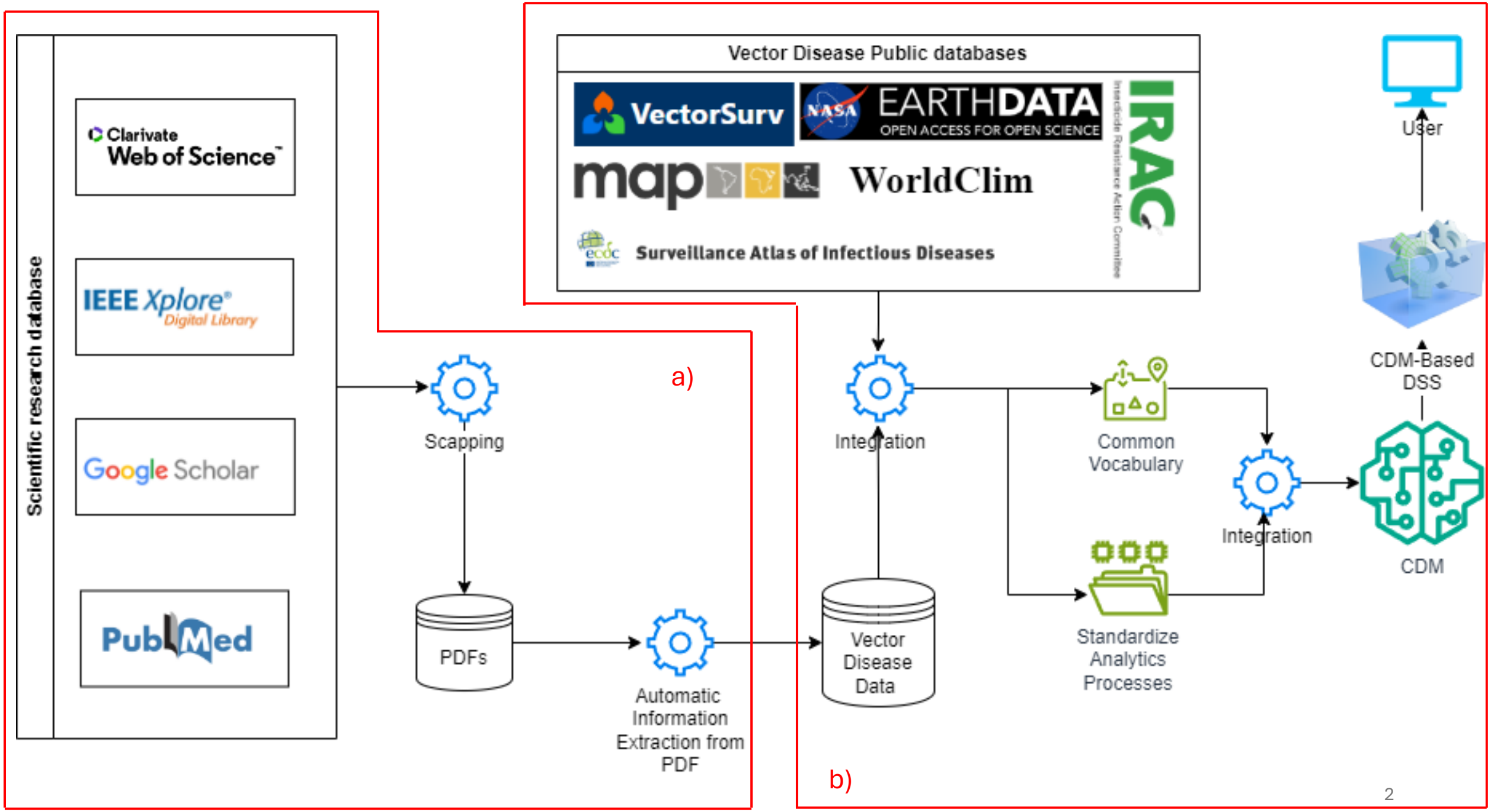
A Common Data Model for Disease vectors

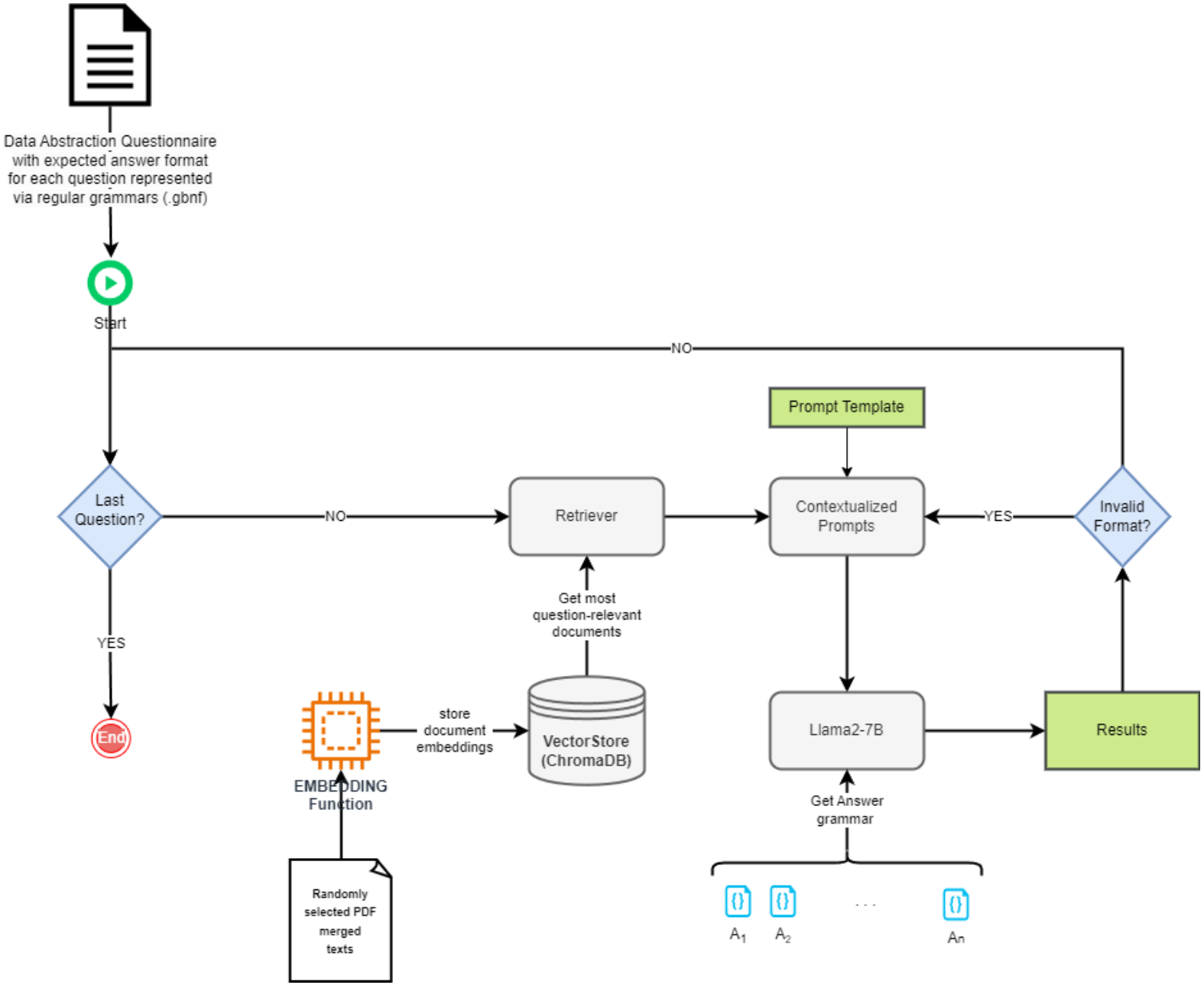
Salvador Desconsciences Atagong

Under the supervision of

Dr. Henri Tonnang, Dr. Kennedy Senagi, Pr. John Odindi


Scientific research database
Clarivate Web of Science
IEEE Xplore Digital Library
Google Scholar
PubMed
Scapping
PDFs
Automatic Information Extraction from PDF
a)
Vector Disease Public databases
VectorSurv
NASA EARTHDATA
OPEN ACCESS FOR OPEN SCIENCE
map
WorldClim
IRAC
Insecticide Resistance Action Committee
ecdc
Surveillance Atlas of Infectious Diseases
Integration
Vector Disease Data
Common Vocabulary
Standardize Analytics Processes
Integration
CDM
CDM-Based DSS
User
b)

Automatic Data Abstraction From PDF documents



Evaluation

Current approach is to compare human extraction on the same document with the models' outputs and consider the success rate.

Success rate = S / N
S: total of right answers
N: Total of questions

Mean success rate is considered for a batch of documents.

Similarity between models output and human outputs are evaluated using cosine similarity, with a threshold of 0.5.

Major Limitations

- PDFs documents preprocessing and extraction accuracy