



Innovative Artificial Intelligence Solutions

Xen Document360.AI - an Artificial Intelligence (AI), Machine Learning (ML) and Deep Learning based solution for text processing and text extraction from images and documents.

Overview

In recent years, OCR (Optical Character Recognition) technology has been applied throughout the entire spectrum of industries, revolutionizing the document management process. OCR has enabled scanned documents to become more than just image files, turning into fully searchable documents with text content that is recognized by computers. With the help of OCR, people no longer need to manually retype important documents when entering them into electronic databases. Instead, OCR extracts relevant information and enters it automatically. The result is accurate, efficient information processing in less time.

By integrating AI and Machine Learning based technologies, a number of important business problems can be solved very effectively without human intervention. They include scanning of boxes of documents and storing retrieved textual and numeric data them in database. It allows to effectively keep track of records, simplifies data collection and following analyses.

Goals of Xen Document360.AI Solution

1. AI based integrated solution to extract textual and numeric information from PDF and image files in the form of text and present this information in a convenient (e.g. tabular) format.
2. Organize storage of this information in a database for following quick extraction and possible analytics.

Web: www.xen.ai
Email: support@xen.ai

USA
501, Gibson Dr, #2624
Roseville, California - 95678, USA
Phone: +1 408 221 6976

INDIA
ES 11, Heavenly Plaza, Kakkanad
Kochi – 682021, Kerala, India
Phone: +91 701 261 9339



Innovative Artificial Intelligence Solutions

Xen Document360.AI Solution Overview

1. Usage of Open Source tools.

To minimize dependence on vendors, we prefer to use open source tools like Python, OpenCV and machine learning libraries.

2. Parallel processing to speed up.

In a multi-threading setup, a single server of 15-20 cores with SSD storage could process about 1 million images per day. For this purpose, we can rent a computing cluster and pay for just execution time.

3. Building pipeline for real world applications.

Our OCR architecture consists of 4 main blocks:

a) Image Corrector:

- Prepare the input image in order to do better text recognition in OCR component (e.g. removal of borders, background cleaning, text segmentation).

b) Optical Character Recognizer:

- Implementation of state-of-the-art technique used in OCR (Tesseract, LSTM, etc.)
- Tesseract may require training for recognizing new fonts or hand-written texts.

c) Text Processor & Corrector:

- Generated text needs post-processing in order to extract important fields (e.g. spell-checking, use of Regex and text processing libraries).

d) Data Storage and Insight Generation:

- Extracted fields can be populated in Database (unstructured to structured data conversion).
- It will augment the features/variables and improve the data quality.

Web: www.xen.ai
Email: support@xen.ai

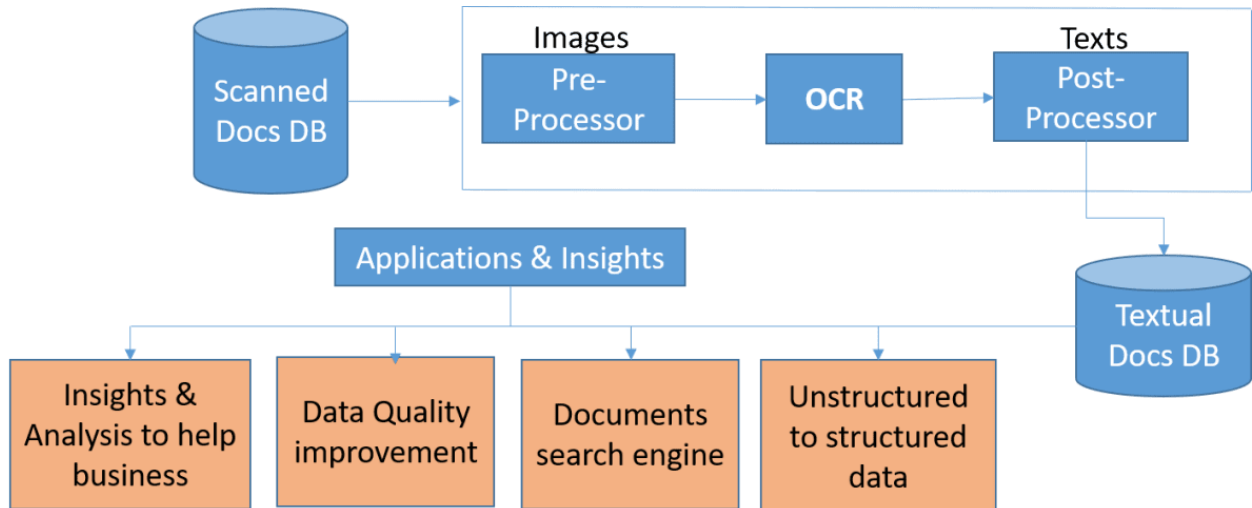
USA
501, Gibson Dr, #2624
Roseville, California - 95678, USA
Phone: +1 408 221 6976

INDIA
ES 11, Heavenly Plaza, Kakkanad
Kochi – 682021, Kerala, India
Phone: +91 701 261 9339

- Insight generation to help business.
- Can be utilized for creating a documents exploration system.

Xen Document360.AI Solution Architecture

A basic architecture of the end-to-end application is given below:



Example of document search

Web: www.xen.ai
 Email: support@xen.ai

USA
 501, Gibson Dr, #2624
 Roseville, California - 95678, USA
 Phone: +1 408 221 6976

INDIA
 ES 11, Heavenly Plaza, Kakkanad
 Kochi – 682021, Kerala, India
 Phone: +91 701 261 9339



Innovative Artificial Intelligence Solutions

Type your request here:

age of revolution

Results

Documents found

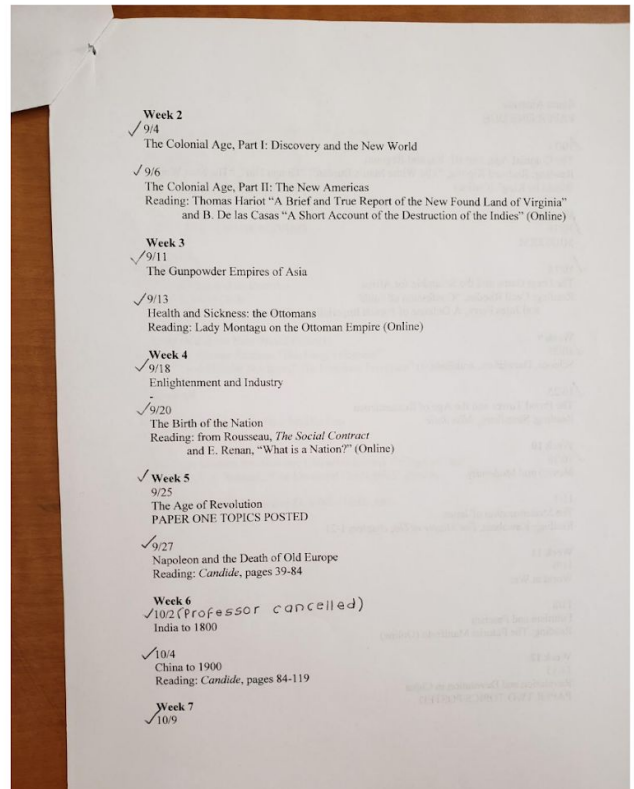
document_name	matching_score
12.jpg	1.404
11.jpg	1.271
14.jpg	1.012
16.jpg	0.755
10.jpg	0.502
17.jpg	0.501

Document to show:

12.jpg

6 documents found

Original document



Web: www.xen.ai
Email: support@xen.ai

USA
501, Gibson Dr, #2624
Roseville, California - 95678, USA
Phone: +1 408 221 6976

INDIA
ES 11, Heavenly Plaza, Kakknad
Kochi – 682021, Kerala, India
Phone: +91 701 261 9339



Innovative Artificial Intelligence Solutions

Xen Document360.AI Solution Features

Quick and effective digitization of paper documents.

Xen Document360.AI solution allows to convert tons of paper documents of various formats to digit, and makes them available online. It saves money used for paper storage and text replication.

Documents storage for fast access.

Use of modern document databases allows to store countless texts and then quickly access them using topics, keywords, etc. It will dramatically reduce time for finding a required information.

Insight generation.

Xen Document360.AI allows to effectively group and then summarize information stored in documents. It may significantly boost business efficiency.

Key Benefits of Xen Document360.AI Solution

The suggested solution can be helpful in many fields, for example, education, healthcare, finance, and government agencies. Possible positive outcomes include:

- Countless paper documents would be converted to digit and be available online.
 - In education, it will save money for students and allow knowledge to be shared.
 - For many businesses, one can easily keep track of all financial records.
 - In government, legal, healthcare agencies, it simplifies data collection and their future analysis, among other processes.
-

Web: www.xen.ai
Email: support@xen.ai

USA
501, Gibson Dr, #2624
Roseville, California - 95678, USA
Phone: +1 408 221 6976

INDIA
ES 11, Heavenly Plaza, Kakkanad
Kochi – 682021, Kerala, India
Phone: +91 701 261 9339



Innovative Artificial Intelligence Solutions

- Overall, it maximizes a business performance, and significantly boosts efficiency of the information accumulated in those documents.

Web: www.xen.ai
Email: support@xen.ai

USA
501, Gibson Dr, #2624
Roseville, California - 95678, USA
Phone: +1 408 221 6976

INDIA
ES 11, Heavenly Plaza, Kakkanad
Kochi – 682021, Kerala, India
Phone: +91 701 261 9339