



## *Advanced OCR in Education*

### **Table of Contents**

Executive Summary.....	2
Background.....	3
Structured Document/Form Processing.....	4
Unstructured Document Processing.....	5
Potential Areas of Implementation.....	6
Enrollments, Transfers & Transcripts Processing.....	6
Table Data Extraction.....	7
Automated Mailroom & Backfile Conversions.....	7
Summary.....	10

## Executive Summary

PROSAR-AIDA from Paradatec is an advanced document recognition technology that was initially designed for the insurance industry and has been successfully implemented across a variety of vertical markets including education. In a mailroom, PROSAR-AIDA eliminates the need to manually sort any mix of documents before scanning. One can literally feed incoming paper or faxed documents directly into a batch and PROSAR-AIDA automatically detects the types of the documents it is fed. PROSAR-AIDA also is able to discover data within documents where the layout of the document may be variable, per institution for example, or unpredictable, such as in the case of transcripts, and correspondence documents.

PROSAR-AIDA is designed to extract data from documents and pass on indexed data to other systems. There are many products in the marketplace today that have the capability to process structured forms but these products typically are difficult to administer with a wide variety of document types and are unable to handle variable document layouts. Most forms processing products need “templates” to define the layout of a known document e.g. the format of a student enrollment form. Templates allow these products to only OCR known regions of a document for specific data. For certain document types encountered within the academic world, variations can be extremely common. These types of documents do not lend themselves well to templates. A point occurs, where the number of unique templates required becomes unmanageable.

PROSAR-AIDA typically does not use templates or make assumptions about the layout of potential documents or forms. It always performs a full-page OCR. It analyzes the OCR output based on rules that define the general properties of various types of documents. It then determines the document type such as: “School Transcript” or “Enrollment Application”. Rules also indicate the general properties of data fields to be extracted (without needing to know locations of data fields). Despite incorporating a process which includes a full-page OCR for each page, PROSAR-AIDA typically processes a page per second or roughly 3600 images per hour per CPU core. This unparalleled performance has proven to be able to meet the needs of organizations processing hundreds of thousands of pages per day.

PROSAR-AIDA can process education-related documents from any source ***without special logic for the layout of documents from that specific source***, i.e. PROSAR-AIDA can:

- Automatically sort mixed batches of documents and attachments without separator sheets needing to be inserted between these documents.
- Extract key data such as: Student Name, Student Date of Birth, Student Number, Institution Name, and even Course Information...without special templates for individual document variations.
- Validate the information extracted against databases to make certain that the data is accurate.
- Accurately determine which of the data extracted needs to be verified by a human being.
- Perform all of the tasks above without the need for scripting or complex programming since PROSAR-AIDA is tuned purely through a graphical user interface. This means that there is no need for IT staff to maintain the system.

PROSAR-AIDA is typically deployed in large-scale, high-volume environments. Our reference sites process up to 250,000 scanned images per day.



## Background

Paradatec has been in the advanced capture technology business for more than ten years and has a proud history of innovation.

Paradatec's own OCR technology (PROSAR) is able to perform a full-page OCR of a page in much less than a second. This speed is several multiples faster than any competing applications on the market today. It is this unique capability which allows our product to process a document very much like a human would. We read the document from top to bottom, and then determine where the data is that we need. Competing applications have slower OCR and therefore need to perform a variety of shortcuts intended to allow them to "anticipate" the location of the data to be extracted. This approach may work acceptably for documents where layout is quite standard or where there are not too many variations. However, for the education market, where a great deal of variation can be common, this approach becomes too difficult to manage.

PROSAR-AIDA has been in production operation for the past twelve years and makes unstructured documents available to back end systems either via the Paradatec Capture Workflow Module, or alternatively, via other popular capture platforms. Paradatec, the creators of PROSAR-AIDA, have collaborated on projects ranging from 5,000 to 250,000 scanned images per day.

PROSAR-AIDA high volume implementations include (but are not limited to):

Company Name	Vertical Market and Application
AZ-based Institution	Education
Allianz	Property/Casualty, Life, Health
Eucon Information Systems	Automobile Claims
TX-based Mortgage servicer	Post-closing Audit & Mailroom automation
Equifax (HR BPO)	Mailroom Automation
Auto Finance (USA)	Financial Contracts (major automaker)
Hanse-Merkur Insurance	Health, Casualty, Life
ERGO Group	Health Insurance

One of the major strengths of PROSAR-AIDA is its scalability. Our customers are currently achieving processing rates exceeding 30,000 images per hour.

## Structured Document/Form Processing

Structured documents or forms are pages with a pre-defined layout. The positions of lines or shapes on the form can be used to identify which, of a known list of forms, this is.

**PARADATEC UNIVERSITY**  
Undergraduate Application for Admission

*Please print or type all information requested and review responses for accuracy.*

**Social Security Number (USA)** 123-45-6789

**Current Legal Name**  
 Last Name: Jones      First Name: John      Middle Initial: J  
 Jack Jones, Jim Jones, Jay Jones  
 Other name(s) by which you may have been identified in relevant academic records

**Permanent Address**  
 333 N. 3rd Ave.  
 Number and Street (include Rd., St., Ave., etc., as well as Apt. or Suite number if applicable)  
 Cincinnati      Ohio      33434  
 City      State      Zip Code      Country (if not USA)

**Phone Numbers with Area Code (please circle "D" for day or "E" for evening):**  
 (Home)      (D/E)      (Business)      (D/E)      (Business Extension)

**Email Address**

**For Office Use**  
 Special Population: P C E  
 Third Party: [ ] [ ] [ ]

Student Name - Jones, John J.  
 Student Address - 333 N. 3rd Ave.  
 City - Cincinnati  
 State - Ohio  
 Zipcode - 33434...

For structured document recognition, known shapes and text may be checked in a **specific area** of a document to determine if a document is of one template type. In this example, it is possible that the area shown might be analyzed to see if this is a standard (known) form. If so, only specific areas are OCR'd to obtain index data.

Form-reading technologies employ "reading zones" to limit the time needed to OCR/ICR a document.


PROSAR-AIDA can be used to process structured forms. Our approach is more reliable than the traditional zone-OCR approach. We read the entire page, look for the labels that identify the data to be extracted, and then extract the data from exactly the right place. In the example above, we would discover the location of the label "Social Security Number" from reading the entire document and then know exactly where to find the value of the Student's Social Security Number itself. This methodology allows for our solution works flawlessly in less than ideal situations. For example, many institutions allow their enrollees to print forms from the Internet, fill them out and fax them in. The forms can be shrunken, skewed or expanded by the printer or fax equipment. In these cases, the traditional approach of looking for data at fixed coordinates on a page will not work. The more flexible "on-the-fly" approach of reading the document from top to bottom guarantees a more reliable result.

## Unstructured Document Processing

The unstructured document processing capabilities of PROSAR-AIDA are useful where:

- Existing technologies do not scale well. PROSAR-AIDA is designed to scale to hundreds of thousands of pages per day.
- The layout of the data within a page changes from time to time (e.g. if the form/document creator updates it).
- Some documents may never have been seen by the recognition system e.g. a new layout of an enrollment or transfer application form is created.
- The variety of the incoming documents is so great that it would be impractical to create a template for each document type e.g. per-institution variation or per-state variations.
- Speed and scalability are a factor.
- The format of the data to be extracted varies e.g. “six-thousand dollars” == “\$6,000.00”.
- The amount of information to extract is unpredictable e.g. tabular data needs to be automatically indexed but the number of rows/columns varies.
- Multi-page documents exist where there is a relationship between fields on different pages.

In larger-scale environments it is not practical to pre-define “reading zones” for OCR/ICR. Instead, it is necessary to perform a full-page OCR and attempt to make sense of every clue provided by the document to determine both its type and index data.

 <b>Utah Valley State College</b>		800 West University Parkway Orem, UT 84058-5999 (801)863-8000, (801)225-4677 FAX www.uvsc.edu		Under The Family Education Rights and Privacy Act of 1974, information from this record may not be released to any other party without obtaining written consent of the student.																																																																																																																									
NAME: Clark Kent		ID: 111-22-3333	BIRTHDATE: 15-MAY-1982		DATE: 21-MAR-2006																																																																																																																								
Course Level: Undergraduate Current Program College : Sch of Humanities Arts Soc Sci Major : Theatre Arts			<table border="1"> <thead> <tr> <th>SUBJ NO.</th> <th>COURSE TITLE</th> <th>CRED</th> <th>GRD</th> <th>PTS</th> <th>R</th> </tr> </thead> <tbody> <tr> <td colspan="6">INSTITUTION INFORMATION CONTINUED:</td> </tr> <tr> <td>RUS 2020</td> <td>Intermed Russian II HH</td> <td>3.00</td> <td>A</td> <td>12.00</td> <td></td> </tr> <tr> <td>RUS 3050</td> <td>Advanced Russian LH</td> <td>3.00</td> <td>A</td> <td>12.00</td> <td></td> </tr> <tr> <td colspan="6">Ehrs: 21.00 GPA-Hrs: 21.00 QPts: 83.10 GPA: 3.95</td> </tr> <tr> <td colspan="6">Dean's List</td> </tr> <tr> <td colspan="6">2002 SPRING</td> </tr> <tr> <td>CNS 1250</td> <td>Obj-Orien Prg I-Java</td> <td>4.00</td> <td>E</td> <td>0.00</td> <td>B</td> </tr> <tr> <td>MATH 1220</td> <td>Calculus II NM</td> <td>5.00</td> <td>AU</td> <td>0.00</td> <td></td> </tr> <tr> <td>RUS 1020</td> <td>Beginning Russian II LH</td> <td>5.00</td> <td>A</td> <td>20.00</td> <td></td> </tr> <tr> <td colspan="6">Ehrs: 5.00 GPA-Hrs: 5.00 QPts: 20.00 GPA: 4.00</td> </tr> <tr> <td colspan="6">2002 FALL</td> </tr> <tr> <td>PHYS 3100</td> <td>Intro Modern Physics</td> <td>3.00</td> <td>C</td> <td>6.00</td> <td></td> </tr> <tr> <td>PHYS 3210</td> <td>Intro Exptl Phys I</td> <td>2.00</td> <td>A-</td> <td>7.40</td> <td></td> </tr> <tr> <td colspan="6">Ehrs: 5.00 GPA-Hrs: 5.00 QPts: 13.40 GPA: 2.68</td> </tr> <tr> <td colspan="6">2003 SPRING</td> </tr> <tr> <td>BIOL 1010</td> <td>General Biology BB</td> <td>3.00</td> <td>B</td> <td>9.00</td> <td></td> </tr> <tr> <td>HIST 1700</td> <td>Am Civilization AS</td> <td>3.00</td> <td>B+</td> <td>10.20</td> <td></td> </tr> <tr> <td>PHYS 3220</td> <td>Intro Exptl Phys II</td> <td>2.00</td> <td>A-</td> <td>7.40</td> <td></td> </tr> <tr> <td colspan="6">Ehrs: 8.00 GPA-Hrs: 8.00 QPts: 26.60 GPA: 3.32</td> </tr> </tbody> </table>			SUBJ NO.	COURSE TITLE	CRED	GRD	PTS	R	INSTITUTION INFORMATION CONTINUED:						RUS 2020	Intermed Russian II HH	3.00	A	12.00		RUS 3050	Advanced Russian LH	3.00	A	12.00		Ehrs: 21.00 GPA-Hrs: 21.00 QPts: 83.10 GPA: 3.95						Dean's List						2002 SPRING						CNS 1250	Obj-Orien Prg I-Java	4.00	E	0.00	B	MATH 1220	Calculus II NM	5.00	AU	0.00		RUS 1020	Beginning Russian II LH	5.00	A	20.00		Ehrs: 5.00 GPA-Hrs: 5.00 QPts: 20.00 GPA: 4.00						2002 FALL						PHYS 3100	Intro Modern Physics	3.00	C	6.00		PHYS 3210	Intro Exptl Phys I	2.00	A-	7.40		Ehrs: 5.00 GPA-Hrs: 5.00 QPts: 13.40 GPA: 2.68						2003 SPRING						BIOL 1010	General Biology BB	3.00	B	9.00		HIST 1700	Am Civilization AS	3.00	B+	10.20		PHYS 3220	Intro Exptl Phys II	2.00	A-	7.40		Ehrs: 8.00 GPA-Hrs: 8.00 QPts: 26.60 GPA: 3.32					
SUBJ NO.	COURSE TITLE	CRED	GRD	PTS	R																																																																																																																								
INSTITUTION INFORMATION CONTINUED:																																																																																																																													
RUS 2020	Intermed Russian II HH	3.00	A	12.00																																																																																																																									
RUS 3050	Advanced Russian LH	3.00	A	12.00																																																																																																																									
Ehrs: 21.00 GPA-Hrs: 21.00 QPts: 83.10 GPA: 3.95																																																																																																																													
Dean's List																																																																																																																													
2002 SPRING																																																																																																																													
CNS 1250	Obj-Orien Prg I-Java	4.00	E	0.00	B																																																																																																																								
MATH 1220	Calculus II NM	5.00	AU	0.00																																																																																																																									
RUS 1020	Beginning Russian II LH	5.00	A	20.00																																																																																																																									
Ehrs: 5.00 GPA-Hrs: 5.00 QPts: 20.00 GPA: 4.00																																																																																																																													
2002 FALL																																																																																																																													
PHYS 3100	Intro Modern Physics	3.00	C	6.00																																																																																																																									
PHYS 3210	Intro Exptl Phys I	2.00	A-	7.40																																																																																																																									
Ehrs: 5.00 GPA-Hrs: 5.00 QPts: 13.40 GPA: 2.68																																																																																																																													
2003 SPRING																																																																																																																													
BIOL 1010	General Biology BB	3.00	B	9.00																																																																																																																									
HIST 1700	Am Civilization AS	3.00	B+	10.20																																																																																																																									
PHYS 3220	Intro Exptl Phys II	2.00	A-	7.40																																																																																																																									
Ehrs: 8.00 GPA-Hrs: 8.00 QPts: 26.60 GPA: 3.32																																																																																																																													
TRANSFER CREDIT ACCEPTED BY THE INSTITUTION: 05/02-12/02 Brigham Young University Ehrs: 10.00 GPA-Hrs: 0.00 QPts: 0.00 GPA: 0.00			08/01-12/01 University Of Utah Ehrs: 0.00 GPA-Hrs: 0.00 QPts: 0.00 GPA: 0.00																																																																																																																										
05/00-05/00 Adv P1 - Math: Calculus AB (3) Ehrs: 8.00 GPA-Hrs: 0.00 QPts: 0.00 GPA: 0.00			05/05-08/05 Univ Of California-Los Angeles Ehrs: 8.00 GPA-Hrs: 0.00 QPts: 0.00 GPA: 0.00																																																																																																																										
INSTITUTION CREDIT:																																																																																																																													

A full-page OCR allows PROSAR-AIDA to read the entire document, then discover in context where the data is. For example, the two-column table boundaries for this transcript document are defined, by **discovering** these identifier headings. In this way we are able to extract the necessary transcript course and credit information in much the same way a human being would. The software maintains a list of synonyms for each label or string so that every synonym for every possible layout of the document is supported. This is one of many unique techniques our technology is capable of using for documents of this kind.

A key strength of PROSAR-AIDA is its ability to recognize and index structured as well as unstructured documents. Most education environments involve a mix of structured and unstructured documents.

## Potential Areas of Implementation

By streamlining processes and reducing complexities that add time, expense and frustration to enrollment, transfers and other administrative processes, Paradatec's technology has the potential to dramatically improve the way educational institutions process information. Paradatec's unique technology can be applied to a number of different areas including:


- Enrollments, Transfers and Transcripts Processing
- Automated Mailroom Processing
- Backfile Conversions

The discussion below is intended to highlight the advantages of using a sophisticated document recognition technology for these common education institutional needs.

## Enrollments, Transfers & Transcripts Processing

Our capture process for Enrollment and Transfers Processing can change business as usual in an area which is often bogged down by much paperwork and many manual processes. It can provide the automation necessary to make processes run more smoothly. PROSAR-AIDA is able to achieve a remarkable level of automation in classifying and capturing data from the type of paperwork found in this process due to the advanced forms reading techniques discussed in the previous section.

In addition to forms such as: Enrollment Applications, Health Insurance Validations, College Major Declaration Forms, and etc... PROSAR-AIDA is designed to process unstructured and semi-structured documents such as School Transcripts, correspondences, and other unstructured documents. These documents are created in many variations, and come from a great many different sources.



Utah Valley State College  
800 W. Central  
(801) 873-3333  
www.uvcc.edu

**Data Value Landmarks**

Under the Family Education Rights and Privacy Act of 1974, information from this record will be released to any other party without obtaining written consent of the student.

**Extracted Data Values**

NAME: Clark Kent      ID: 111-22-3333      BIRTHDATE: 15-MAY-1982      DATE: 21-MAR-2006

SUBJ. NO.	COURSE TITLE	CRED	GRD	PTS	R
<p>Course Level: Undergraduate</p> <p>Current Program: College : Sch of Humanities Arts Soc Sci Major : Theatre Arts</p>					
<p>TRANSFER CREDIT ACCEPTED BY THE INSTITUTION:</p> <p>05/02-12/02 Brigham Young University Hrs: 10.00 GPA-Hrs: 0.00 QPts: 0.00 GPA: 0.00</p> <p>08/01-12/91 University of Utah Hrs: 0.00 GPA-Hrs: 0.00 QPts: 0.00 GPA: 0.00</p> <p>05/00-05/00 Adv Pl - Math: Calculus AB (3) Hrs: 8.00 GPA-Hrs: 0.00 QPts: 0.00 GPA: 0.00</p> <p>05/05-08/05 Univ of California-Los Angeles Hrs: 8.00 GPA-Hrs: 0.00 QPts: 0.00 GPA: 0.00</p>					
<p>INSTITUTION CREDIT:</p> <p>1999 SUMMER ASTR 1010 Astronomy PP 3.00 D+ 4.20 MUS 1010 Intro To Music FF 3.00 C 5.10 Hrs: 6.00 GPA-Hrs: 6.00 QPts: 9.30 GPA: 1.55</p> <p>2000 FALL CSIS 2210 Unix Operating System 3.00 B 9.00 HUM 1010 Intro To Humanities HH 3.00 A 12.00 PHYS 2220 Phys Sci/Engr I PP 3.00 B 9.00 PHYS 2211 Phys Sci/Engr I Lab 1.00 A 4.00 Hrs: 10.00 GPA-Hrs: 10.00 QPts: 34.00 GPA: 3.40</p> <p>2001 SPRING ENGL 1010 Intro To Writing CC 3.00 A 11.10 ENGL 2020 Inter Wrtg: Sci/Tech CC 3.00 A 12.00 PHYS 2220 Phys Sci/Engr II PP 3.00 A 12.00 PHYS 2221 Phys Sci/Engr II Lab 3.00 A 4.00 Hrs: 12.00 GPA-Hrs: 12.00 QPts: 36.00 GPA: 3.00</p>					
<p>2002 SPRING CNS 1250 Obj-Orien Prg I-Java 4.00 E 0.00 E MATH 1220 Calculus II MH 5.00 AU 0.00 RUS 1020 Beginning Russian II LH 20.00 Hrs: 8.00 GPA-Hrs: 5.00 QPts: 20.00</p> <p>2002 FALL PHYS 3100 Intro Modern Physics 2.00 A 7.40 PHYS 3210 Intro Exptl Phys I 13.40 GPA: 2.68 Hrs: 5.00 GPA-Hrs: 5.00 QPts: 13.40</p> <p>2003 SPRING BIOL 1010 General Biology BB 3.00 B 9.00 HEAT 1709 Am Civilization AS 3.00 B 10.20 PHYS 3220 Intro Exptl Phys II 2.00 A 7.40 Hrs: 8.00 GPA-Hrs: 8.00 QPts: 26.60 GPA: 3.32</p> <p>2003 SUMMER HGH 1100 Pers Health/Wellness TE 2.00 W 0.00 PES 1300 Fitness For Life FE 1.00 A 3.70 PHIL 2050 Ethics And Values HH 3.00 B 10.20 PHYS 1010 Elementary Physics PP 3.00 A 12.00 Hrs: 7.00 GPA-Hrs: 7.00 QPts: 25.90 GPA: 3.70</p> <p>2003 FALL SENG 5710 Circuit Theory 4.00 C 6.80 HIST 2030 History Of Civ I SS 3.00 C 7.20 HFW 1100 Pers Health/Wellness TE 2.00 B 6.00 JPNS 1010 Beginning Japanese I LH 5.00 C 10.00 Hrs: 14.00 GPA-Hrs: 14.00 QPts: 30.00 GPA: 2.14</p>					

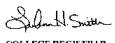
\*\*\*\*\* CONTINUED ON NEXT COLUMN \*\*\*\*\*

PAGE: 1

A BLACK AND WHITE TRANSCRIPT IS NOT AN ORIGINAL

OFFICIAL SEAL      RAISED SEAL NOT REQUIRED

This official college transcript is printed on sturdy paper and processed by Malcom Copy security to seal, does not require a raised college seal.



COLLEGE REGISTRAR

Undergraduate Admissions  
212 Hamilton Hall  
1150 Amsterdam Ave.  
New York, NY 10027

251568

The ability to reliably read these types of documents is something that sets PROSAR-AIDA apart in the industry. PROSAR-AIDA will discover column layouts “on-the-fly” without a template for each layout (just as a human decision maker would). This unique process provides the flexibility needed to support potentially tens of thousands of layouts.

## Table Data Extraction

The Paradatec table module can automatically detect tabular data and export that data in a normalized and consistent format. Table detection is set up via a GUI and has a number of important features such as:

- Multiple tables per page.
- Tables spanning pages.
- Math operations are used to help determine which column is which (e.g. for GPA values, the software is able to calculate a GPA value on semester sub-total line. This information helps to determine on-the-fly (without a template) which columns and lines are which.
- De-interlaced columns (e.g. on a transcript document, GPA value might be physically placed in the same column as, say, the Credits for a course. The table module will separate these values into two columns.
- Unstructured content columns (e.g. description columns) can be accurately identified.
- Can split a column into two (e.g. we can separate a course number from somewhere in the middle of an item description and return it as a separate column).
- Use of OMR (checkboxes or bubbles) objects in tables:
- Line items that are wrapped over more than one line (we will unwrap and deliver a single line).

It is important to understand that all of these features are inherent to our table module and require no programming or scripting. Everything regarding setup is via a GUI and is NOT done on a layout-by-layout basis, but rather a single set of table rules applies to every layout of that document type. This is not how most other technologies work. This allows the Paradatec solution to require very little in the way of maintenance due to layout changes over time.

## Automated Mailroom & Backfile Conversions

A great deal of information (everything from W-2s and tax returns for financial aid, to high school and college transcripts, correspondences, and applications) comes into an educational institution’s mailroom. When archived, this documentation can quickly become difficult to manage in its paper form.

Not only does this information take up a great deal of space over time, but it is subject to loss by disaster or simply misfiling. When a particular document needs to be retrieved, it often may not be found or requires a manual request which can cause unreasonable delays. Moving these types of documents to a content management system, can greatly streamline administrative processes and sharply reduce the risk of record loss.

PROSAR-AIDA can have significant value as part of an institutional backfile conversion solution. The technology is able to automatically:

- Sort mixed batches of hundreds of different document types as one would find in a typical student applicant or undergraduate file, thereby eliminating manual sorting.

- Determine document boundaries in batches with documents comprising say: an undergraduate application for enrollment, a corresponding application for financial aid along with associated attachments such as W-2s and bank statements, and then another several documents such as an applicant essay and letters of reference. In other words we can automatically correlate each applicant (folder) with its associated attachments (documents) with automation rates often exceeding 95%.

One of the challenges of the incoming documents in the education marketplace is that each educational institution will require different forms and different documents be present for each student applicant. School transcripts, which are often used in the process of evaluating potential enrollees, can come in an almost limitless variety of formats which often change over time.

An approach that tries to apply a fixed template to extract data at a certain location becomes unmanageable with, say, fifteen hundred document layout variations. As you should see in the illustration below there are 3 different documents from 3 different institutions, each with its own unique (and often changing) format. The challenge is to capture the information which may be in very different places from one variation of the form to the next.

**TRANSCRIPT**


Student No: 820-16-116      Date of Birth: 12-JUN-1985      Date Issued: 28-FEB-2005  
 Record of: Betty Rubble      10 Limestone Way      Bedrock, AZ 98765      Page: 1

Issued To: NYU  
 Office of Undergraduate Adm.  
 212 Hamilton Hall  
 1130 Amsterdam Ave. NC 2807  
 NY, NY 10027-2807

Course Level: Undergraduate  
 High School: Karachi American School 01-MAY-2003  
 Only Admit: Fall 2003  
 Current Major(s): Economics  
 Events: First Year Distinction

SUBJ NO.	COURSE TITLE	CRED	GRD	PTS	R
IN PROGRESS WORK continued:					
In Progress Credits      4.00					
***** TRANSCRIPT TOTALS *****					
TOTAL INSTITUTION	12.00	12.00	12.00	43.35	3.61
TOTAL TRANSFER	0.00	0.00	0.00	0.00	0.00
OVERALL	12.00	12.00	12.00	43.35	3.61
***** END OF TRANSCRIPT *****					

INSTITUTION CREDIT:  
 Fall 2003

  
**WELLESLEY COLLEGE**  
 WELLESLEY, MASSACHUSETTS  
 02481-8203

**Utah Valley State College** | 800 West University Parkway  
 Orem, UT 84058-5999  
 (801)863-8000, (801)225-4677 FAX  
 www.uvsc.edu

Under The Family Education Rights and Privacy Act of 1974, information from this record may not be released to any other party without obtaining written consent of the student.

NAME: Clark Kent      ID: 111-22-3333      BIRTHDATE: 15-MAY-1982      DATE: 21-MAR-2006

Course Level: Undergraduate  
 Current Program: College : Sch of Humanities Arts Soc Sci  
 Major : Theatre Arts

SUBJ NO.	COURSE TITLE	CRED	GRD	PTS	R
INSTITUTION INFORMATION CONTINUED:					
RUS 2020	Intermed Russian II HH	3.00	A	12.00	
RUS 3050	Advanced Russian LH	3.00	A	12.00	
Ehrs: 21.00 GPA-Hrs: 21.00 QPts: 83.10 GPA: 3.95					
Dean's List					
2002 SPRING					
CNS 1250	Obj-Orien Prg I-Java	4.00	E	0.00	B
MATH 1220	Calculus II MM	5.00	AU	0.00	
RUS 1020	Beginning Russian II LH	5.00	A	20.00	
Ehrs: 5.00 GPA-Hrs: 5.00 QPts: 20.00 GPA: 4.00					
2002 FALL					
PHYS 3100	Intro Modern Physics	3.00	C	6.00	
PHYS 3210	Intro Exptl Phys I	2.00	A-	7.40	
Ehrs: 5.00 GPA-Hrs: 5.00 QPts: 13.40 GPA: 2.68					
2003 SPRING					
BIOL 1910	General Biology BB	3.00	B	9.00	
HIST 1700	Am Civilization AS	3.00	B+	10.20	
PHYS 3220	Intro Exptl Phys II	2.00	A-	7.40	
Ehrs: 8.00 GPA-Hrs: 8.00 QPts: 26.60 GPA: 3.32					
1999 SUMMER					
WTH 1100	Perh Health/Wellness TE	2.00	W	0.00	

INSTITUTION CREDIT:  
 1999 SUMMER      ACTS 1010      Astronomy BB      3.00 B+      4.20

OFFICE OF THE REGISTRAR / SAIS			UNIVERSITY OF CALIFORNIA, IRVINE		
123-456-789	11111	MARVIN GAYE	*SOCECOL	*UNAFFILIATED	
SOCIAL SECURITY NO.	STUDENT NO.	NAME OF STUDENT	COLLEGE	MAJOR PROGRAM	
RESIDENT	SEP 2004	WALNUT CREEK, CA	01/08/XX	UNDERGRADUATE	
RESIDENT STATUS	DATE ADMITTED	PLACE OF BIRTH	DATE OF BIRTH	ACADEMIC RECORD	
		CFC	06/30/05	05/24/06	
SECONDARY SCHOOL - DATE GRADUATED - MISSION SAN JOSE HIGH SCH JUN 2004					
UNIVERSITY REQUIREMENTS:	09/04 UC ENTRY LVL WR REGT REGT SATISFIED				
	09/04 AMERICAN HISTORY REGT SATISFIED				
	09/04 AMERICAN INSTITUTION REGT SATISFIED				
	MEMORANDA				
	05/06/05 MAJOR CHANGED FROM UNDECIDED/UNDEC				
AP ENGLISH LIT	05-04				
OHLONE COLLEGE	2 TRM TO 08-03				
TOTAL	17.0*				
FALL QUARTER 2004					
HUMANITIES CORE LEC HUMAN	1A	4.0 B+	13.2		
HUMANITIES CORE WRT HUMAN	1A	4.0 B-	10.8		

PROSAR-AIDA does not have this limitation since it reads the entire document in context and determines the correct location for the Date of Birth from the label "Birthdate" or any other synonyms on the paper itself.

## Summary

Increased automation can have positive effects throughout an educational institution's administrative processes, from reduced operating costs to quicker and better enrollment and financial aid decisions.

Expenses and time saved from increased administrative efficiency leaves more resources available for instructional expenditures and a better focus on an institution's core mission.

PROSAR-AIDA provides the capability to process documents that are hard to handle with traditional zone-based systems. Our software eliminates the need to manually sort any mix of documents before scanning. One is able to literally feed incoming mail directly into a batch. From here, PROSAR-AIDA automatically detects the types of the documents it is fed.

Our product is configurable by the end-user. This allows an organization to leverage the knowledge and experience of their own people, who in every case, will have the best knowledge of what makes one document different from another.

PROSAR-AIDA is a proven technology used in large volume mailroom and data capture processing environments worldwide, since 1997. It can, and has been used in environments processing from several thousand to hundreds of thousands of scanned images per day.

