

IRPersonalNews

Information Retrieval Personalized News System

Final report of
Andrea Guzzo
Matricola: 761818

Academic year: 2019-2020

Table of Content

Introduction



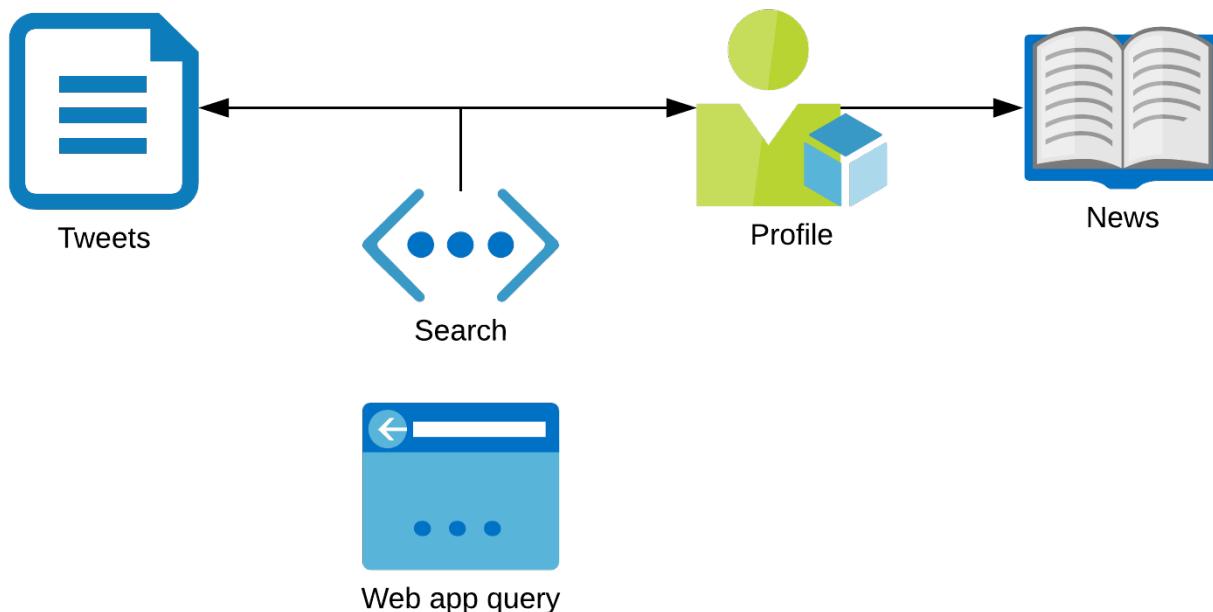
1. Introduction
2. Architecture
3. Steps: Indexing, Retrieval, Personalization
4. Conclusion

Introduction

Introduction



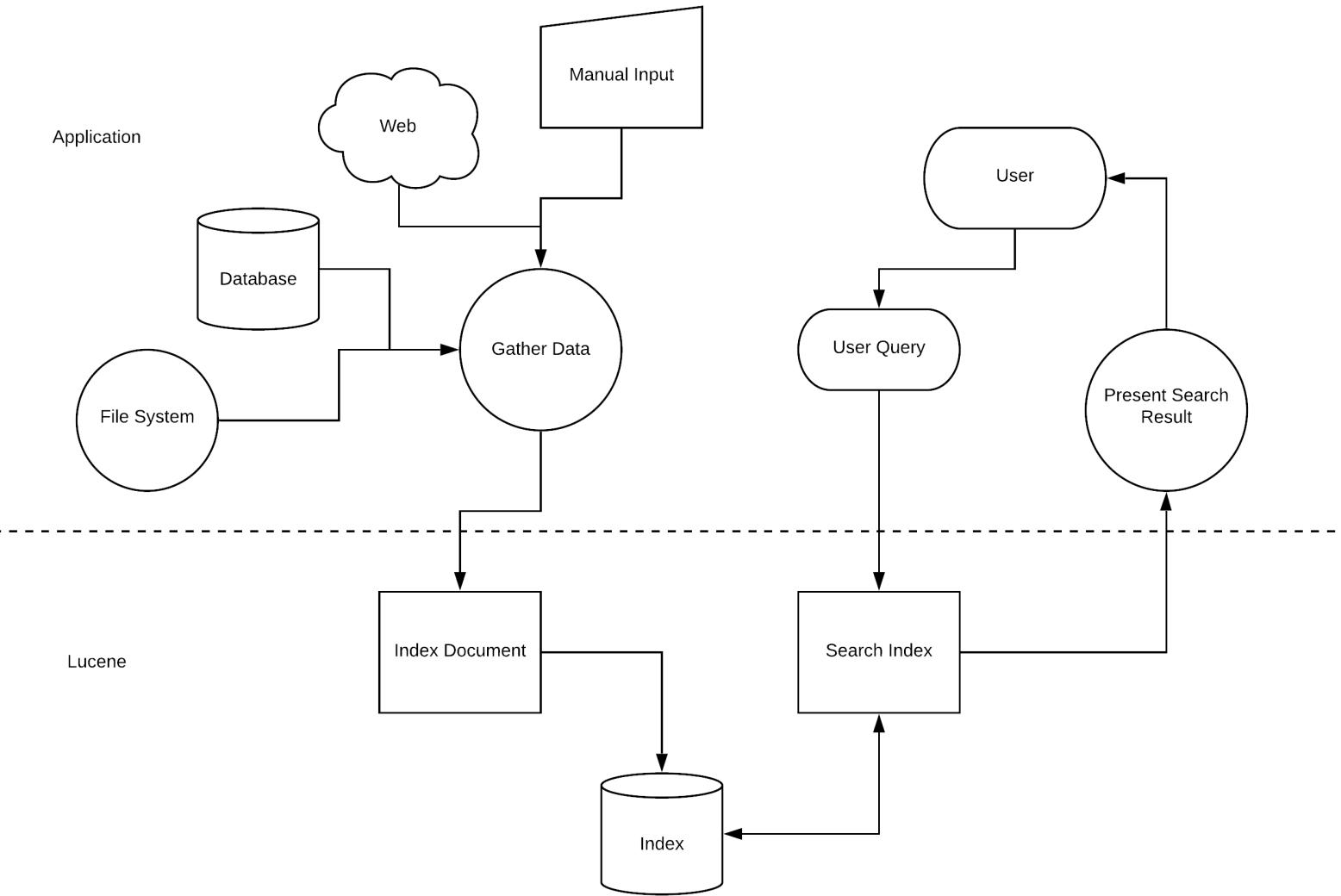
Webapp Recommended System for retrieving and searching tweets based on custom user profile built from past news reading.



<https://github.com/JeyDi/IRPersonalNews>

A Generic Recommended System

Introduction

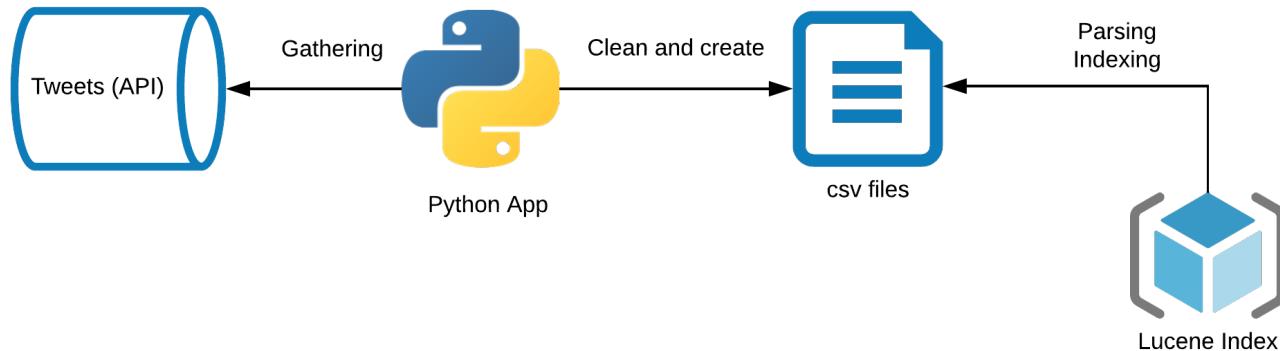


Data

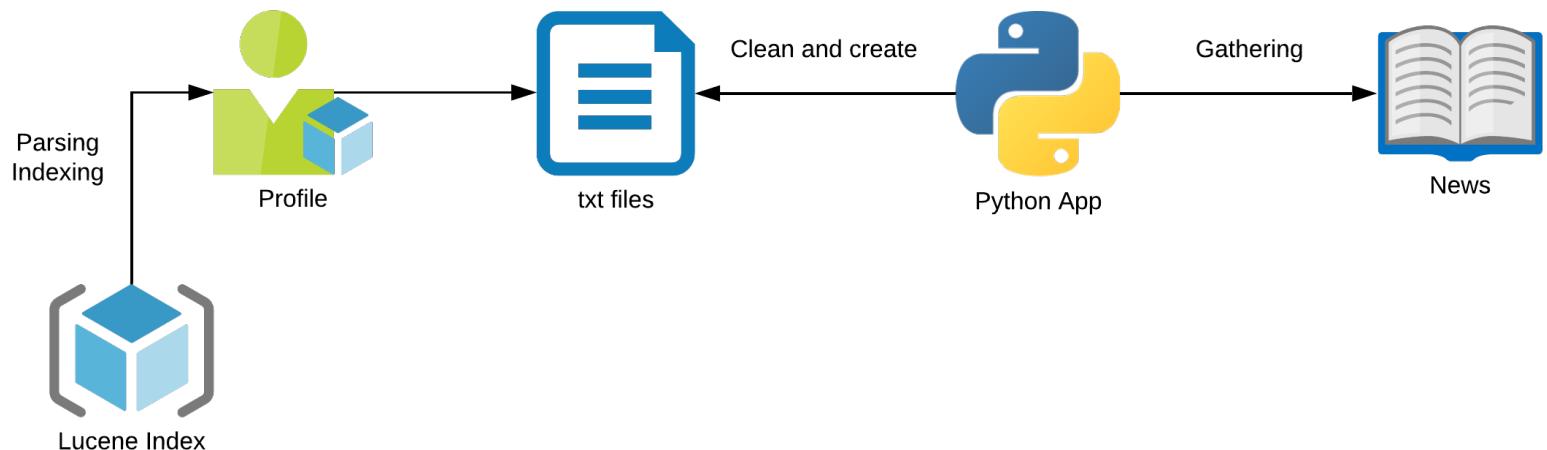
Introduction



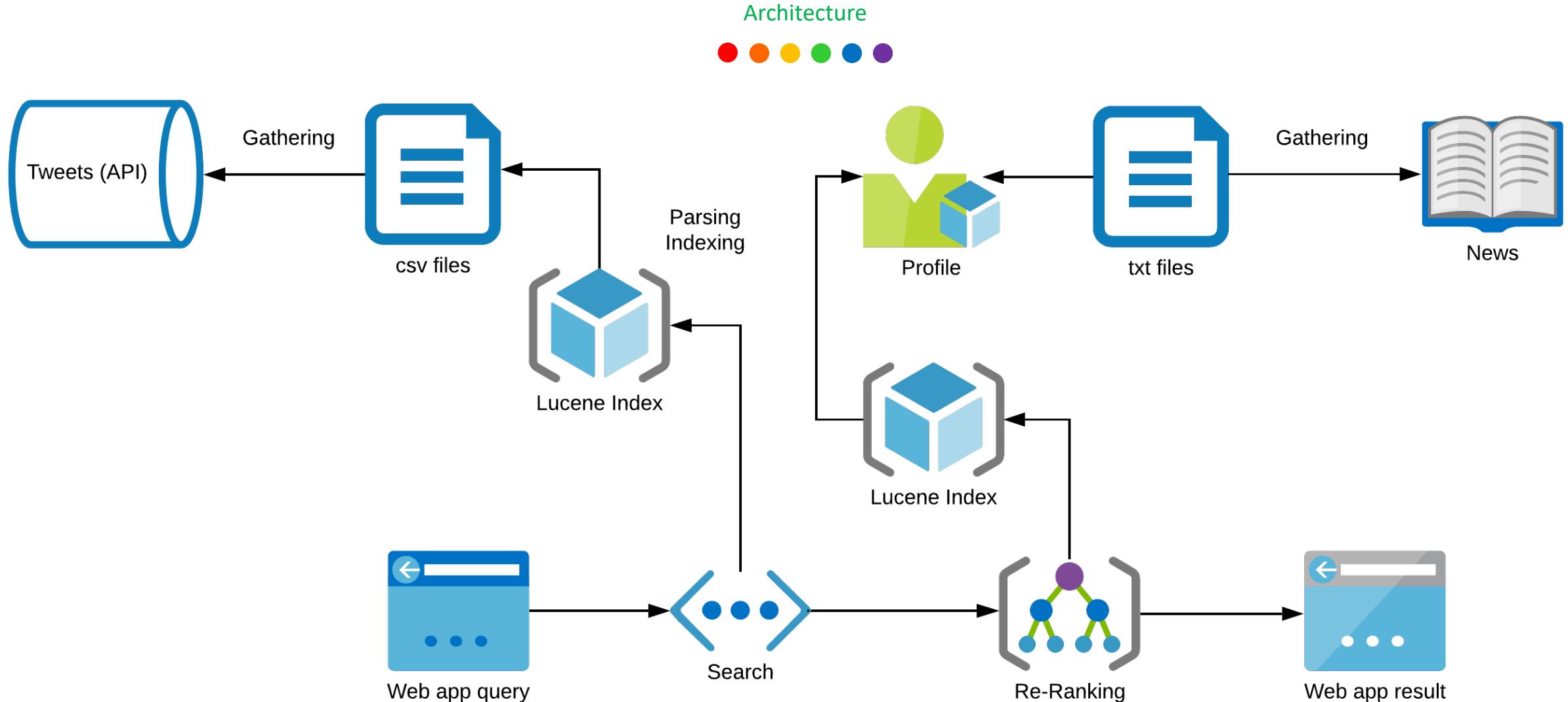
Tweets Gathering Pipeline



News Gathering Pipeline



Proposed solution



Indexing

Detailed Steps



Tokenizer

Patterns
Replacements

Case
Transformation

Stopwords
Removal

Stemming

Standard Tokenizer

Removed URLs

Lowercase

Porter Stemmer

Unicode UX29 Text

Retrieval and Personalization

Detailed Steps



Retrieval



Tweets

Popularity boosting

Recent documents boosting

Okapi BM25

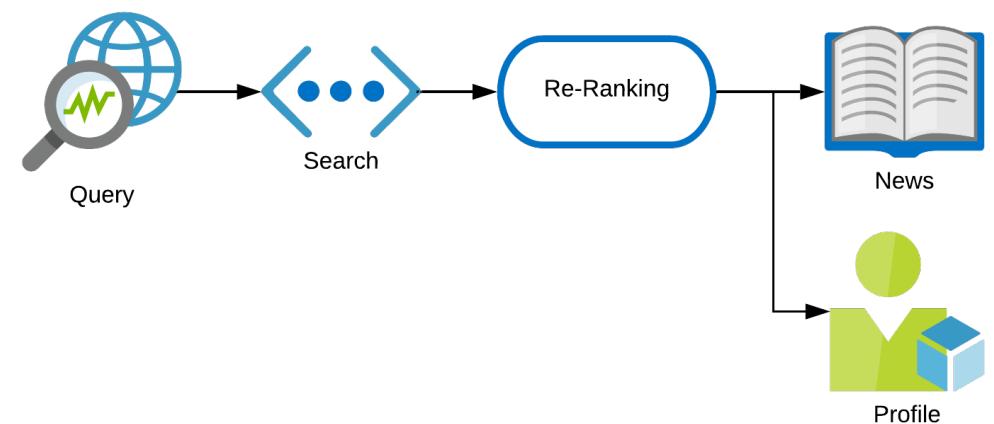
$$score(D, Q) = \sum_{i=1}^n IDF(q_i) * \frac{f(q_i, D)(k_1 + 1)}{f(q_i, D) + k_1 * \left(1 - b + b * \frac{|D|}{avgdl}\right)}$$

Personalization



News

Cosine Similarity



Webapp

Conclusions



IRPersonalNews

POLITICS SPORT SCIENCE TECHNOLOGY LIFESTYLE

space

Personalization On/Off

RANK: 1

Top 10 pictures taken by the Hubble Space Telescope in its 25 years: <http://ow.ly/M2gwX> #science #TBT

2015-04-23 23:02:03+00:00

RETWEETS: 239

RANK: 2

Watch the first artificial gravity experiment: <http://ow.ly/LtOnm> #science #space

2015-04-11 23:03:12+00:00

RETWEETS: 167

RANK: 3

Meteorite chemicals may have started life on Earth - and space: <http://ow.ly/LIlsq> #science #cosmos

2015-04-17 04:03:06+00:00

RETWEETS: 99



intercooler.js

AJAX With Attributes: There is no need to be complex

Maven™



Next Steps

Conclusions



- Automatization of the pipelines
- Much more data
- Custom profile creation
- Better UI
- More robust similarity for personalization

Thanks!

