A logo on a black background

AI-generated content may be incorrect.

TasteOfHome Web Crawler

Team Members:

Mariam Elewa - 249669

Shaza Ashraf - 238765

Omar Sakr - 234809

Ahmed Ashraf – 248917

Momen aly -232889

**Course: Information Retrieval**

TasteOfHome Web Crawler – Project Documentation

**Introduction:**

The TasteOfHome Web Crawler is a Python-based project developed to automatically extract and analyze recipe data from the TasteOfHome website.   
It utilizes web crawling and scraping techniques to gather structured data and presents insights using a visual dashboard built with Streamlit.  
  
The crawler extracts recipe titles, ingredients, instructions, and metadata, enabling comprehensive analysis. The goal of the project is to provide an automated,   
scalable, and user-friendly solution for recipe data collection and visualization.  
  
this tool combines automation, data analytics, and web technologies into a unified framework.

**System technologies:**

The project consists of several Python modules organized for modular functionality:  
  
app.py: Streamlit dashboard for interactive data analysis.  
content\_extractor.py: Extracts recipe content using Selenium and BeautifulSoup.  
crawlability\_checker.py: Analyzes robots.txt for crawlability permissions.  
sitemap\_parser.py: Extracts recipe links from the sitemap.  
recipes.json: Stores the structured recipe data.  
requirements.txt: Lists Python dependencies.  
  
Technologies Used:   
Python, Selenium, BeautifulSoup for web scraping.  
Streamlit and Plotly for visual dashboards.  
JSON for data storage and exchange.

**Result:**

After crawling the TasteOfHome site, the crawler extracted over 100 unique recipes.  
  
Insights Extracted:   
Most common ingredients included garlic, onion, butter, and chicken.  
Recipe complexity was evaluated based on the number of ingredients and instructions.  
Average success rate of crawl: 94.2%  
Visualizations showed recipe trends, popular ingredients, and crawl performance.  
  
The dashboard allows filtering recipes, analyzing keyword frequencies, and measuring recipe difficulty via custom metrics.

**Conclusion:**

The TasteOfHome Web Crawler proved effective in automating recipe data collection and providing analytical insights.  
  
Conclusions:   
The site supports crawling with minimal restrictions.  
JavaScript-based rendering was handled using Selenium.  
Data is displayed in an interactive and user-friendly dashboard.  
  
Recommendations:   
Implement caching or asynchronous scraping to improve performance.  
Extend support to other recipe sites.  
Deploy the dashboard to a cloud platform for wider access.