Smart Algorithms for Information Retrieval

About us

Yapı Kredi Teknoloji A.Ş. is a technology company established in 2015 to deliver innovative and patentable products and solutions with high added value.

We develop innovative and R&D-oriented comprehensive software projects and mobile applications applicable to the industry, especially banking and finance. For the solutions and products, we mostly employ artificial intelligence, machine learning, natural language processing, data mining and mobile software development techniques.

We cooperate with academic institutions, domestic and foreign R&D companies for the projects developed. We share the project outcomes in scientific articles and at national/international conferences with our industry partners and academia.

Customer focused smart products

Safir

Safir, is an artificial intelligence application that classifies transaction orders that come from various channels i.e.fax and e-mail, and removes manual data entry by automatically providing transaction information from these orders.

Thanks to Safir you can increase work efficiency and minimize human errors.

What problems does Safir solve?

Every day, tens of thousands of customer orders reach to banks from various channels. At first step, types of these transaction orders are determined by operation teams, afterwards the transaction information is entered manually into the system by looking at the instruction image. In the last stage, another employee checks and approves the transaction data entered.

The data entry process by looking at the transaction order images takes a lot of time for operation teams and carries the risk of error due to human nature.

Safir is an artificial intelligence and Turkish natural language processing product that automatically classifies customer orders written in free format according to the transaction type and automatically extracts transaction information. This way Safir removes the need of entering transaction details manually. Achieved as Safir digitizes transaction types classification and extraction of transaction detail phases.

50% — 50% reduction in manual data entry
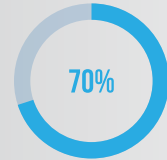
70% — up to 70% reduction in processing time

A customer order can consist of more than one page, and each page may have one or many different transaction types.

With Safir, manual data entry has come to an end!

Which transaction types does Safir support?

Safir supports 4 transaction types most intensively communicated to bank operation departments:

· Money transfers
· SGK payment (Social Security Institution)
· Tax payments
· Import payments

Safir can learn from past data and be trained according to your company's needs to produce the best results.

How does Safir work?

Transaction Orders
First transaction orders in free text format with TIFF, PDF, PNG, or JPEG extension are sent to Safir as image.

Structured as "Software as a Service" architecture, Safir sends the incoming request to the Optical Character Recognition (OCR) module.

Optical Character Recognition
Via OCR software, Safir extracts characters from the image.The OCR module has features such as "barcode reader" and "character location detection" and it is optimized according to various parameters, some as below

· Transaction Order Types
· Document Extensions
· File Size
· PC Performance

Transaction Order Classifier
Transaction Order Classifier detects the type transaction, both page by page and on the basis of order itself.The classifier determines transaction order types with 3 different methods:

Rule Based
Machine Learning
Client Behavior Analysis
Then it integrates the results from the 3 methods into single correct one.

Information Retrieval
Safir Information Retrieval module retrieves information required for the execution of transaction with models customized for each transaction order type.

Sender's account number
Receiver's account number
Amount
Currency, etc.

Rule Based
Pattern recognition
Machine Learning

Information Retrieval

Transaction Results
All data coming from the Information Retrieval module is used to create transaction orders. This phase is where the transactions leaving the Safir system are generated.
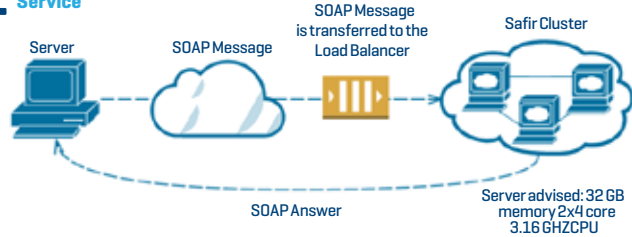
Validation and Correction
Noisy results are improved by using client's past data. Therefore, logic errors and OCR originated problems are overcome and maximum correction is achieved.
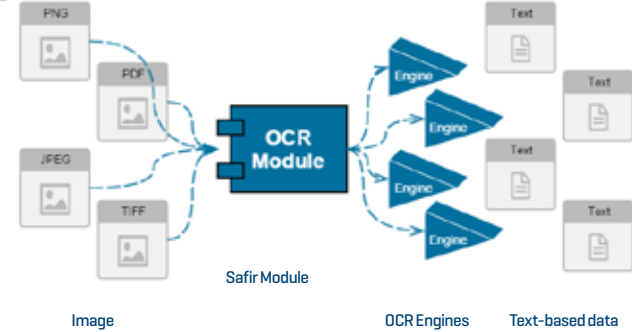
Efficiency:
Safir

Smart Algorithms for Information Retrieval

1 Service

Server — SOAP Message — SOAP Message is transferred to the Load Balancer — Safir Cluster

SOAP Answer

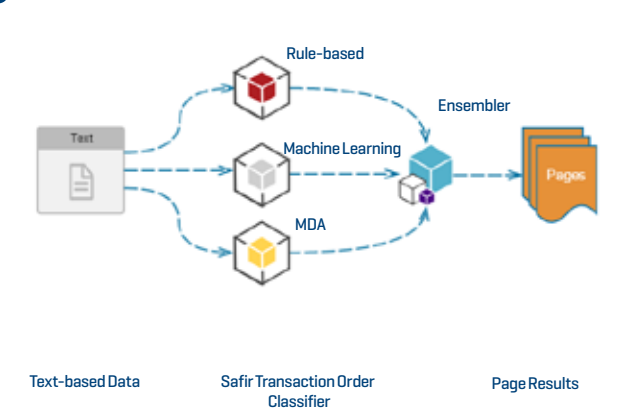Server advised: 32 GB memory 2x4 core 3.16 GHZ CPU

- As a "Software as a Service", Safir has the capacity of processing requests that come from the web without making the client wait thanks to its asynchronous structure.
- Thanks to its expandable structure, each Safir server added to its batch increases the system capacity and the performance.
- Requests reaching Safir in their order are forwarded to the best fitting machine in the network.
- Safir services, developed with Java and built upon the Tomcat web host, can perform independent of platforms in different environments

2 Optical Character Recognition

PNG, PDF, JPEG, TIFF — OCR Module — Engine / Engine / Engine / Engine — Text / Text / Text

Image — Safir Module — OCR Engines — Text-based data

- Safir Optical Character Recognition module was developed with ABBYY. It supports TIFF, PNG, JPEG, BMP and PDF formats.
- As the OCR engines can be used as 'out-of-process' during the course of intense amount of work and request, it can also be used as 'in-memory' when the priority is minimal use of resources. The 'out-of-process' mode is advised for modern systems.
- Optimum number of OCR engines = Number of CPU cores
- The file size, number of pages to OCR and various other parameters can easily be configured during execution.
- Has multiple additional features - some as:
  Precise character recognition
  Barcode reading
  Character location detection
  File Format conversion
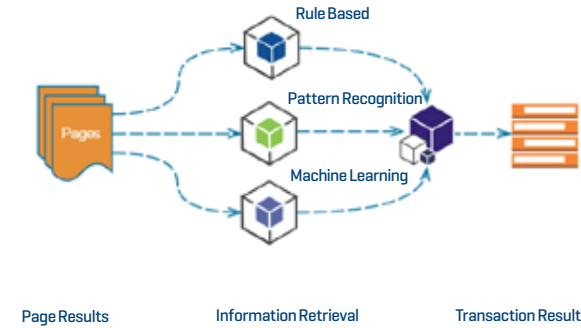- Safir OCR module communicates with the ABBYY software through Java Native Interfaces.

3 Transaction Order Classifier

Text — Rule-based / Machine Learning / MDA — Ensembler — Pages

Text-based Data — Safir Transaction Order Classifier — Page Results

- Safir Transaction Order Classifier aims to estimate the most accurate transaction order type by using 3 different methods.
- Rule Based: Classifies by identifying keywords that belong to certain processes in the transaction order text.
- Machine Learning: Makes automatic classification thanks to Machine Learning methods unaffected by noisy text using expressions with different lengths from the past transaction orders.
- Customer Behavior Analysis (MDA*): Classifies based on customer profiling by evaluating their past transactions without the need for analyzing current order's content.
- Integrator: Brings everything to a single result by using a decision tree on the estimation of the 3 classification methods.
- Safir allows adaptation to new data environments with the re-training of the classification methods.
- The page-based results from classifications are brought together and the transaction order of each page is executed.

*This module has to be re-trained for new clients. Without this part the system can work but the success is affected. More details on its success can also be provided.

---

4 Information Retrieval

Rule Based / Pattern Recognition / Machine Learning

Pages — Page Results — Information Retrieval — Transaction Result

- Safir Information Retrieval Module's main role is to extract, from the documents classified, information such as the sender's/receiver's account number, receiver's name (company/person), transaction date, balance and currency.
- Rule Based (REGEXP): Extracts entities such as the IBAN, account number, ID Number, tax registry, accrual numbers and etc.
- Pattern Recognition: Extracts patterns from the transaction order without being effected by noisy text and reveal entities within patterns.
- Named Entity Recognition (NER): Extracts entities by using "Entity Extraction" and "Entity Relation Extraction" algorithms of Natural Language Processing methods.
- A "Safir Transaction" object is created with all the information retrieved. Transaction information objects are kept as (key, value) pairs.

Example:
Sender Account Number: 101XX903
Receiver IBAN Number: TRXX994301122XX0900002913
Explanation: Rent
Balance: 1.200
Currency: TL

5 Validation and Correction

ORACLE / MySQL / MS SQL

Transaction Validation and Correction
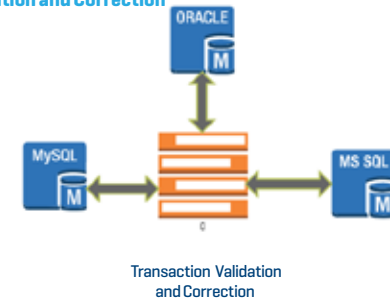
- Safir Validation and Correction module is developed to increase the validation success rate and to correct the noisy data produced due to low resolution, insufficient/incorrect OCR results, noisy text, logical errors and etc.
- Safir, which builds up from past data, updates and corrects Safir transaction objects by analyzing clients' past transaction data from different sources.
- In order to make database connections and access data JDBC and JPA technologies are used and DAO structures are built inside the module. This way it is easier to integrate different data sources.
- From here on, perfected results are brought to the server.