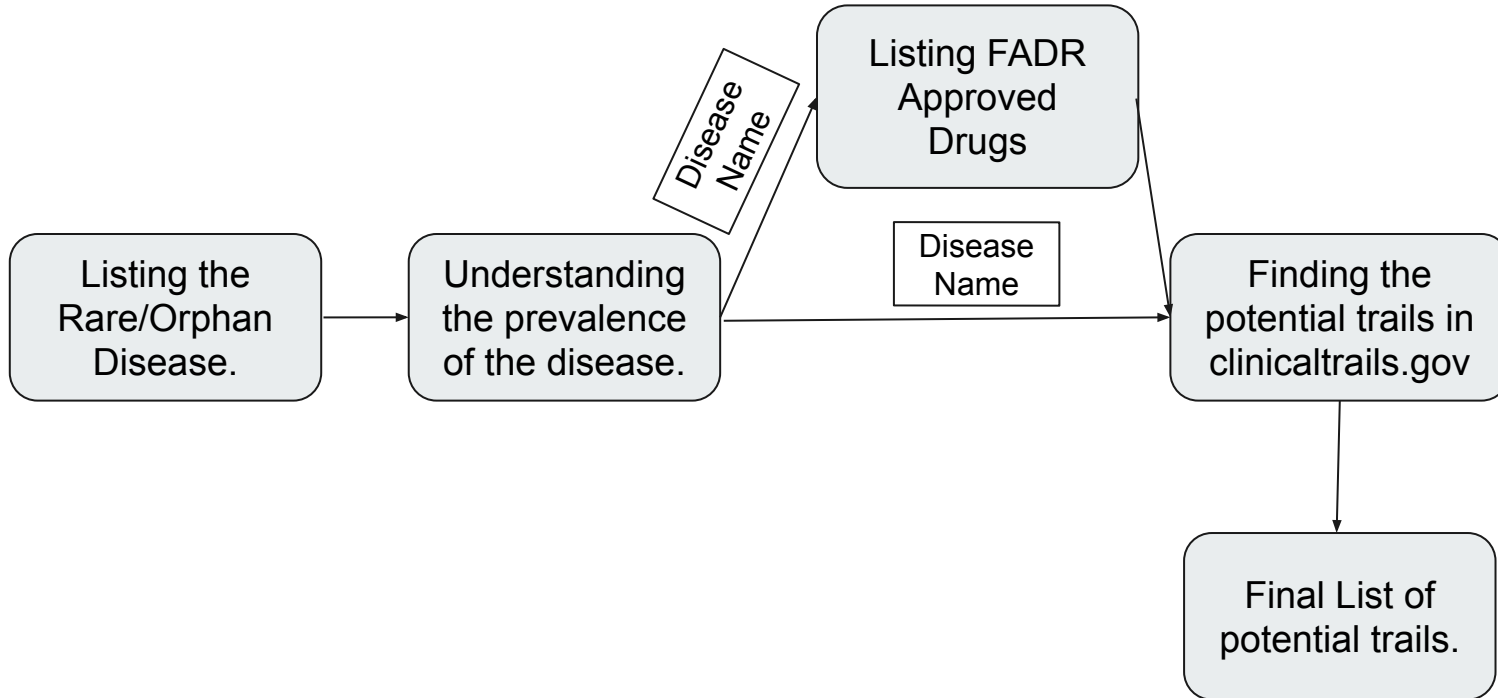




Workflow Assignment

Venkatesh Surampally

Executive Summary



Getting the List of Rare Orphan Diseases

- List of rare orphan diseases can be retrieved by scrapping these two websites.
 - Orphadata
 - Rare Diseases (NORD)

The logo for Orphadata, featuring the word "Orphadata" in a sans-serif font. The letters "o", "r", "p", "h", "a", and "d" are black, while "a", "t", "a" are colored in orange, yellow, and green respectively.

Implementation Output

	Orpha Code	Disorder Name	Summary
0	166024	Multiple epiphyseal dysplasia, Al-Gazali type	A rare primary bone dysplasia characterized by...
1	58	Alexander disease	A rare neurodegenerative disorder of the astro...
2	166032	Multiple epiphyseal dysplasia, with miniepiphyse...	Multiple epiphyseal dysplasia, with miniepiphy...
3	61	Alpha-mannosidosis	An inherited lysosomal storage disorder charac...
4	166029	Multiple epiphyseal dysplasia, with severe pro...	Multiple epiphyseal dysplasia, with severe pro...
...
10670	619284	Narcolepsy	No Info
10671	619340	Inherited hematologic cancer-predisposing synd...	No Info
10672	619363	Neonatal-onset severe multisystemic autoinflam...	No Info
10673	619360	NON RARE IN EUROPE: Isolated hereditary persis...	No Info
10674	619233	Hereditary persistence of fetal hemoglobin-int...	No Info



Prevalence of the disease

- Prevalence of the disease in US can be scrapped from NORD and further text processing can lead us to have exact numbers.
- Further popularity of the disease can be retrieved from google trends.

Example Implementation output of disease prevalence.

' 47, XXY (KS) is the most common human sex chromosome disorder and occurs in approximately 1 in 500-1,000 males. It is estimated that 3,000 affected boys are born each year in the United States.'

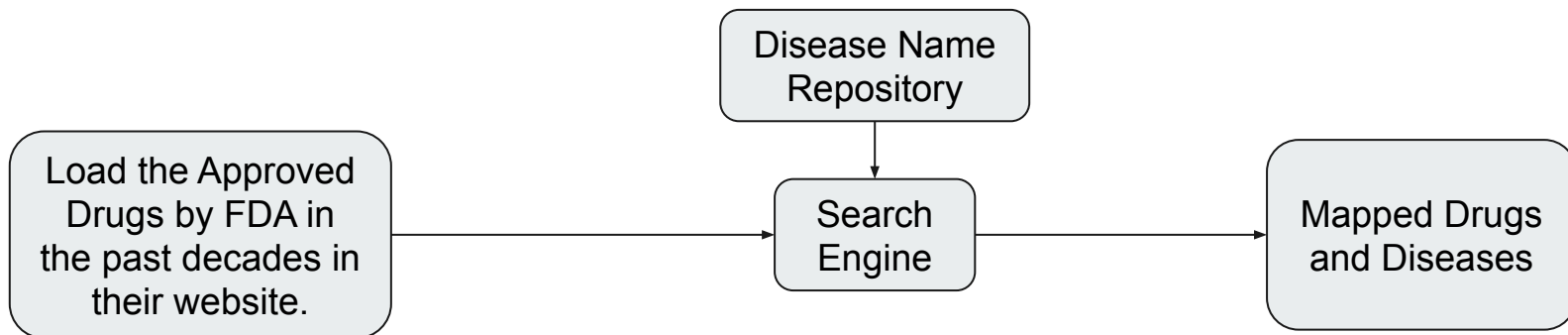
' AGL and APL generally affect women more than men, although this may be due in part to ascertainment bias because women tend to be more severely affected and more easily recognized. APL has been reported in approximately 250 individuals with a male to female ratio of 1:4. It has been reported in individuals of various different ethnicities. AGL has been reported in approximately 100 individuals with a male to female ratio of 1:3. Most cases have been reported in Caucasians. LD-HIV is estimated to affect approximately 100,000 individuals in the United States. Consistent with the increase prevalence of HIV in males, LD-HIV is also more prevalent in males.'

Popularity of example disease by google trends.

Arterial Tortuosity Syndrome isPartial		
date		
2017-09-24	0	False
2017-10-01	0	False
2017-10-08	0	False
2017-10-15	0	False
2017-10-22	0	False
...
2022-08-21	0	False
2022-08-28	0	False
2022-09-04	0	False
2022-09-11	0	False
2022-09-18	58	True

Approved Drugs by FDA

- List of all the drugs that are approved in the past 2 or 3 decades will be collected from FDA website.
- We will develop a small search engine that maps the approved drugs with the diseases in our repository.
- Finally a map of drugs and diseases will be generated.



Retrieving the Trail Data

- ClinicalTrials.gov is offering an API for retrieving the trail data.
- Available study fields are present in https://clinicaltrials.gov/api/info/study_fields_list
- Required data can be dumped in to the dataframe with the required attributes.
- Dataframe can be filtered with the following constraints.
 - i) Phase 2 or phase 3 trials that are active/recruiting/completed within 1 year
 - ii) Industry sponsored
 - iii) Interventional
- Develop a system that automates the whole process.

Implementation Example Output (Please visit the code for exact details)

	Rank	NCTId	Condition	Phase	StudyType	OverallStatus	LastKnownStatus	IsFDARegulatedDrug	InterventionType
9	10	NCT04085523	Achondroplasia	Phase 2	Interventional	Enrolling by invitation		Yes	Drug Drug Drug
11	12	NCT05246033	Achondroplasia	Phase 2	Interventional	Recruiting		Yes	Drug Drug
14	15	NCT02724228	Achondroplasia	Phase 2	Interventional	Active, not recruiting			Drug
15	16	NCT03424018	Achondroplasia	Phase 3	Interventional	Active, not recruiting		Yes	Drug
18	19	NCT03989947	Achondroplasia	Phase 2	Interventional	Active, not recruiting		Yes	Drug



Please visit the link below for automation implementation:

<https://github.com/venkatesh7979/workflowtask/blob/main/Process%20flow%20task.ipynb>

Thankyou