Optical Character Recognition

Optical Character Recognition (OCR) is the mechanical or electronic conversion of photo's or scanned images of either printed text or typewritten into encoded machine text. OCR is a field of research in pattern recognition, artificial intelligence and computer vision. The purpose of this document is to define the scope of the project and the solution provided to the client.

Technosoft has worked on following client projects using OCR.

Nutri Online App

Description

The purpose of this document is to define the scope of the project and the solution provided to the client.

1.1 Requirements

The client was looking to develop an internet-based application to order and provide meals for school children in private schools. This application was intended to gather information from various parties and applications, consolidate the meals, and then provide the data to all concerned parties in an efficient manner.

An administrator would create daily lunch menus, and then the student's parents would make meal selections on a week-by-week or month-by-month basis for each of their children. The meal schedules were then provided to the teachers, staff, cafeteria, and/or caterer either with the online application or in Excel files. The cafeteria or caterer then would prepare and deliver the lunches as necessary.

Solution Provided

The basic flow of the application was as following

1. Nutri-Link Sets up District.
2. District sets up.
    a. Schools
    b. Teachers
    c. Pantry
    d. Lunch menus
3. School provides the parents with an ID number for of each of their children.
4. School districts maintain daily menus for their schools.
5. The parents log in and set themselves up in the Accounts table.
6. The parents set each of their children up in the Students table.
    a. Parents update their children.
7. Parents select lunches for each of their children for each day of school.
    a. Parents make changes as necessary to the lunch schedule for their children.
8. School administrators confirm and place the meal order with the appropriate caterer.

www.techno-soft.com | info@techno-soft.com | (203) 676-8299

9. School administrators process payments and log against Accounts.

Important Technical Achievements

Following are some important technical achievements in this project:

1. Implementation of project source code in three layers consisting of User Interface, Business logic and Data access layers.
2. We have used WCF (Windows Communication Foundation) services to send email reminders.
3. We have used CRAP extensively to make website look better and noted this technique helped us much.
4. JQuery has been implemented in multiple modules.
5. We have used RDLC reports.

2 Application

This application is split into two applications. One application serves the school administrators in the role of coordinating lunches with caterers. The other application serves the parents in scheduling lunches for their children.

2.1 Application Overview

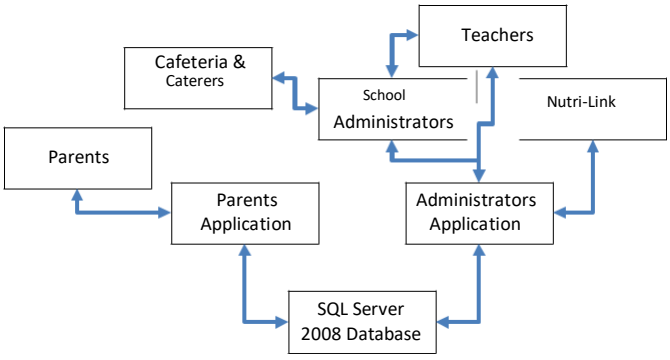
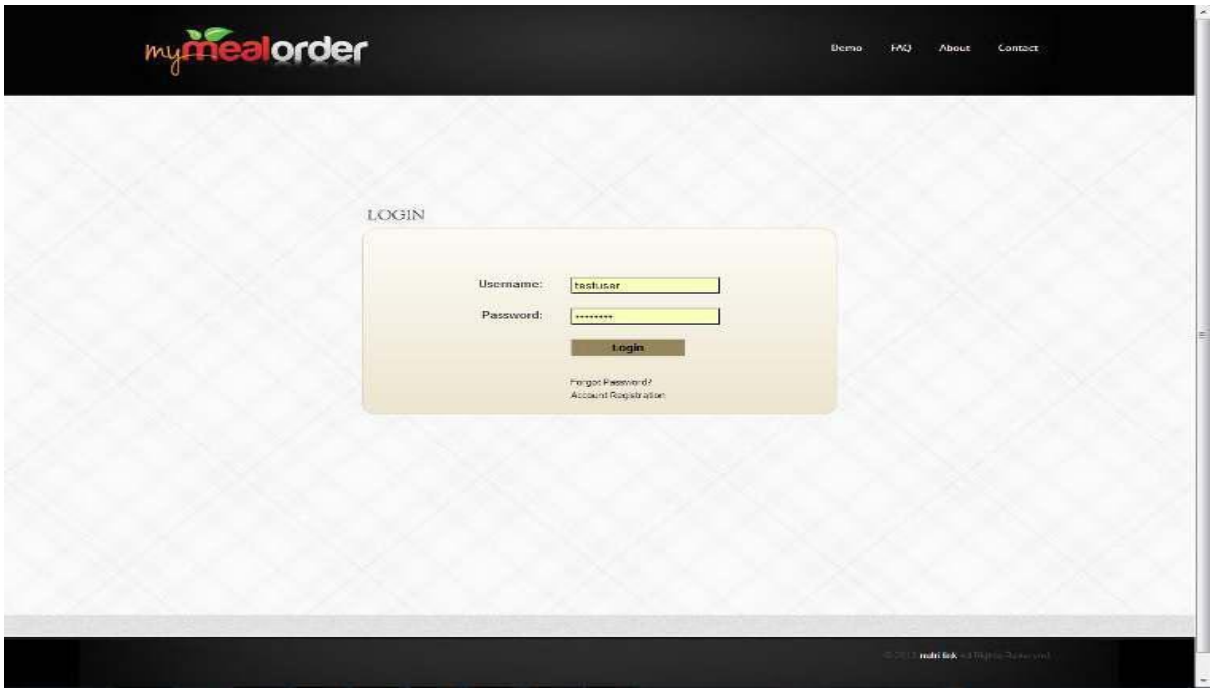


Figure 1: High-level overview of the two separate applications and the user groups they serve.

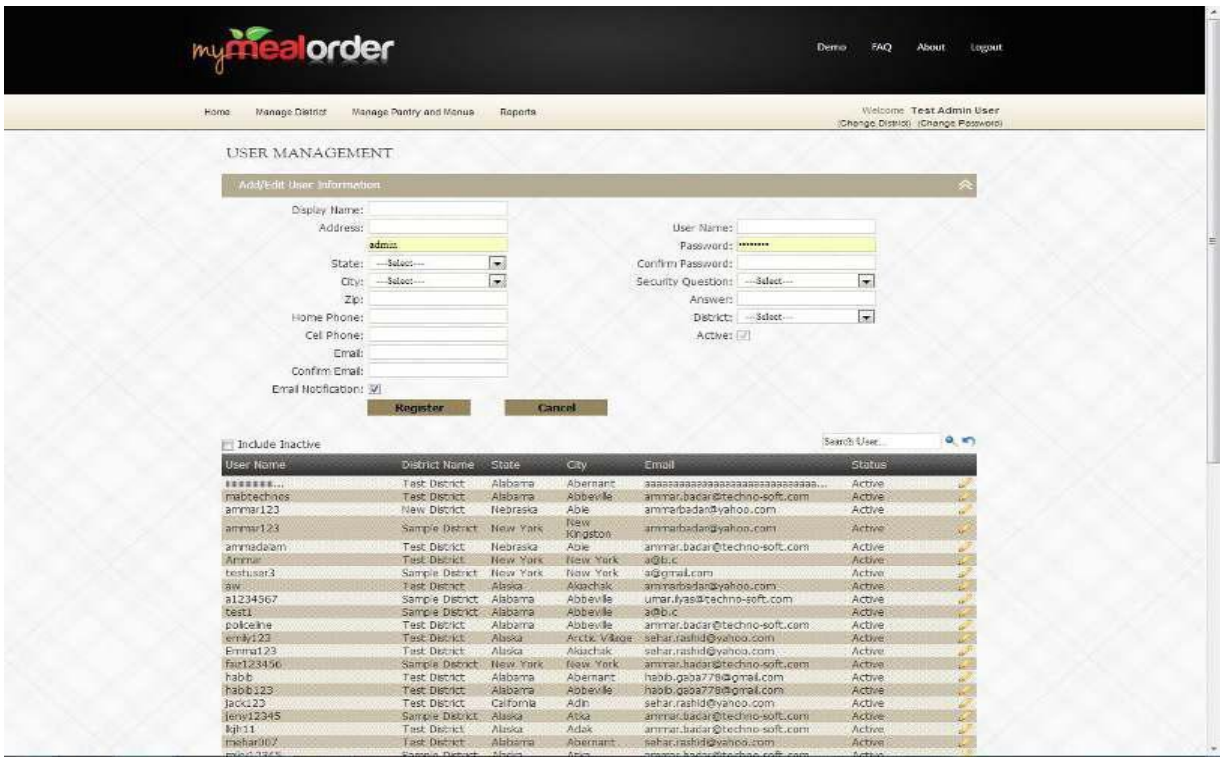
www.techno-soft.com | info@techno-soft.com | (203) 676-8299

2.2 Key Screens

Application Login Screen

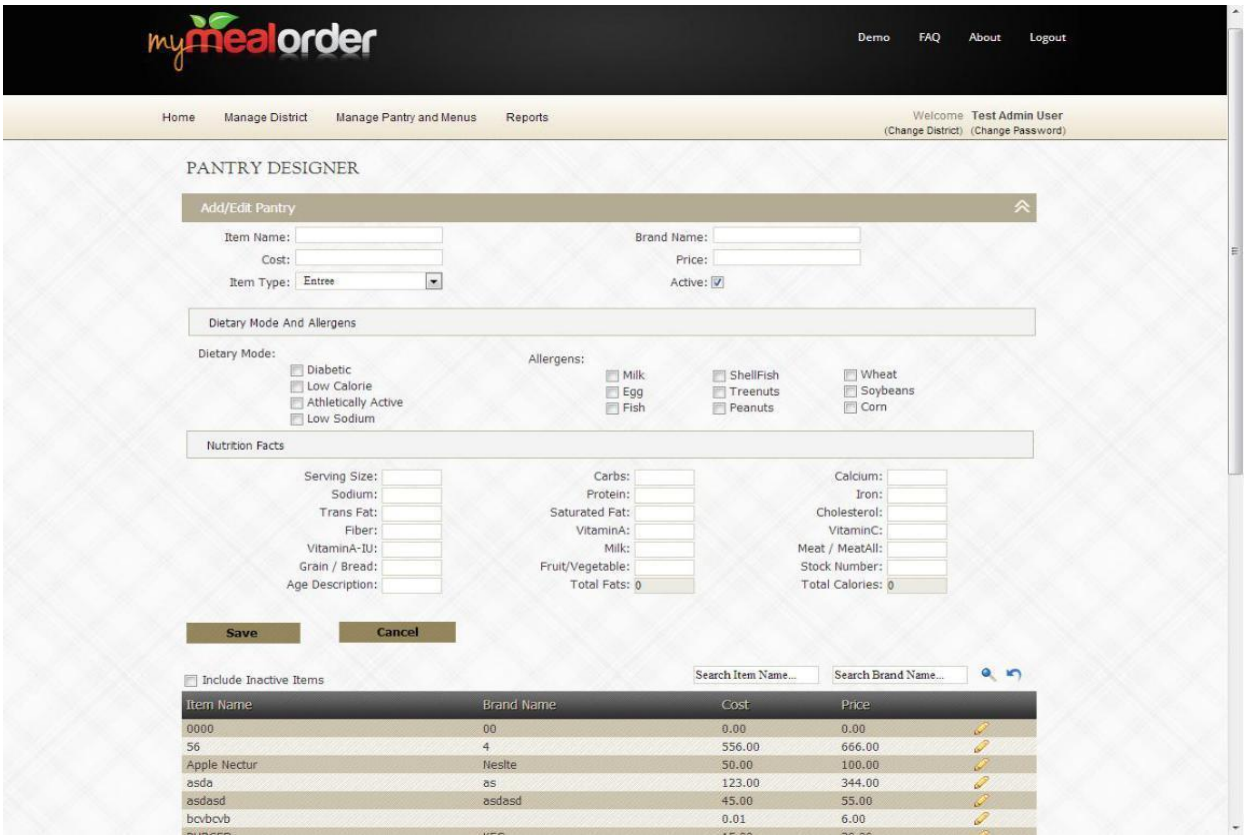


Admin Management Screen

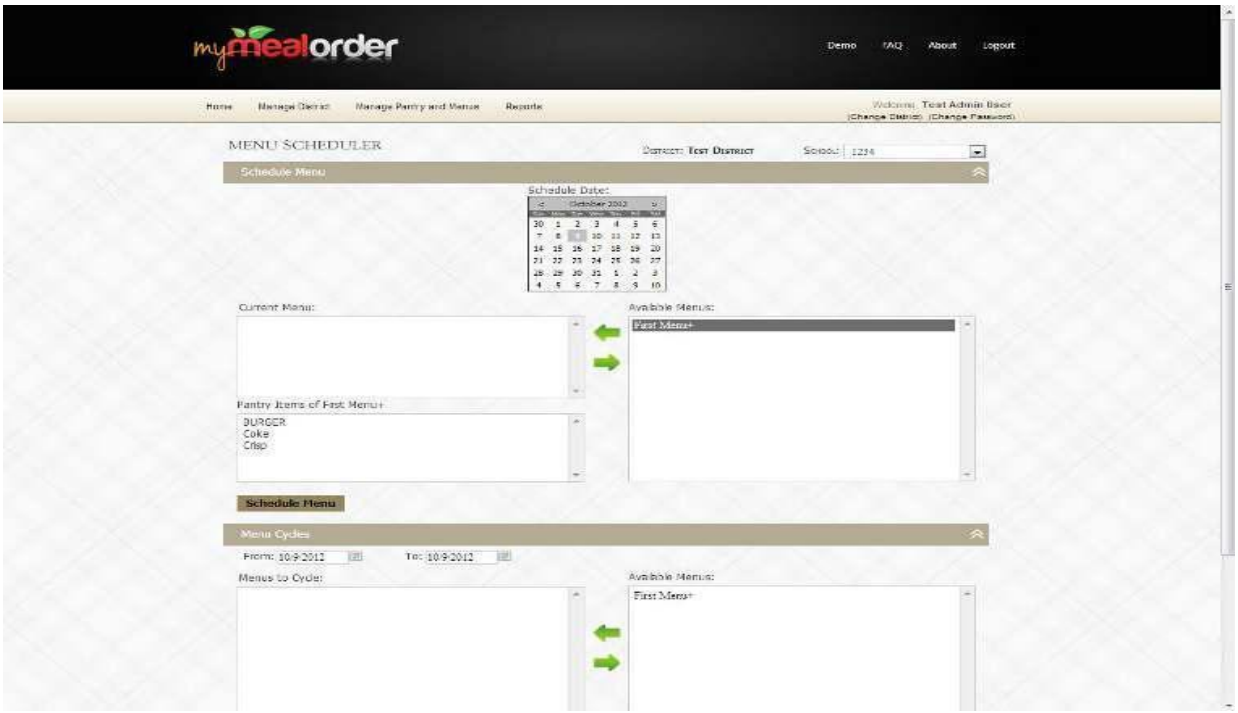


www.techno-soft.com | info@techno-soft.com | (203) 676-8299

Admin Pantry Designer Screen

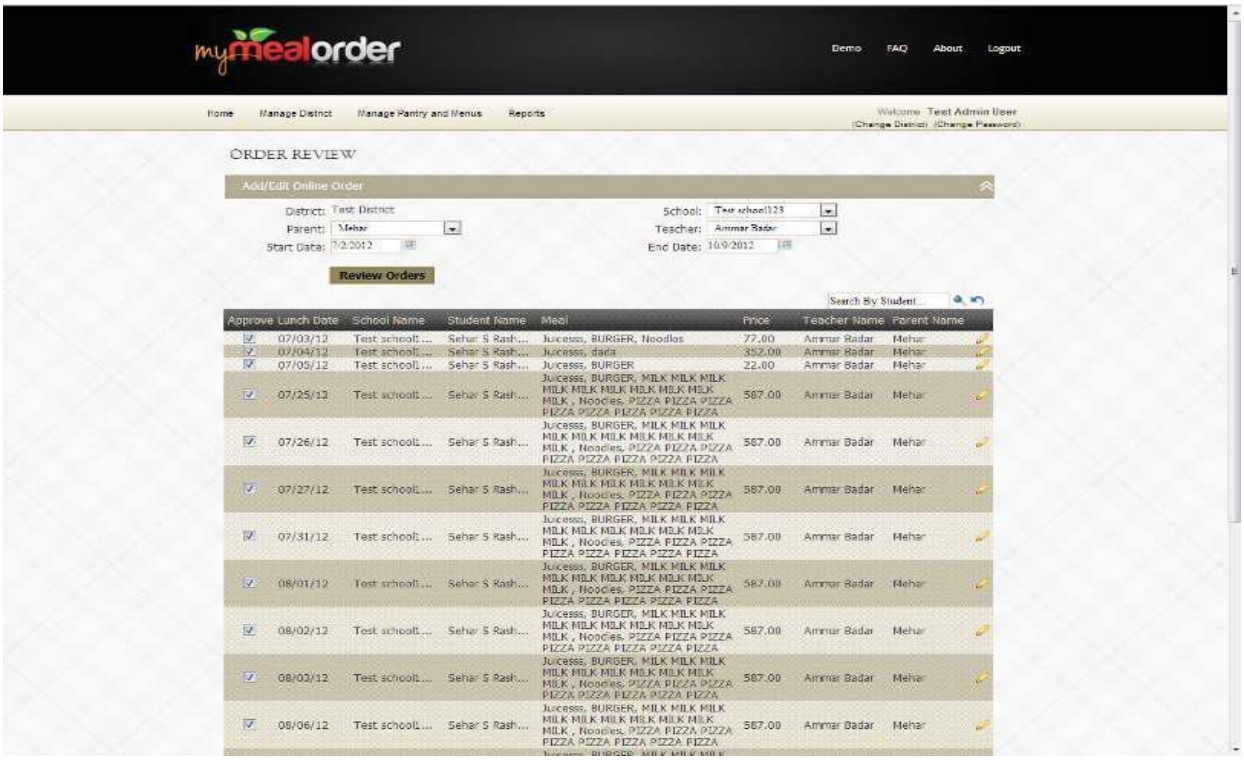


Admin Menu Scheduler Screen

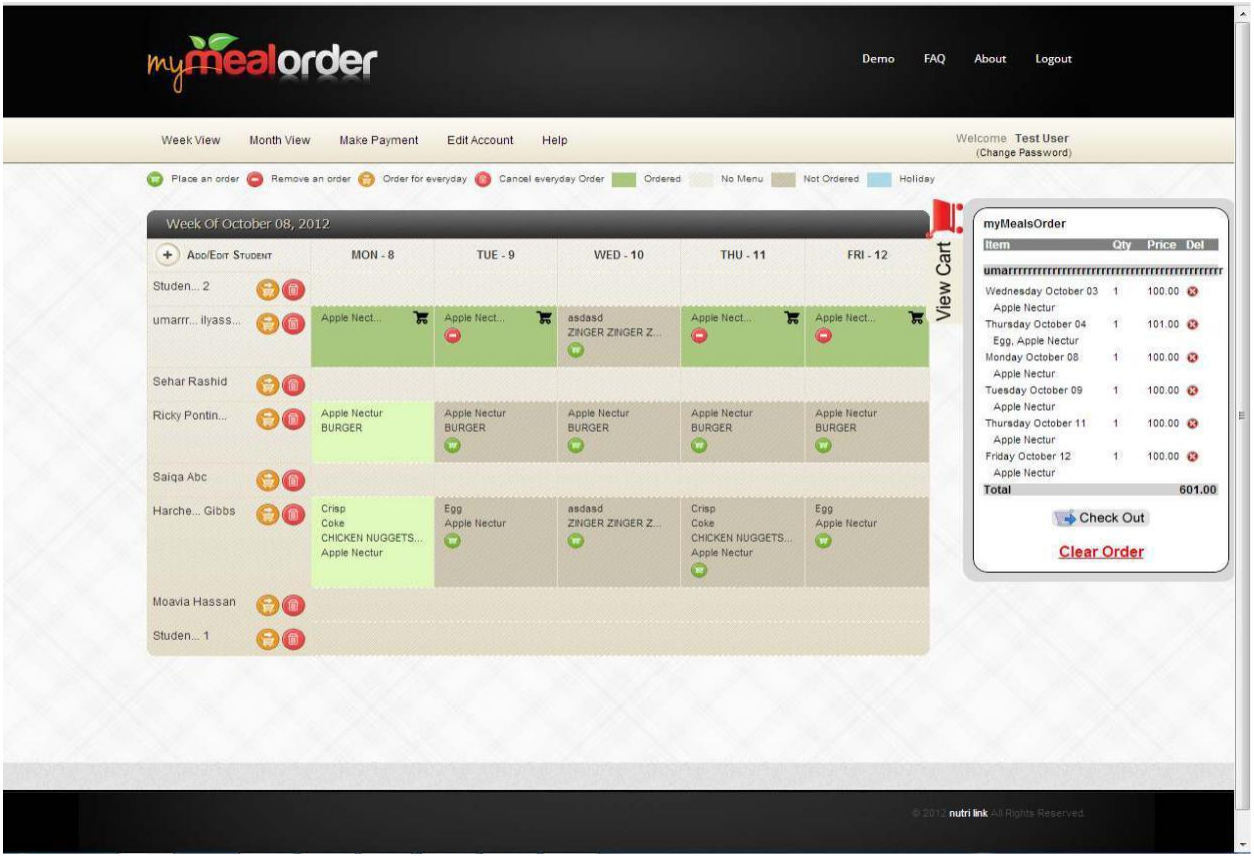


www.techno-soft.com | info@techno-soft.com | (203) 676-8299

Admin Order Review Screen

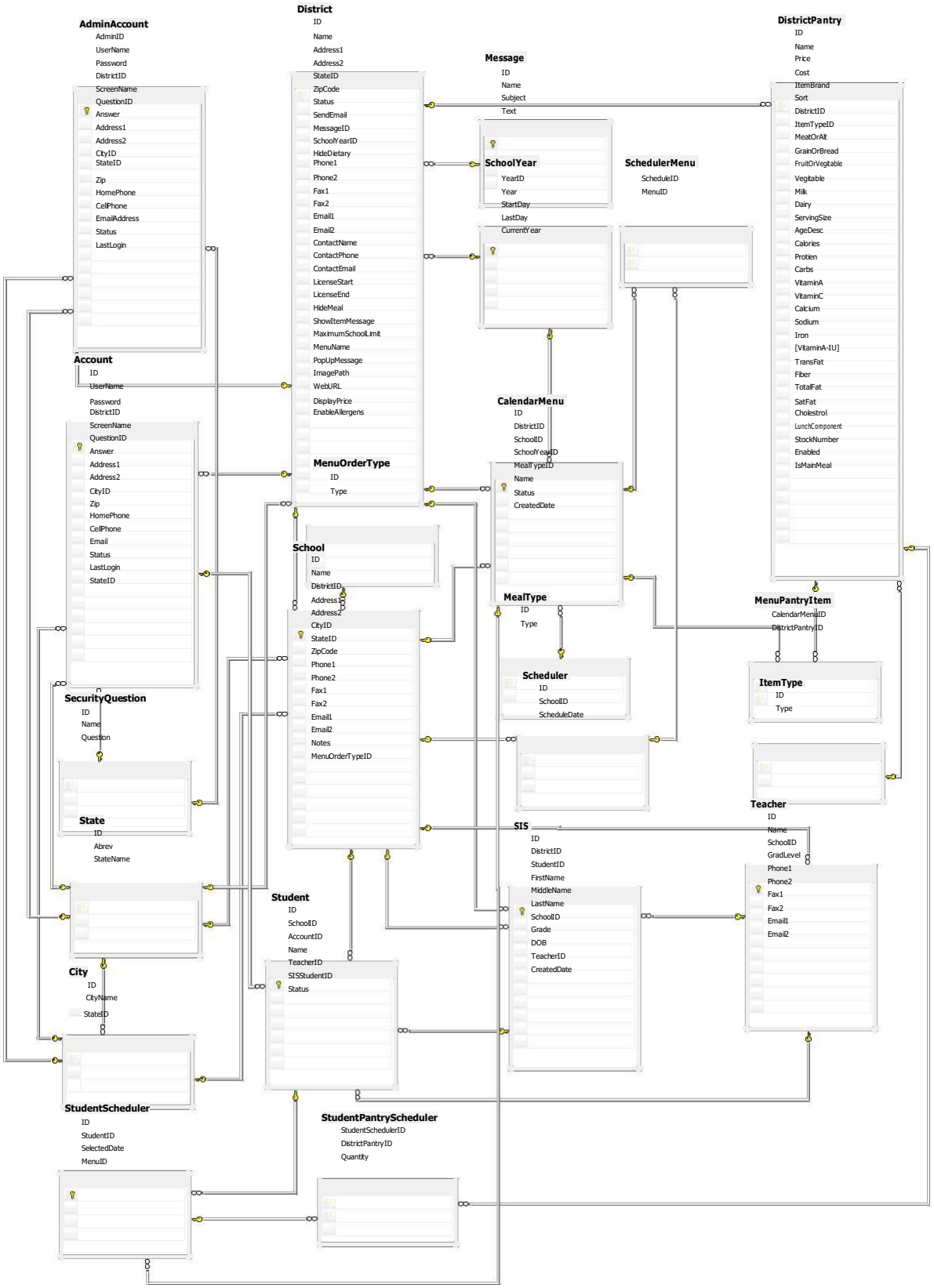


Parent Cart View Screen



www.techno-soft.com | info@techno-soft.com | (203) 676-8299

3 Detailed Design



www.techno-soft.com | info@techno-soft.com | (203) 676-8299

3.1 Detailed Class Diagram of Package/Layer/Component 1 (Optional)

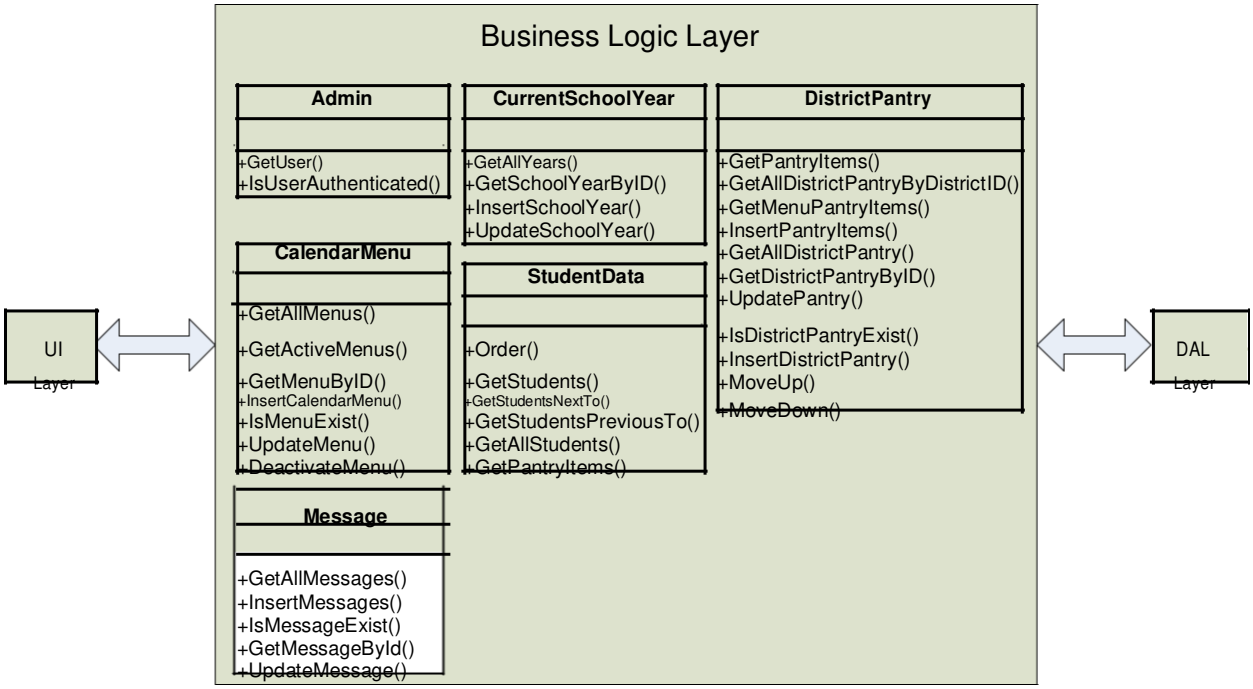


Figure 2: Class Diagram

3.2 System Requirements and Dependencies

1. Stand-alone web app using current .NET technologies
2. The application must run on all common and supported versions of web browsers
    a. Microsoft Internet Explorer
    b. Mozilla Firefox
    c. Google Chrome
    d. Safari (Mac and Windows)
3. Compatible with current and supported Windows operating systems
    a. Windows
    b. Apple
    c. Linux
4. Database to reside on the current CAFÉ 2008 SQL Server 2008 (nutridb.mdf, nutri-cafe.com)
5. Use as much of the current Nutri-Café application as possible (http://nutri-cafe.com/Admin/Login.aspx)

www.techno-soft.com | info@techno-soft.com | (203) 676-8299

a. Admin Panel
    i. add items
    ii. create menus
    iii. schedule cycles
b. Import file configurator

Future Relationship

Our client continues with us with great success and happiness and we are planning to execute multiple projects for them.

www.techno-soft.com_| info@techno-soft.com | (203) 676-8299

AbbyyOCR

Description

The purpose of this document is to define the scope of the project and the solution provided to the client.

3.3 Requirements

The purpose of the application was to parse the OCR data that had been generated by the Fine Reader application. The output of the application was an Excel document that contained the OCR data. System was designed to identify errors as RED color. Error identification was done using Regular Expressions. C# Application also split and merged the columns where it would find it necessary and mark the corrections as YELLOW color. The resulted output file would be very much near to the PDF document.

4 High Level Features

- **How to Import a Task in Fine Reader**

1. **Click on "New Task" bu**tton on Main Screen.

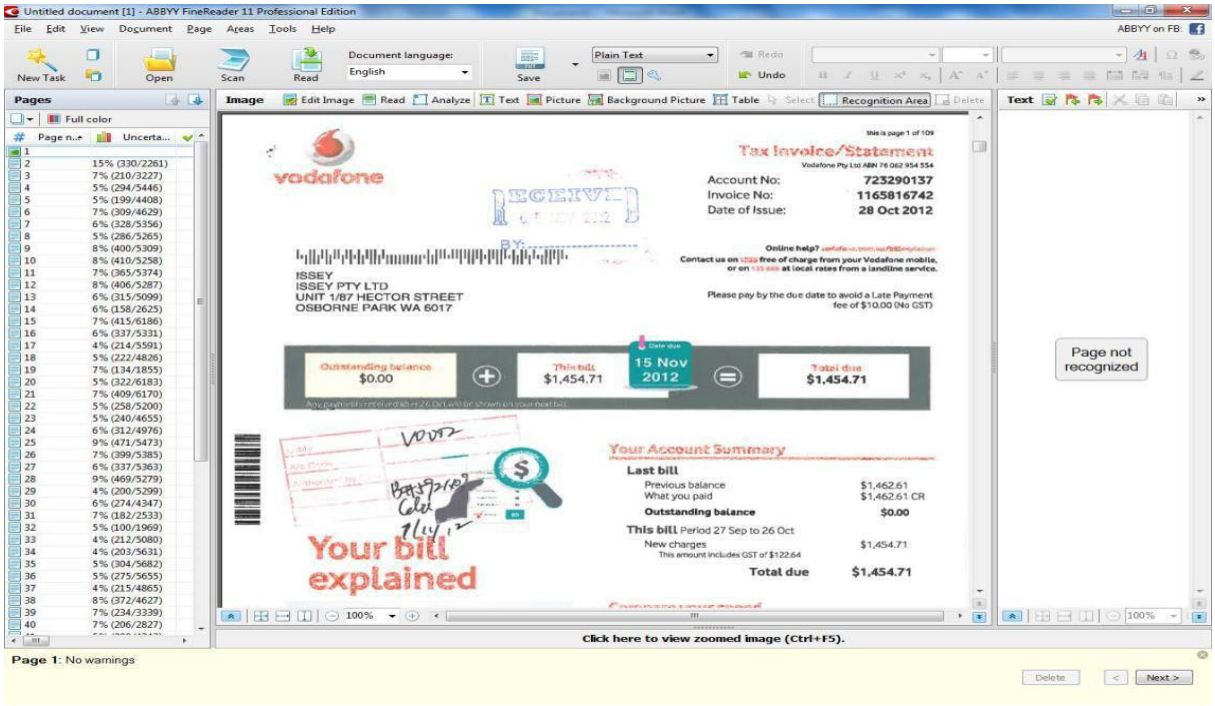


Figure 3: Main Screen

2. **"New Task" screen appears. This screen listed all the tasks that had been created/imported by user. To import a task, click on "Import" button.**

www.techno-soft.com | info@techno-soft.com | (203) 676-8299

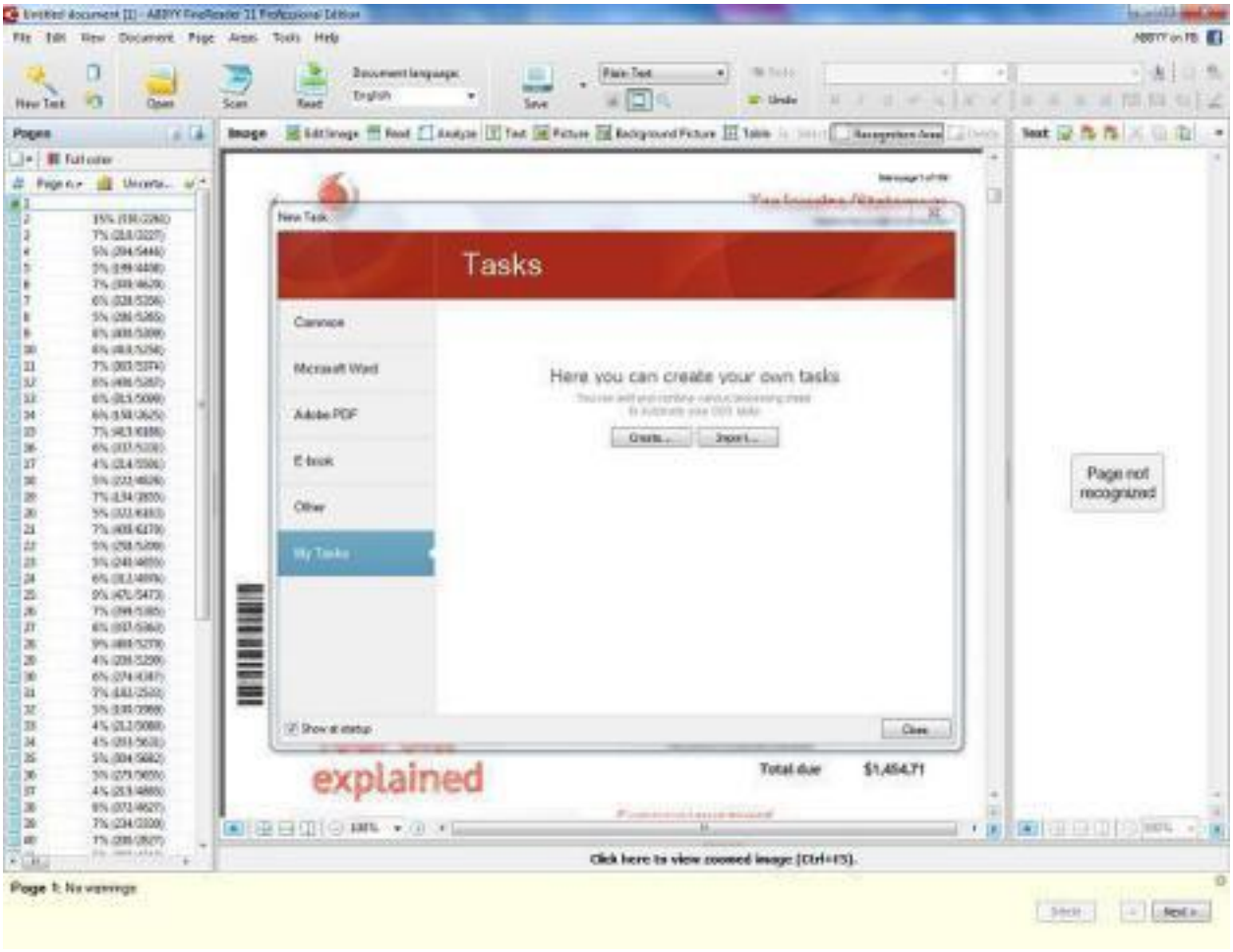


Figure 4: New Task Screen

- **How to run the task to generate OCR data**

1. **Select "AbbyyOCRTask" file from the Task folder, provided as a sample task file.**
2. **The task would be listed in "New Task" screen.**
3. **Now open "Vodafone_Sample_PDF.pdf" from the "OCR Sample" folder.**
4. Click on **"New Task" button in Main Screen. Click on the task with the name AbbyyOCRTask".**
5. **User would be prompted to select an "Area Template". Select "Vodafone_Sample_PDF.blk" file from "Area templates" folder.**
6. Fine Reader would start analyzing the opened document and after analyzing recognition process would start.
7. After recognition process was completed, user would be prompted to provide the name and type of file. In File name in**put box, provided the output file name. Select "Text" in "Save as type" option.**

www.techno-soft.com | info@techno-soft.com | (203) 676-8299

8. **In File options, select "Create a separate file for each page" option. Click on Save button.**
Fine Reader would save the output file in text format with the numbers added as postfix.

See the following screen:

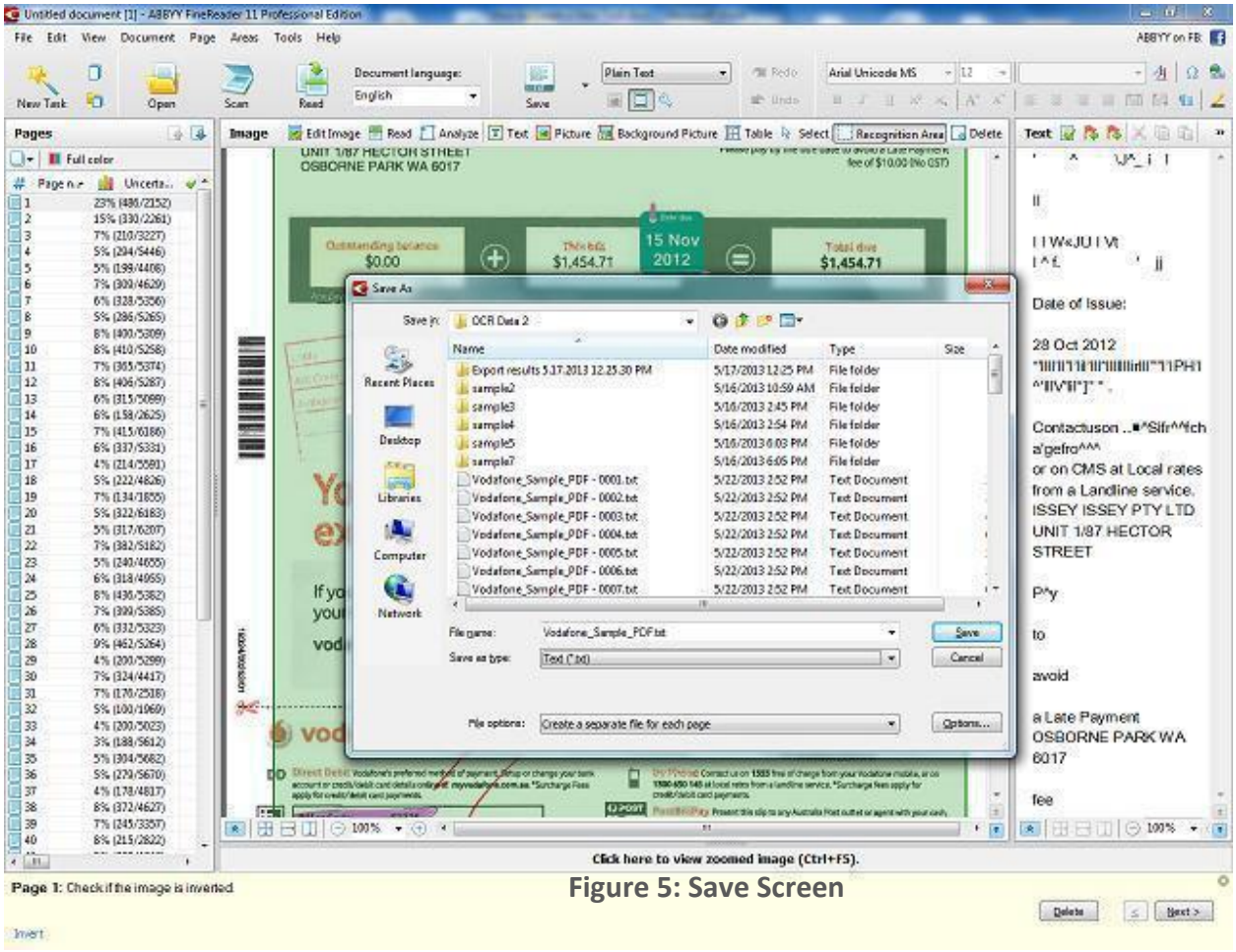
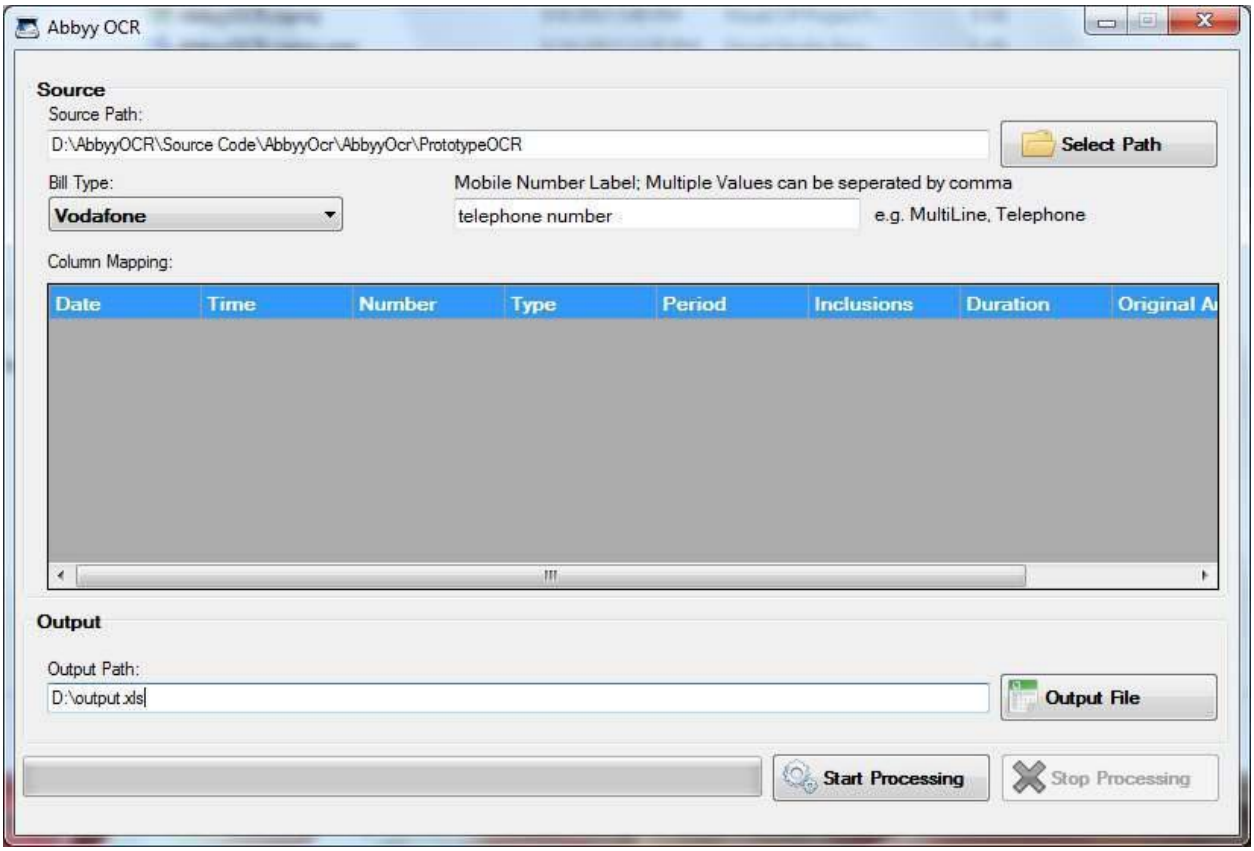


Figure 5: Save Screen

www.techno-soft.com | info@techno-soft.com | (203) 676-8299

- **How to Create an Excel output**

1. Install and run AbbyyOCR Application.



2. Browse for the folder that contained the data that was generated by Fine Reader using **"Select Path" button.**
3. **Provide the output path for the excel file using "Output File" button.**
4. **Click on "Start Processing" button. Application would start parsing the text files in the** selected folder and at the end generated an excel document in output file. The progress indicator showed the number of files that were being processed.
5. **You can stop processing using "Stop Processing" button.**

Technologies Used

- .NET Framework 3.5
- C#
- LINQ
- Excel Interop 12.0
- Microsoft Visual Studio 2010

www.techno-soft.com | info@techno-soft.com | (203) 676-8299

5 Application Prerequisites

1. Windows 7 32 bit.
2. .NET Framework 3.5.
3. Microsoft Office Excel 2007.

6 Application Configuration

Configuration.ini file contained the application configuration. This file would be located in **application's installer directory. It contained the items that are** in Period, Inclusions, and MobieType columns of the Telephone bill PDF document. We needed these items to correctly **parse the OCR data. See the following sample "Configuration.ini" file. It contained all the items** that you would need for the samples you have provided us. You need to modify it if the PDF document had different items that are not listed in it.

Configuration.ini

About Technosoft Solutions Inc.

Technosoft Solutions Inc. is leading outsourcing software and application development Company dedicated in Healthcare Software Development and Integration Services. We are offering a full range of IT solution and support services to companies developing healthcare solution for the U.S market. For more than 10 years, our commitment to providing reliable healthcare system software solutions has remained constant and stable.

Many of our clients are well versed in software development and have healthcare solutions deployed at main hospitals. We augment their resources with our HL7 integration, HIPAA privacy & security services.

We are endowed with quality certifications and standards, including CMMI ML2, ISO 9001:2008, Sun, MySQL and as IBM Certified Partner.

Technosoft Solutions Inc. | Offshore Development
w: http://www.techno-soft.com
e: info@techno-soft.com
t: (203) 676-8299