



CROWDSOURCING OF RECORDS

Team Name:Runtime Terror





INTRODUCTION

- DATA CROWDSOURCING FROM THE INTERNET USING PHONE NUMBERS OR EMAIL ADDRESSES INVOLVES TWO SEPARATE PROCESSES THAT MUST BE AUTOMATED AND COMBINED.
- THE FIRST STEP IS TO SEARCH THE INTERNET FOR INFORMATION RELATED TO THE EMAIL ADDRESS OR PHONE NUMBER IN PUBLIC DOMAINS, AND THE SCRAPE THOSE PUBLIC DOMAINS WITH WEBSCRAPERS ENABLED WITH PROXY CHAIN TO OBTAIN DETAILED INFORMATION.

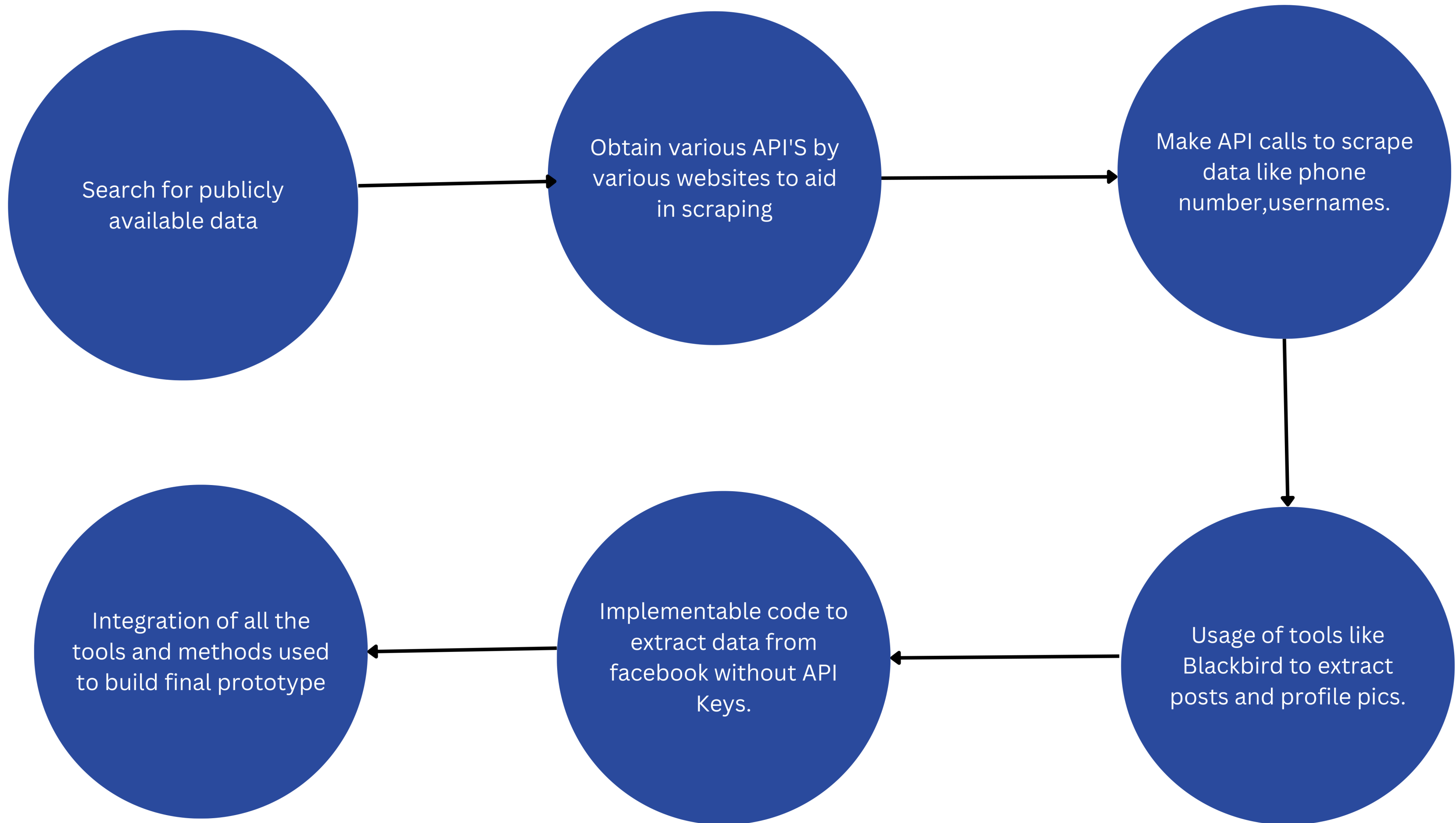


WORKING RULE

- USAGE OF A KALI LINUX TOOL CALLED BLACKBIRD TO SEARCH FOR USERNAMES OR PUBLIC NAMES AVAILABLE IN VARIOUS WEBSITES.
- SCRAPE PUBLICLY AVAILABLE DATA USING VARIOUS API'S PROVIDED BY VARIOUS WEBSITES.
- INSTALOADER TOOL TO SCRAPE POSTS OF DIFFERENT ACCOUNTS ON INSTAGRAM GIVEN THE USERNAMES.
- OBTAINED DEVELOPER KEYS FOR TWITTER TO SCRAPE PUBLICLY AVAILABLE POSTS AND TWEETS.

WORKING RULE

- USED TF-IDF COSINE SIMILARITY TO OBTAIN TOP MATCHES FOR A CERTAIN USER IN ORDER TO NARROW DOWN SEARCH.
- TRACK DOWN PHONE NUMBERS OF USERS USING TRUECALLER API.
- INTEGRATION OF VARIOUS METHODOLOGIES TO PRODUCE FINAL PROTOTYPE.
- USAGE OF PROXY CHAINS TO RANDOMIZE IP ADDRESS AND PREVENT BLOCKAGE OF IP ADDRESSES.
- DATA EXTRACTION LIKE POSTS, VIDEOS AND USERNAMES FROM FACEBOOK BY GIVING USERNAMES.



BUSINESS FEASIBILITY

- The application makes use of open-source libraries and tools, hence it requires less investments.
- We can get developer account from Meta which helps scrape large amounts of data for free by agreeing with the usage policies of websites.



THANK YOU