TOPICS

Intro
Statement
Approaches
Flow Chart
Algorithm
Relevance
Applications

Intelligent Chatbot to Analyze and Monitor Websites

Jose Kurisingal
George Puthanangadi
Akshay Krishna

Introduction

TOPICS

Intro
Statement
Approaches
Flow Chart
Algorithm
Relevance
Applications

CHAT BOT

- A chatbot is a piece of software that conducts a conversation via auditory or textual methods.
- Such programs are often designed to convincingly simulate how a human would behave as a conversational partner, although as of 2019, they are far short of being able to pass the Turing test.
- Chatbots are typically used in dialog systems for various practical purposes including customer service or information acquisition. Some chatbots use sophisticated natural language processing systems, but many simpler ones scan for keywords within the input, then pull a reply with the most matching keywords, or the most similar wording pattern, from a database.

Problem Statement

TOPICS

Intro
Statement
Approaches
Flow Chart
Algorithm
Relevance
Applications

to create a basic chat bot using python

Approaches

TOPICS

Intro
Statement
Approaches
Flow Chart
Algorithm
Relevance
Applications

MACHINE LEARNING= Decision Tree Algorithms

- In this algorithm a decision tree is used to map decisions and their possible consequences, including hances, costs and utilities.
- This method allows the problem to be approached logically and stepwise to get to the right conclusion.
- An important algorithm that evolved from this algorithm is the Random Tree algorithm. This algorithm uses multiple trees to avoid overfitting that often occurs with using decision trees.

TOPICS

Intro
Statement
Approaches
Flow Chart
Algorithm
Relevance
Applications

Building a Chatbot using NLTK Natural Language Processing with Python provides a practical introduction to programming for language processing.

- Downloading and installing NLTK
  Install NLTK: run pip install nltk
- Installing NLTK Packages
  import NLTK and run nltk.download().
- Text Pre- Processing with NLTK
  Converting the entire text into uppercase or lowercase, so that the algorithm does not treat the same words in different cases as different

Relevance

TOPICS

Intro
Statement
Approaches
Flow Chart
Algorithm
Relevance
Applications

- Chatbot applications streamline interactions between people and services, enhancing customer experience. At the same time, they offer companies new opportunities to improve the customers engagement process and operational efficiency by reducing the typical cost of customer service.
- To be successful, a chatbot solution should be able to effectively perform both of these tasks.
- Human support plays a key role here: Regardless of the kind of approach and the platform, human intervention is crucial in configuring, training and optimizing the chatbot system.

Applications

TOPICS

Intro
Statement
Approaches
Flow Chart
Algorithm
Relevance
Applications

- 1. Product Suggestions
- 2. Customer Support
- 3. Weather
- 4. Personal Finance Assistance

Intro
Statement
Approaches
Flow Chart
Algorithm
Relevance
Applications

Scraping webpages using scrapy

November 19, 2019

Objective

- **Scrap information from webpages of a given website using Scrapy**

What is Web scraping ? What is it's relevence ?

Web scraping refers to the process of collecting and organizing data from a web page.The steps involved in web scraping are Fetching web pages and Extracting data from it

Applications of web scraping

- For contact scraping
- For web indexing
- For data mining
- For comparing and doing researches on data

Existing methods

1.Human Copy-Paste

A human being manuals examines the webpage and copy paste required data. Failure rate is almost zero in this traditional technique

2.Text pattern matching

An approach to extract information from webpage using UNIX **grep** command or regular expression matching facilities of programming languages

3.Using web-scraping framewors/softwares

These frameworks/softwares gives the facilities to create web scraping bots(spiders/crawlers) to scrape data from webpages.These spiders try to recognize structure of the page automaticaly and extracts the data and gives the facility to process and store them in specific format

4.Computer vision webpage analysis

Employing the features of machine learning and computer vision to idenetify and extract data from webpage. This can be described as computerised version of human copy-paste with a failure rate ¿0

Proposed Methodology

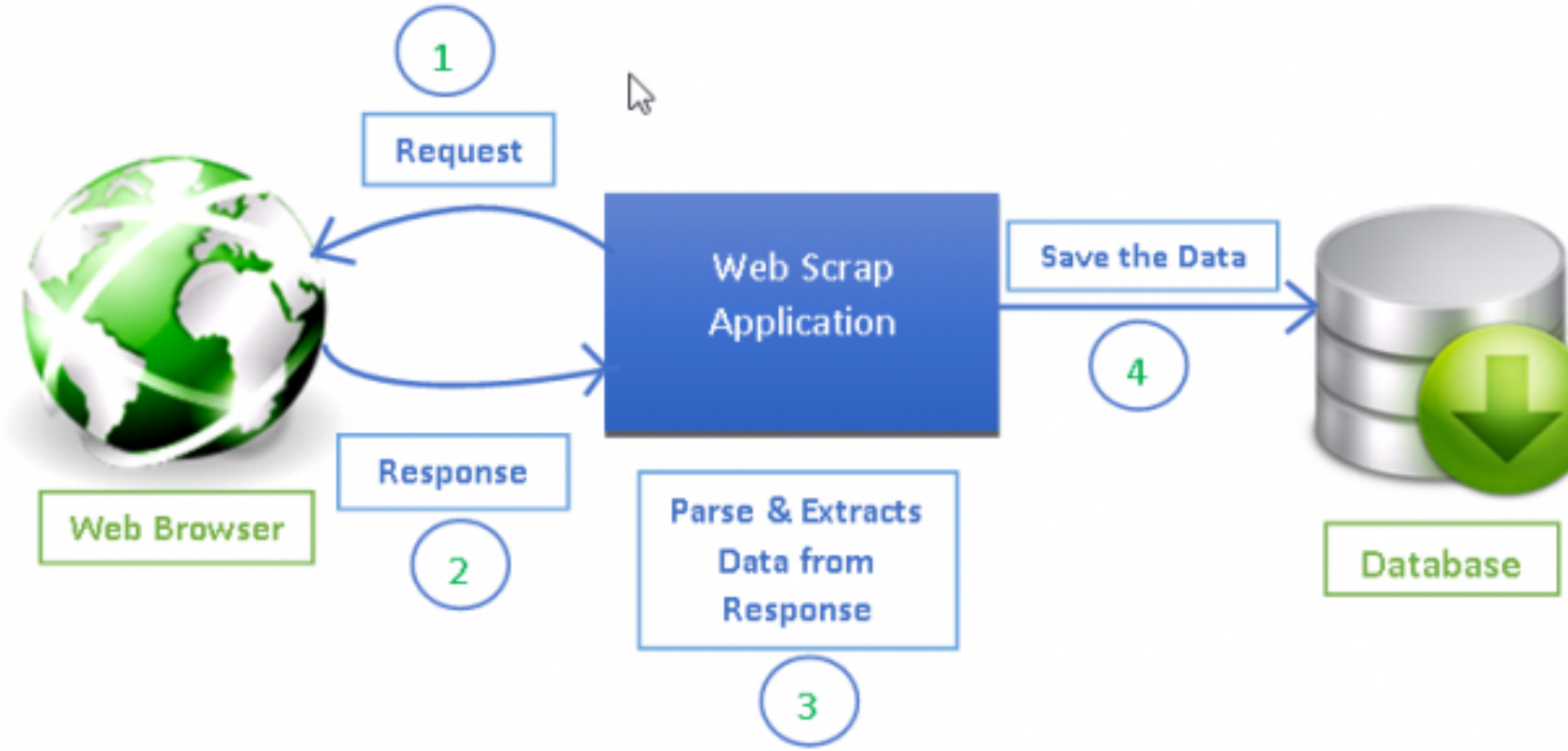
Using web-scraping framework

The proposed methodology is to scrape webpages using a software/framework. Here we are using "Scrapy", a framework written in python to generate and deploy spiders to extract data from a webpage

Why Scrapy ?

1. Faster and automated way to scrape webpages
2. Easy to use & implement
3. Have selectors to specify data field needs to be extracted
4. Have pipelines to refine extracted data

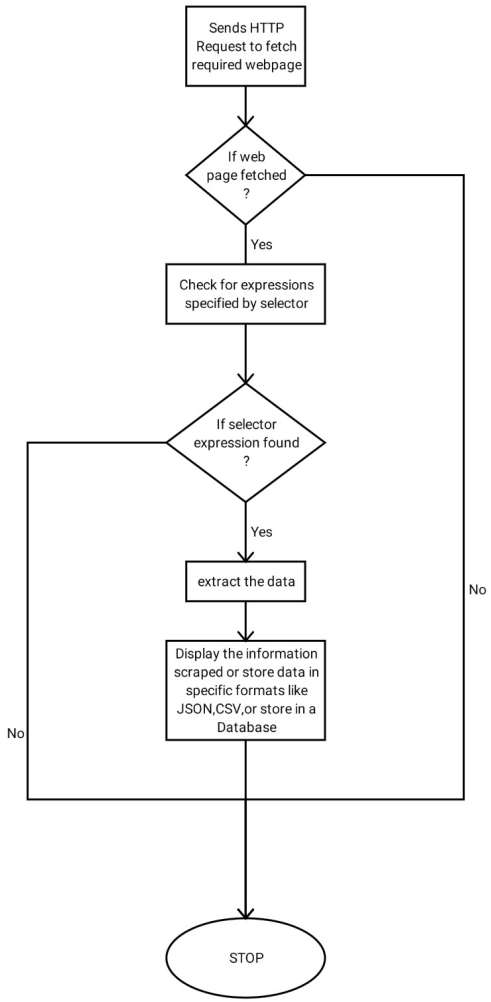
How Scrapy works



Barriers to Web Scraping

- Not all web pages allows data scraping
- Captchas block the scraper from proceding further
- Frequent structural changes made to website
- IP blocking

Implementation Algorithm



Spell Checker

1. Aim
2. Introduction
3. Existing Algorithm-1
4. Existing Algorithm-2
5. Proposed Method
6. Proposed Architecture
7. Proposed flowchart
8. Result
9. Conclusion

AIM

writing a spell checker program to check the correctness of a given document

INTRODUCTION

We have to create a user program that checks the correctness of a document and makes neccesary changes if needed.Spell-checking features are often embedded in software or services, such as a word processor, email client, electronic dictionary, or search engine.

EXISTING ALGORITHM PAGE 1-Grammerly

Grammarly is an AI powered tool.

1. So it's basically a computer program or algorithm which studies and compares your sentence against hundreds of similar ones that are listed in their database.
2. Grammarly, as mentioned earlier, is an AI powered tool that weighs your writing by comparing it against authentic rules and sentence patterns
3. Grammarly's algorithms flag potential issues in the text and suggest context-specific corrections for grammar, spelling and usage, wordiness, style, punctuation, and even plagiarism using patterns in its database.

Intro
Statement
Approaches
Flow Chart
Algorithm
Relevance
Applications

EXISTING ALGORITHM PAGE 2-GINGER

Ginger Software is an Israeli start-up company.

1. Ginger Software that has developed language enhancement technology that uses statistical algorithms in conjunction with natural language processing
2. Ginger Software uses patented software algorithms that deal with natural language processing. The company claims that the algorithm allows it to correct the written sentences with relatively high accuracy (eliminating up to 95 percent of writing errors) compared to standard spell checkers.
3. Its unique algorithm allows the software to understand the context of the sentence rather than correcting based solely on a word.
4. The software operates on the logic of sentence context in addition to the memory of a database of words.

Intro
Statement
Approaches
Flow Chart
Algorithm
Relevance
Applications

PROPOSED METHODOLOGY

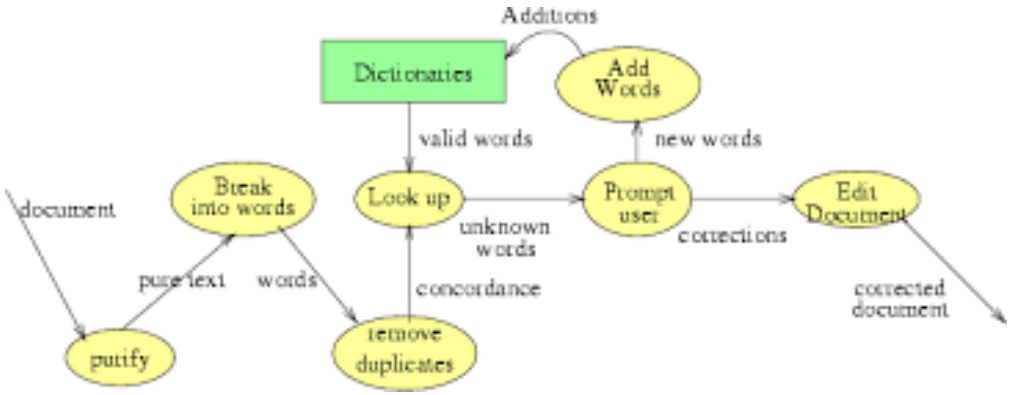
To do the User Authentication we use

1. online Dictionary
2. where we cross check word by word with its counterpart in the dictionary using the first letter as hint.

Proposed algorithm

1. Given document is broken down into words and duplicate words are removed.
2. each word is checked up in the dictionary module and if found is kept or else added to the dictionary if needed or changed by the user if not needed.
3. the final edition is done by the user and recorrected document is received

FLOW CHART


Additions
Dictionaries
Add Words
valid words
new words
document
Break into words
Look up
Prompt user
Edit Document
unknown words
corrections
pure text
words
concordance
corrected document
purify
remove duplicates

PROPOSED ALGORITHM

1. Input URL from user using chatbot
2. run scrapy process using os library
3. scrap main page of website. Output scraped data to data.json file
4. If user wants to follow to assoiciated links .scrape the follow links too.
5. After scraping, open data.json using file handler.
6. Check for spelling and offensive words in json file
7. If found listed to chat bot as website analysis report
8. end

CONCLUSION

A bot was created which analyses and monitors a website periodically. It uses help of three sub programs that is chat bot , spell checker which also checks for offensive words and a scrapy tool which will scrap the desired web page These sub programs is later on implemented together to create bot which creates a bot that that checks for the content of a website given and detects for spelling and offensive words if present in it and also makes sure to notify of any outside activity.

Initially chat bot was executed and acts as user friendly interface and can be used to communicate to the bot and later on used to analyze the website which is our main concern. Th chat bot makes sure to pick up the URL of the website needed and is given to spell checker to be executed. The chat bot can also generate response on questions asked by the user based on keyword matching.

The scrapy package used here to scrap data from desired websites. It also provides an option to follow scrapping to associated links and stores the data into a document.

Spell checker checks the output file where the scrapped data is stored and checks for spelling if present and also checks for offensive words If found the words are returned to chat bot and displayed as the analysis report . It also checks for any external activities.