees the report of the uncorrect spellings what are the places that the mistakes was done and finally a report with presentage of mistakes that divided into abow mention catagories.

In the second step is to get the whole web site and go through all the website that related to the main website and give a report on the mistakes sapartely

As the thired step we can extend our project to check the grammer mistakes on the website also the attached documents like .pdf, .txt, .xlsx, .doc, .docx, and .ppt. then we can give a separte reports on that.

3.1 Functionality

3.1.1.1 Check Spellings

Use Case Name	**Check Spellings**
XRef	Sec 2.2.1.1 and Sec 2.2.1.7 Check Spellings
Trigger	The User selects to Check Button
Precondition	The User view the report of the current site
Basic Path	1. Type the URL of the wanted site 2. Click the Check Button
Alternative Paths	In step 1, if the URL entered is not valid get a massage In step 3, if any required field is blank, the User is instructed to add an entry.
Postcondition	The system presents the report of the website
Exception Paths	If there is no such URL or cannot access to the specific URL
Other	1. we are using the Microsoft BING Spell Cherk API (https://dev.cognitive.microsoft.com/docs/services/56e73033cf5ff80c2008c679/operations/56e73036cf5ff81048ee6727) 2. It removers the known words form the content 3. We get the whole txt in the URL to one String Variable 4. Check the String using the spell chek api 5. Uses DataView in JS to store the URL, total words,tags, misspelled words, this location of the incorrect word 6. From this DataView Collection model we can get the results and calculate the Error rate and the accuracy rate 7. *Error rate* = $\frac{total\ number\ of\ errors}{otal\ number\ of\ words}$ 8. *accuracy rate = 1 − error rate*

3.1.1.2 Derict Link Search

Use Case Name	**Derict Link Search**
XRef	Sec. 2.2.1.2 Direct link search
Trigger	The user access the target URL
Precondition	The User has accessed the Report Manager main screen and the errors are in the database
Basic Path	Get to the URL that the User was entered
Alternative Paths	-
Postcondition	-
Exception Paths	We cannot access to the website that user entered show a massage Box
Other	-

3.1.1.3 Through Link Search

Use Case Name	**Through Link Search**
XRef	Sec 2.2.1.3 Through link search
Trigger	The User Access the whole website contain all the pages
Precondition	Get the report of the all pages
Basic Path	1. Map all the web site URLs to Array 2. Get one by one and get the report for each page 3. Get the report for per page and the whole report for the website
Alternative Paths	-
Postcondition	-
Exception Paths	If the all URL that we are mapping exceed the limit of our expected limit we have to go for it for two levels
Other	This feature is for pro Users

3.1.1.4 SignIn

Use Case Name	**Sign In**
XRef	Sec 2.2.1.3 Sign In;
Trigger	The User selects to sign in because he was registered before
Precondition	The User has accessed the Login screen.
Basic Path	1. The Editor type a email address. 2. User type the password 3. Click the sign In Button 4. Check from the database fields . 5. The system checks that required fields are not blank.
Alternative Paths	In step 1, if the email entered is not valid sign In Button is not active In step 3, if any required field is blank, the User is instructed to add an entry.
Postcondition	The system presents the database information and get access to the pro account.
Exception Paths	If there is no such person is in the database get a massage for check or no user found massage
Other	This use case is not used when direct link is selected

3.1.1.5 SignUp and Paymnt Method

Use Case Name	**SignUp and Paymnt Method**
XRef	Sec 2.2.1.5 Signup
Trigger	The User selects Signup
Precondition	The User view the Signup Form
Basic Path	1. Type the Name (First Name and Last Name) 2. Type the e-mail Address 3. Having a Website 4. Enter a password 5. Re Enter the password 6. Select the payment method
Alternative Paths	In step 2, if the email already exist get a massage In step 2, if the email is not a valid one get a massage In step 4, Must match with the password In step 6, if any required field is blank, the User is instructed to add an entry. In last step, Select a payment method PayPal or a VISA /MasterCard
Postcondition	Crate the Pro Account
Exception Paths	If the Credit Card no or the PayPal account is invalid get a massage to check the numbers once again
Other	To store the password, credit card no and the code we use our encryption method for that

3.1.1.6 Set the level for go through

Use Case Name	**Set the level for go through**
XRef	Sec 2.2.1.6 Set the level for go through;
Trigger	The User selects the level for go through
Precondition	Get the report for the whole pages
Basic Path	Get the correct depth for the website (Check with the account type wither this is a pro account or not)
Alternative Paths	If the the depth was too large get a massage Box
Postcondition	-
Exception Paths	-
Other	-

3.1.1.7 Print/Get pdf/ Save Page

Use Case Name	Print/Get pdf/ Save Page
XRef	Sec 2.2.1.7 Print/Get pdf/Save Page;
Trigger	The User selects to export or Print
Precondition	The User has accessed the print view
Basic Path	1. Print the report 2. Export as a pdf
Alternative Paths	-
Postcondition	-
Exception Paths	-
Other	-

3.1.1.8 Review and compare with previous results

Use Case Name	Review and compare with previous results
XRef	Sec 2.2.1.8 Review and compare with previous results;
Trigger	The User selects to assign a reviewer to an article.
Precondition	The User has accessed the Report Manager main screen and the errors are in the database
Basic Path	1. We store the above mention data (In User Case : Spelling Check) on the database. 2. Get the Current URL results 3. Compare the values of two reports
Alternative Paths	-
Postcondition	-
Exception Paths	-
Other	This is not use for the free users

3.2 Usability

Usability is more simple and the reports of previous and current report compairing is so much important as a website owner.The web site is very simpale is like this Jest past the URL on the txtField.

3.3 Reliability

The development team should put effort on reusability of ideas as well as codes. Two components that reusability is important include the word segmentation and the string matching algorithms.

3.4 Maintainability

Maintanabilty is highly consernd about us becase there are differnet stages on the project we have to extend different levels so that the coding is done in a layered architechure so the Extending the project and the maintanabilty is very easy.

3.5 Purchased Components

We have to get a server with a level of security and a database for store the results reports for the pro version

3.6 Interfaces

3.6.1.1 User Interfaces

The user interface is important. Since the users tend to be normal computer users, we decide to use Graphical User Interface (GUI) as the interface. GUIs are easier to use than any other interfaces. The Spell rater is a application contains One static page with 3 Sub deviation's Logging/Register, A deviation's to Past the URL, Reporting devition. Mainly we are focusing on the reporting part.

3.6.1.2 Software Interfaces

This software must have a Login because we give some extra options for the users that have paid for the Pro account. For that parpous we need a interface for selection the payment method and to add a VISA /MasterCard card or sync a PayPal account. Then in the user side user have to do some modifications on their accounts so for that we need another Interface.

For the users if they want see previously report and compare them with current reports for that we need another one.