Web Scrappy: From Concept to Completion

Project Timeline Table

Week	Dates	Tasks
Week 1: Project Planning & Setup	Feb 20 – Feb 26, 2025	- Define project objectives & scope - Research web scraping techniques - Set up project structure - Install required libraries (Scrapy, Selenium, Pandas, etc.)
Week 2: Data Input & URL Processing	Feb 27 – Mar 5, 2025	- Develop `excel_operations.py` for reading/writing Excel files - Implement `url_processing.py` for extracting and processing URLs - Test URL handling and filtering
Week 3: Web Scraping Implementation	Mar 6 – Mar 12, 2025	- Configure Scrapy project settings (`settings.py`, `items.py`) - Implement `content_spider.py` for extracting webpage content - Develop middleware for request handling - Integrate Selenium for dynamic content extraction
Week 4: Data Storage & API Integration	Mar 13 – Mar 19, 2025	- Develop `file_operations.py` for saving extracted data - Implement `api_operations.py` for API interactions - Secure API keys with `key_manager.py`
Week 5: Testing & Debugging	Mar 20 – Mar 26, 2025	- Perform unit testing using `pytest/unittest` - Handle errors & exceptions (e.g., timeouts, missing data) - Implement `logging_config.py` for structured logging

While the table provides a structured breakdown of weekly tasks, a **Gantt Chart** helps visualize the overall timeline and dependencies between tasks. Below is a progress tracker to gauge the estimated development timeframes effectively.

Tracking Development Progress

Task	Week 1	Week 2	Week 3	Week 4	Week 5
Project Planning & Setup	■				
Data Input & Processing		■			
Web Scraping Implementation			■		
Data Storage & API Integration				■	
Testing & Debugging					■