THE
DOWNLOAD

eClerx

Welcome to the inaugural issue of The Download, eClerx's Technology Newsletter.

The Download is a semiannual digest of timely, must-read posts by our SME's and analysts that provides an in-depth look at the latest trends making headlines within our business.

At eClerx many of the technology products and workflows we have developed are a direct result of customer engagements. Having a comprehensive understanding of our clients' processes enables us to develop "intelligent products" which are embedded in the solutions we offer.

In this edition you will be exposed to three cases in which we used Kaiza, our AI framework, to conduct Text Mining, Image Processing and Competitive Intelligence. Each of these use-cases delivered benefits across productivity, helped us as well as the client to garner new insights and even laid the foundation for new business models.

We hope you enjoy this issue and are eager to hear your feedback.

Sanjay Kukreja

HEAD OF TECHNOLOGY
www.eclerx.com
Twitter: @sanjaykukreja, @eClerx

Table of Contents

eClerx

Kaiza 101

Our technology products journey began with building workflow tools to help process information in a rule based manner. Today, our AI based framework, Kaiza, governs most of the tech related work we deliver.
Kaiza is an AI-based framework for enabling digitization and information processing. The framework covers automation, as well as data and infrastructure and is modular in nature. The framework leverages the domain knowledge we have acquired across all the industries we serve.

As you can see from the above figure, Kaiza is a mesh that incorporates technologies from RPA to Deep Learning, and feeds on the data generated by the various processes that we run for our clients. This framework also leverages visualization techniques to help interpret the output for our clients, and set the stage for their transformation journey.

AI and digital disruption go hand in hand. I am not sure if Alan Turing would have thought about the impact of "thinking machines" when he wrote his infamous paper "Computer Machinery and Intelligence" in 1950. Today, the combination of volumes of data and high-end computing, means that machine and deep learning are becoming the norm. For eClerx, AI is a must however, we need to be sure that:

• A good use case that lend themselves to AI and benefits our client

• Data is readily (labeled data)

• Tools, techniques and wrappers for data processing and visualization exist

• Complementary skillsets that can be implemented and run as BAU
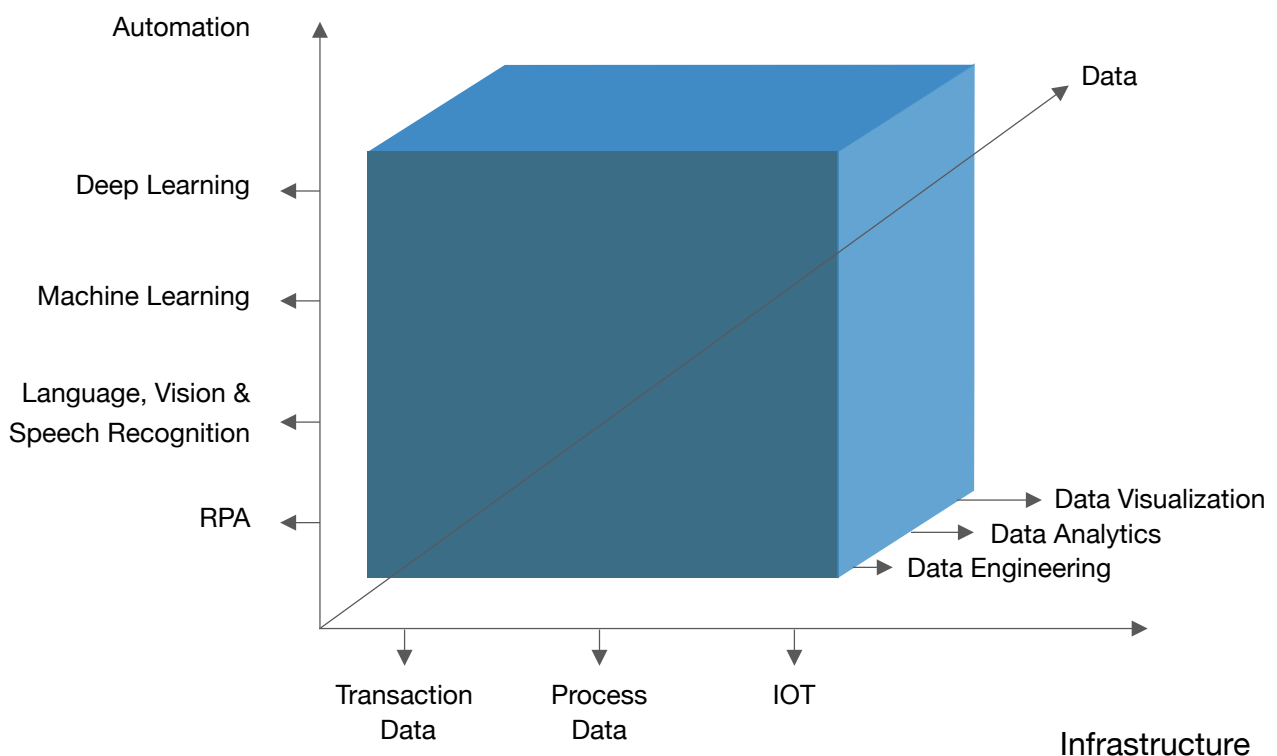
Applying Machine Learning for Strategic Competitive Intelligence

Insight from competitors' websites can change everything. In today's market, sourcing competitive data from the internet provides companies with exceptional visibility into market trends, customer preferences and intelligence on competitors' products and pricing. Companies can source material from competitors' website portal and digital channels that flows every second of each day. This access allows companies, not only extract and interpret competitive data from multiple sources, but also to synthesize and link the information to maintain a competitive edge.

The impact of machine learning, artificial intelligence and robotics is here and almost every industry is impacted. Companies now have at their disposal the full set of building blocks to begin embedding in their businesses. Add to this reductions in the cost of data storage and computing combined with advancements in Machine Learning (ML) and Artificial Intelligence (AI) and the conversation takes on a whole new dimension. According to a recent survey completed by MIT Technology Review, last year 50% of organizations planned to use ML to better understand competitors' pricing and product catalogues. This year, 48% plan to use it to gain greater competitive advantage. In the Competitive Intelligence arena, ML leverages the ability to analyse and interpret large volumes of data in a short timeframe in order to make fast, strategic decisions. ML - or "data-driven" – algorithms render machines more powerful than ever for decision making.
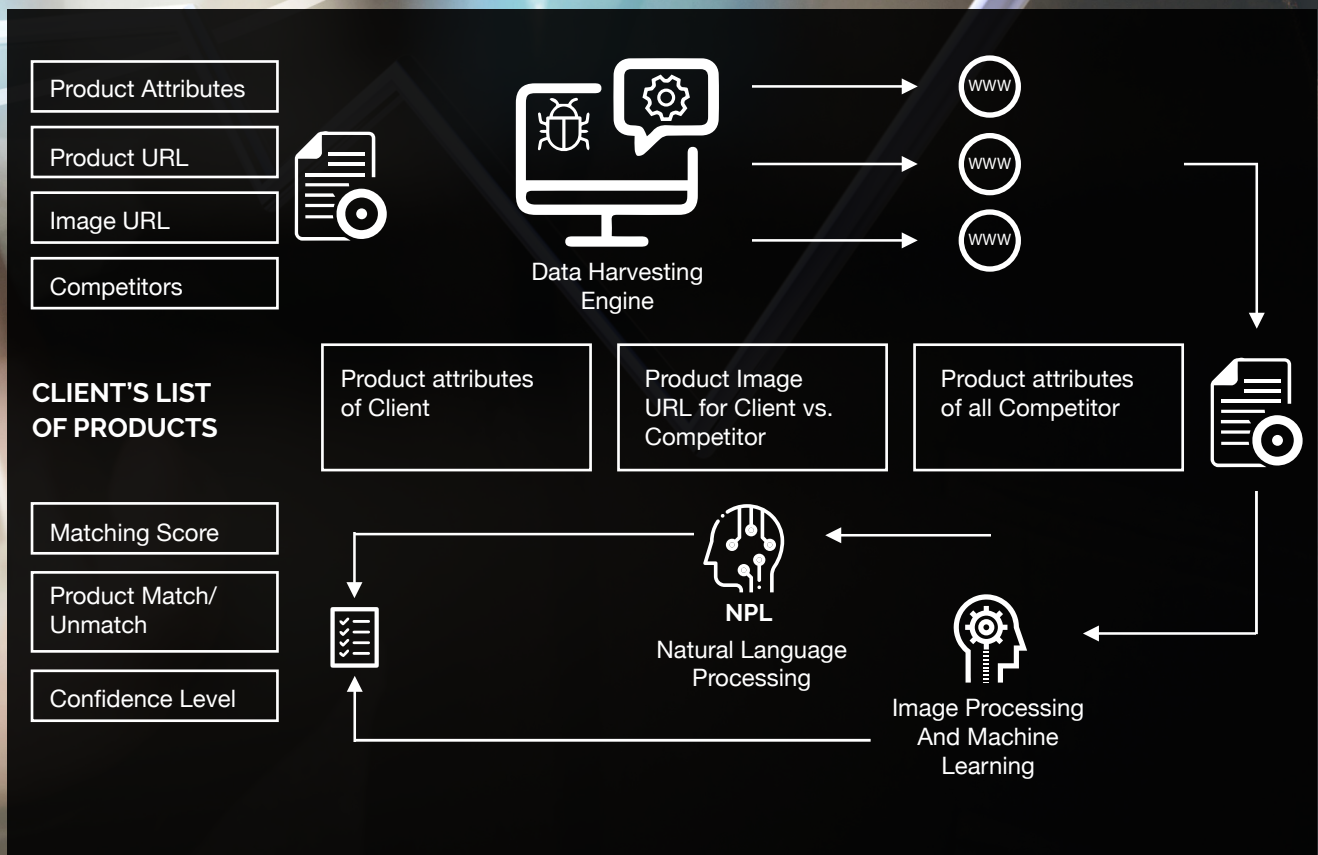
A great source of information on competitor product catalogues is a company's website. But extracting meaningful information from an unstructured source such as websites involves substantial time and effort. Using ML results in significant improvements in speed and accuracy.

At eClerx, one of the important challenges ML solves is matching products listed on competitor websites. Previously, this activity was difficult to automate through traditional programming because the attributes used for matching are vastly different for each category of products. The next level of sophistication was the use of fuzzy algorithms to match strings. This proved ineffective as the product attributes are not available in a consistent, structured format across multiple competitor websites. Large online retailers rely on manual effort to establish matches, which means a long turnaround time and matches that quickly become irrelevant as new products or variants are introduced.

By using ML on prior manually established matches, eClerx can identify relevant attributes for matching each distinct product family. Next, we extract these attributes from available product descriptions, using various techniques including Computer Vision and Natural Language Processing. If descriptions are inadequate, image matching provides a feasible solution.

ML automation helps compress the complete data analysis lifecycle by reducing the manual effort required to establish matches. This, in turn, has a direct impact on the speed of making decisions based on intelligence gathered on competitors.

Increasingly, organizations are investing in and adopting machine learning initiatives. eClerx offers strategic Competitive Intelligence solutions that combine innovative quantitative and qualitative tools allowing companies to obtain extensive competitor catalogue information and insights that can be used to improve their core businesses. Putting ML techniques to work means that insights that used to require weeks of manual work can now be derived within the same business day.

Product Attributes

Product URL

Image URL

Competitors

CLIENT'S LIST OF PRODUCTS

Data Harvesting Engine

Product attributes of Client

Product Image URL for Client vs. Competitor

Product attributes of all Competitor

Matching Score

Product Match/ Unmatch

Confidence Level

NPL
Natural Language Processing

Image Processing And Machine Learning

Artificial Intelligence Brings a New Paradigm in Document Digitization

Most businesses are required to properly store and archive legal agreements. These documents are scanned as pdfs making them difficult to search. This means that when a business leader needs to find and extract pertinent information regarding legal risk, they must navigate the pages of every document, scrolling through pages in hopes of isolating a specific clauses of interest.

One solution to the document storage challenge has been capturing information in a relational database within distinct fields. This requires extensive human effort to read through the documents then interpret and capture the data in an attempt to extract value from various records. Whenever a new piece of information is needed, the effort must begin from scratch to open documents and capture the relevant details.

A new form of document digitization is about to render the old approach obsolete. The solutions that make digital forms of documents more easily navigable are improving. While Optical Character Recognition (OCR) has been in the headlines for several years, Pattern Matching based automated capture is the next generation in document digitization. Together these technologies are blueprints for the way digitized providers can deliver business solutions that involve legal issues. OCR helps convert scanned images into machine readable text. Fields are captured using pattern matching based rules that reduces the overall effort required for robust document analysis. With OCR, businesses can make scanned documents fully text-searchable and indexed in their document repositories. Instant keyword searches save time and effort when searching for key details. OCR allows employees to efficiently search client information, correspondence, invoices and discovery documents. At present OCR cannot be used for a long tail of low volume documents because the setup effort required to automated capture is still too robust.

The combination of Machine Learning, Natural Language Processing and Text Analytics with Pattern Matching is furthering the effort to deliver digitised fit for purpose output in three different forms.
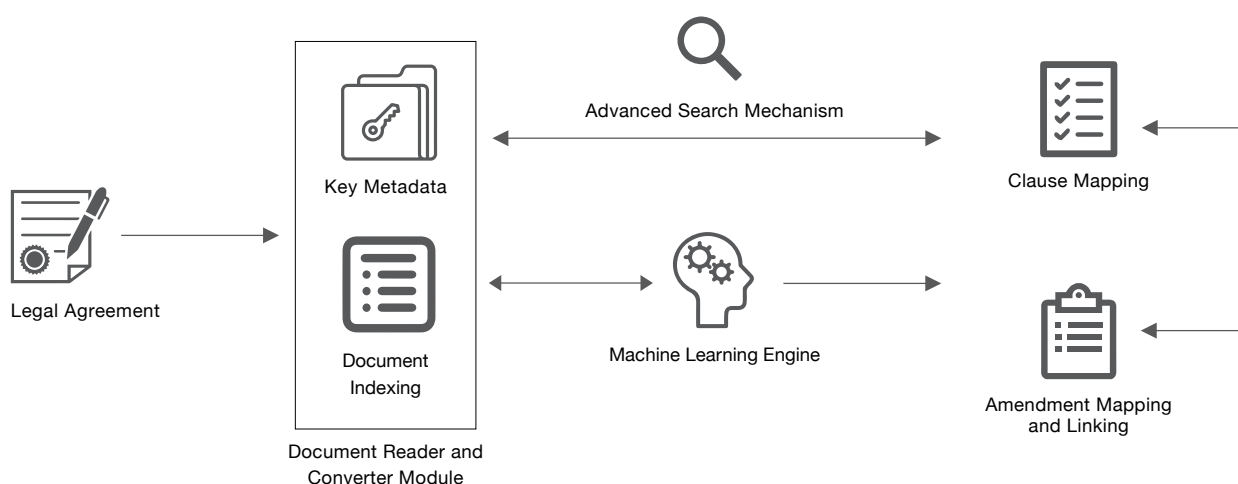
1. Fully parameterized, relational data model form: This form Is produced only where necessary, based on number of documents to be processed. It requires significant training effort for automated capture, but significantly less effort compared to the pattern matching, rule-based capture.

2. Clause database: Legal documents are stored with a mapping of standard clauses, allowing users to search through specific clauses. This output can be produced with very little effort compared to the fully parameterized form, and the only training needed is the ability to classify different sections into standard clauses.

3. Searchable documents: This document is stored in a searchable format with sections identified within the documents. No training effort is required for this output, as automation identifies the sections within the document based on formatting cues.

The document digitization program generates specific business results including:

- Combines automation with manual review to deliver almost 100% accuracy in complex legal documents
- Runs and streamlines BAU process and large remediation projects
- Reduces capture effort by 35-40% for complex documents, over 80% for forms and templates
- Faster turnaround response to digitization requests

Machine Learning and Natural Language Processing have accelerated the process of analysing hidden risks in legal document archives. The solution can be rolled out to progressively extract the information, starting with searchable clause repository becoming available within a few days, comparison of clauses to identify outliers becoming available within a few weeks and a fully parameterized data extract, where necessary, becoming available within a few months.

The many benefits of the new and improved solutions in document digitization can increase productivity and quickly generate ROI, particularly given the increasing demand for better legal services and customer experience. Making the switch to streamlined digital forms and digitized legal documents is a step towards a more competitive future.

Legal Agreement → Key Metadata / Document Indexing (Document Reader and Converter Module) ↔ Advanced Search Mechanism ↔ Clause Mapping; Machine Learning Engine → Amendment Mapping and Linking

Efficient Image Retrieval for Publishing and Marketing

Companies in the publishing industry have numerous images in their repository, which are untagged and unclassified. These images consist of events, personalities, products. The graphical nature of his content makes it difficult and time consuming to search images. Given the speed and turnaround time imposed on publishers, the inability to retrieve images quickly and accurately creates mayhem.

- In order to maintain the competitive edge publishers are seeking an economical solution to enrich the image repository with metadata.

- With the advancement of image analysis using machine learning techniques, along with the availability of online services that help in aggregating the abundance of information available on the web, a more effective solution can be implemented to address the challenges of image retrieval.

- eClerx has devised a human-in-the-loop automation that uses multiple principles to reduce the effort required to add curated annotations to images. The figure below represents a simple process flow.
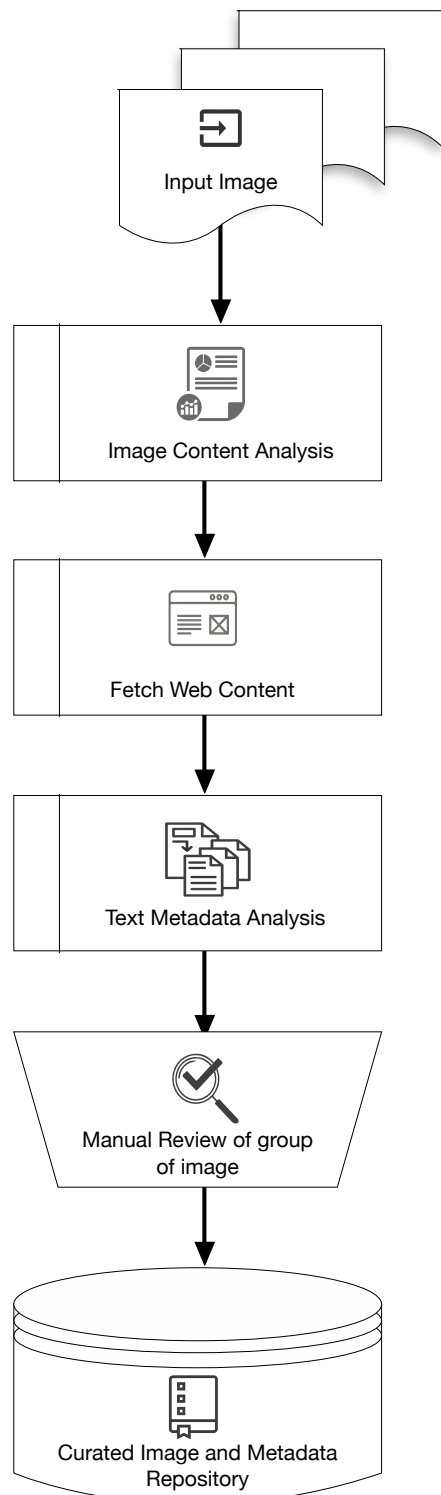
Traditional approaches of image content based search or text metadata based search are either too slow and inaccurate or cost inefficient.

Input Image

↓

Image Content Analysis

↓

Fetch Web Content

↓

Text Metadata Analysis

↓

Manual Review of group of image

↓

Curated Image and Metadata Repository

1. Image clustering and similarity scores to generate metadata: This is particularly useful in marketing where you have multiple images for numerous products. This step can help in grouping together images of the same product, or find similar types of images for all products (e.g. all motorbike images that face towards the left). Based on this criteria metadata labels can be added to newly-uploaded images.

2. Use multiple approaches to retrieve metadata on the images, based on captions and term frequency-based measures. In addition to providing web-based metadata, such services also provide additional metadata.

3. Based on the metadata that is automatically generated in the previous steps, clustering of images is performed to bring together images that have a high degree of overlap in the metadata.

4. To improve the quality and consistency of the metadata, the automatically-generated metadata is reviewed by a human. The review interface is optimized to enable reviewer to finalize or curate the metadata. Figure 2 on the page below provides a schematic of how the information is presented to the end-user for a quick review.

By utilizing this approach we reduced the manual effort required to add metadata to images by 95%.

At eClerx we recommend storing the metadata in logically-segregated columns/buckets. Moreover, in case of product images, a standard set of attributes needs to be defined with specified values.

We then created a repository comprised of images, image content-based attributes, and textual metadata. Next, we defined the record structure based on the context of the use case. For example, to publish the use case, the text metadata was stored in broad buckets – web, image content, personalities, and event; for marketing, the attributes were standardized with a single distinguishing value – product name, model number, image backdrop, picture angle, etc.

A combination of visual- and text-based techniques provides a very efficient and effective method of image retrieval. This method addresses both scenarios – based on query or keyword. Metadata service providers, like Google Cloud Vision API, can help overcome some of the shortcomings of the manual effort involved in creating high-quality text metadata. Through image content analysis, specific metadata attributes may be used as standard attributes within the repository. The manual effort can be further reduced using an intelligent workflow that presents similar images to the reviewer, on a single page. Through this human-in-the-loop automation, the task of curation of large digital asset libraries becomes a feasible one.

Marketing functions can use this approach to build image repositories of products, this can be used to automate web page rendering or expedite the searching during marketing collateral creation. Publishing companies can use this approach to store images with metadata that is not only based on the content of the image but also relates to the event and news associated with the image. This solution approach can be extended to other use cases such as locating closely resembling images from the repository to speed up creative or engineering design work, and can save hours of human effort in creating the models from scratch. Deep learning has potential to truly revolutionize the creatives production industry, and the solution that we have developed is a small step in this direction.

Figure 2: Curator's interface
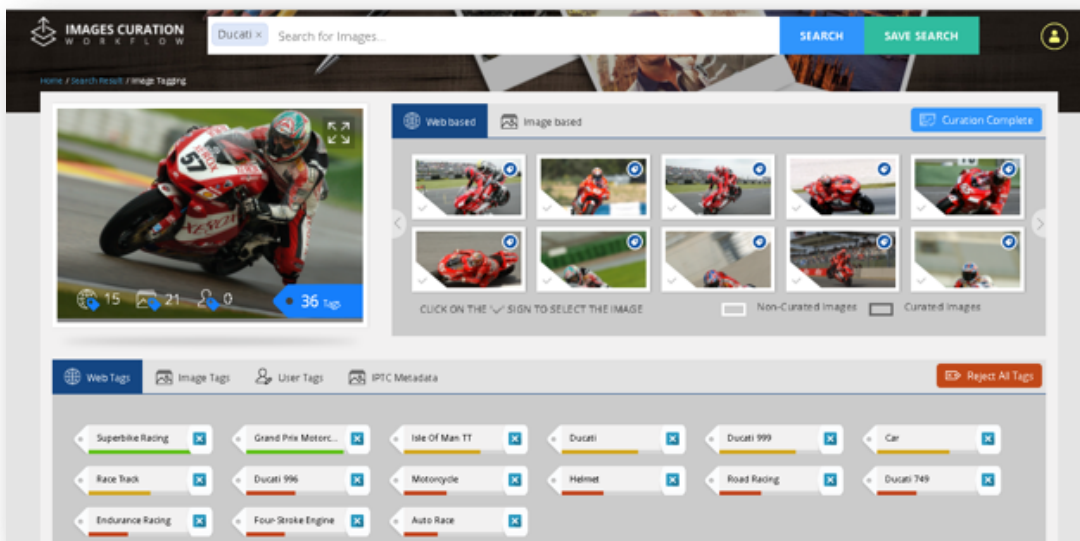
Customer Experience Hub

We provide analytics, insights and real-time recommendations to support your customer-facing teams and elicit greater value from your engagements.

By creating an easy to use the repository of digital media clips demonstrating excellent contact handling skills, your agents can make a significant impact on your business and help reduce customer complaints, repeat customer interactions, handling time and cost per handle. With an easy to use interface, intuitive dashboards, and end-to-end reporting capabilities, your teams can easily recognise what good should look like.

Customer Experience Hub helps you resolve issues associated with:

- Multiple Data Sources

- Defining & Rewarding Best Practices

- Inconsistent Channel Management

- Erroneous Data

- Poor Training Material for New Joiners

Customer Experience Hub

PRODUCT SPOTLIGHT

Customer Experience Analytics and Insights

- 7% reduction in transfers reduced costs and increased first time resolution

- Up to 30% increase in upsell value

- 40 % reduction in customer complaints

- Significant reduction in repeat calls

- 8% reduction in churn

- 20% increase in CSAT scores

eClerx