

|  |
| --- |
| Automated Data Filtering |
|  |
| January 3  A joint paper by JP Tokyo & Co and MRIIRS  Authored by: Nishant Choudhary Reviewed by: Rajesh Nath, Mrs. Bindiya Ahuja |

**Contents**

**1.) Introduction…………………………………………………………………………………………….4**

**2.) Description, Platforms and Technologies…….……………………………………………5**

**3.) Findings and verification….……………………………………………………………………….6**

**4.) About……………………………………………………………………………………………………….7**

**5.) Project artifacts….…………………………………………………………………………………….**

**6.) References………………………………………………………………………………………………9**

Introduction

In this white paper we are providing information about the algorithms to automate the data filtration process. Data Filtration is used to separate the searchable information like images, headings, dimensions etc. from the raw information that contains all kinds of data. In this process we make use of technologies like Python, ElasticSearch, ES ingest attachment plugin,. Python is high level general programming language which is well known for its high readability and similarity to the English language. The Python libraries are so vast that it provides a tool or module for almost everything. Elasticsearch is a highly scalable open-source full-text search and analytics engine. It allows you to store, search, and analyze big volumes of data quickly and in near real time. It is generally used as the underlying engine/technology that powers applications that have complex search features and requirements.

Description

Platforms and Technologies

* ElasticSearch 6.4.2
* Python 2.7.15
* ElasticSearch Ingest Attachment Processor Plugin

Findings & Verification

The project was under constant guidance under of our mentors. Updates were provided weekly with the work done to the mentors from MRIIRS and JP Tokyo and co. Also group member reviews were taken into consideration.

# References

https://www.elastic.co/guide/en/elasticsearch/guide/current/index.html