

Whitepaper: PROSAR-AIDA and EOB Processing

Executive Summary

Paradatec founded in 1988, focuses its engineering skills to delivering the most powerful and flexible Unstructured and FreeForm data capture solution in the ECM market.

Our flagship product, PROSAR-AIDA, seamlessly integrates with industry leading document capture platforms such as InputAccel from Captiva, and Ascent Capture from Kofax, to deliver a Best-In-Class document and data capture solution. Through a select group of systems integrators and OEM partners, PROSAR-AIDA has been deployed in a variety of applications across all major industry verticals including: Healthcare, Insurance, Banking, Automotive and Retail. With nearly a decade of production use, PROSAR-AIDA has been proven to be a mature robust solution that can scale from thousands to hundreds of thousands of images per day, and deliver a substantial ROI.

PROSAR-AIDA reduces the need to manually sort any mix of unstructured Free Form documents, and classify them without the use of templates or Zone OCR. Through sophisticated business rules it will easily classify your document type and extract and export key data fields. With our uniquely powerful table and line item reading functionality, we can interpret complicated layouts of tabular data found on documents such as Invoices and multi-patient EOB's. Customer benchmarks have demonstrated that our processing speed is significantly higher than any other Free Form technology, thus allowing same day processing of documents in high volume installations.

Headquartered in Braunschweig Germany (not far from Hannover), **Paradatec** has regional offices in England and Ohio (USA).

PROSAR-AIDA and EOB processing

By improving the speed and efficiency at which a company captures and processes patient Explanation of Benefits information, PROSAR-AIDA from **Paradatec** can produce measurable benefits at every level of the organization. The benefits resulting from PROSAR-AIDA encompass additional revenues as well as savings from reduced costs.

Hard-dollar benefits include:

Labour savings

- Reduced need for data entry operators; document prep workers, sorters, etc.,
- Increased accuracy of data extracted from EOB's reduces the time spent on error correction

Improved cash management

- Faster payment postings and access to more accurate payment data, means increased control over billing and accounting processes, which leads to better overall cash flow management
- Faster, more accurate generation of critical payment data expedites the process of Secondary Claim submissions.

Rapid return on investment

In mid to high volume organizations, the payback on PROSAR-AIDA is less than a year.

Soft-dollar benefits are measurable but become visible when PROSAR-AIDA changes the work process. Thus, they must be measured on a departmental level and beyond.

Soft-dollar benefits include:

Increased worker productivity

- More EOB's are processed in a shorter time.
- Other healthcare management processes dependent on content extracted from EOB documents, are shortened. Resulting in truncated business cycles from increased throughput

Improved customer service

Faster and more accurate EOB processing means that more customers get served more rapidly, with fewer payment errors.

Strategic benefits originate from EOB work process improvement; therefore they are measured on an enterprise-wide basis. These benefits are long-term rather than short term, and are measured yearly rather than quarterly. Their effects become rapidly apparent because they transform the nature of the business itself.

Healthcare Document Data Capture Overview

PROSAR-AIDA eliminates the need to manually sort any mix of documents before scanning. You can literally feed incoming mail directly into a batch and PROSAR-AIDA automatically detects the types of the documents it is fed. PROSAR-AIDA also is able to discover tables within unstructured documents "on-the-fly" with no predefined layout knowledge.

PROSAR-AIDA is designed to extract data from unstructured paper documents or forms, and pass on indexed data to other back end products via capture platforms from Kofax and Captiva.

There are many products on the marketplace today that have the capability to process structured forms but these products are typically difficult to administer with a wide variety of document types and are unable to handle unstructured data. Most forms processing products need “templates” to define the layout of a known form. Templates allow these products to only OCR known regions of a form for specific data (since OCR can be slow). For sites with a wide mix of documents or forms, templates become hard to maintain. Changes in form layout may require reprogramming.

PROSAR-AIDA does not use templates or make assumptions about the layout of potential documents or forms. It always performs a full-page OCR. It analyses the OCR output based on rules that define the general properties of various types of documents. It then determines the document type such as “EOB” or “dropped HCFA”.

Rules also indicate the general properties of data fields to be extracted (without needing to know locations of data fields). Rules are provided without programming via a simple user interface. Despite a full-page OCR for each page, PROSAR-AIDA typically processes a page per second or roughly 3,000 images per hour on a single PC. A unique feature of PROSAR-AIDA is the ability to extract large volumes of tabulated data, such as line items on an EOB form, to a high degree of accuracy and with a minimum amount of effort by the end-user.

One vertical market segment where PROSAR-AIDA shines is the health care market. Forms such as HCFA 1500's and UB92's are semi-structured in that they combine standard fixed fields with other fields that are variable in length (such as section 24 on a HCFA). PROSAR-AIDA is valuable in this market in that it can not just read these forms but also automatically: Sort mixed batches of HCFA's, dropped HCFA's, EOB attachments, UB's and dropped UB's (whether typed or handwritten). Thereby eliminating manual sorting.

Determine document boundaries in batches with documents comprising say a UB form, several attachments and then another UB form i.e. we automatically correlate the EOB attachments for a form with the form itself.

Determine that a form is a multi-page HCFA (say) by reading the insured's name/ID from each page and finding a match.

And perhaps most importantly: We can automatically extract data from unstructured EOB forms. We do this without the need to create a special template for any particular EOB variety. Our unstructured document processing is smart enough to be able to extract data from EOB variations that it has never seen!

How PROSAR-AIDA works

Rules

A user interface allows rules to be defined that determine: Which type of document is being processed (e.g. invoice, Healthcare form, insurance claim,) then once the document type is determined, rules for that document type determine how to identify and extract the required data.

This is all fairly standard across the industry. The flexibility of rules is more unique to PROSAR-AIDA. The hierarchical nature of rules in PROSAR-AIDA allows extremely fast processing to determine document type.

Rules for a) above, determine the general properties of a document type. For example, an invoice may have the following rules: The word Invoice may be found. The words "Account#" or synonyms like "Account Num.:" or "Acc No.:" or "Customer Num.:" may be located to the left of a series of numbers and letters (that represent the account number). The word "Amount Due" or synonyms such as "Pay This.:" or "Balance Due" or "Total Due.:" may be located near some dollar amount.

A table of dollar values may be located with headings such as "Item Number", "Quantity" or "Unit Price". Similar rules may apply to items such as "Invoice date", "Invoice Number" etc.

PROSAR-AIDA allows rules to be combined such that more complex rules can be built. We could combine rule 2) above with a rule that locates the words "Claim Number" and have a higher level rule that knows that "Claim Number" is typically a key word for a different document type hence reducing the likelihood that this document is an invoice.

The result of each of the rules above is a "fitness" level that indicates how likely it is that this document matches this rule (e.g. invoice rules in this case). Typically PROSAR-AIDA combines the fitness levels from all of its invoice rules and compares those to the fitness levels of each other high level document type. If the document is determined to be an invoice, other rules are then applied to determine which general classification of invoice it is (e.g. a healthcare provider invoice). Nowhere is there a template that matches "Invoice from vendor X". Hence general classes of document types support unlimited numbers of vendors & document types. We have customers where we detect hundreds of document types, all with a single set of rules.

Rules very similar to those above are employed to extract data from a known document type. For the document type "Healthcare provider Invoices" we may look for words "Amount Due" or a long list of synonyms and then attempt to identify a dollar amount near this word. Sometimes there may be several numbers near our key words hence rules usually say .We prefer the amount to be to the left of the key words, but if not found here, look directly below the key words.

Fuzzy Logic

Fuzzy Logic is crucial to the flexibility of PROSAR-AIDA. We just learned that every rule generates a “fitness” value that indicates how closely the item matches the rule. Fuzzy logic is also heavily used in recognition of the keywords that PROSAR-AIDA uses to learn about most documents. For example, to locate the “Claim Number” for a healthcare claim we may look for a series of synonyms for “Claim Number”. On many recognition systems, poor image quality can result in a wrong result. For example, suppose an OCR located the text “C1aim Nunnber” i.e. the letter .l. is OCR’d as the number “1” and the letter “m” is OCR’d as the letter “n” twice. PROSAR-AIDA uses “fuzzy matching” to determine if this text matches any of those sought. It will match this imperfect text with “Claim Number” but will give the result a slightly lower “fitness” than if it was sure that this is “Claim Number”.

Support for wildcards in our search strings makes things even easier. We can search for “Clai?” instead of “Claim” and the “?” matches any character.

This use of fuzzy logic allows PROSAR-AIDA to be flexible and accurate even with a poor image quality. Since the text matching is fuzzy, an imperfect OCR is easily tolerated.

Verification/validation with PROKEY

PROSAR-AIDA typically interfaces with PROKEY, its own verification/validation module. PROKEY is designed specifically for unstructured documents and allows:

- An operator to create/move document borders
- Multiple results
- Compound data (tables, addresses ...)

Ascent’s own correction/indexing tool may sometimes be used in place of PROKEY but, obviously, the features above may be missing in some cases.

Table Data Extraction

PROSAR-AIDA can automatically detect tabular data and export that data in a normalized and consistent format. This capability can be applied to Healthcare forms and is a significant differentiator of PROSAR-AIDA. Table detection is easy to set up & can even handle line items that are wrapped over more than one line.

Performance

PROSAR-AIDA is specially designed for high performance in environments with complex data requirements and high document volumes. Even with several hundred different document types and extensive data extraction (including tables), the processing times is typically around one second per document using standard PC hardware. This allows volumes between 2,000 and 3,000 images per hour to be processed on a single computer.

In addition, PROSAR-AIDA supports operation on SMP architectures and multi-server operation, so that using a cluster of recognition PCs will result in appropriately higher volumes. Installations with volumes of between 1,000 and 200,000+ images per day are in running operation and are available for reference visits. Some installations achieve throughputs of up to 25,000 images per hour.

Availability

PROSAR-AIDA is available for Windows and as an integrated module for the Ascent Capture and InputAccel document imaging platforms.

For contact information please visit our website at <http://www.paradatec.com>