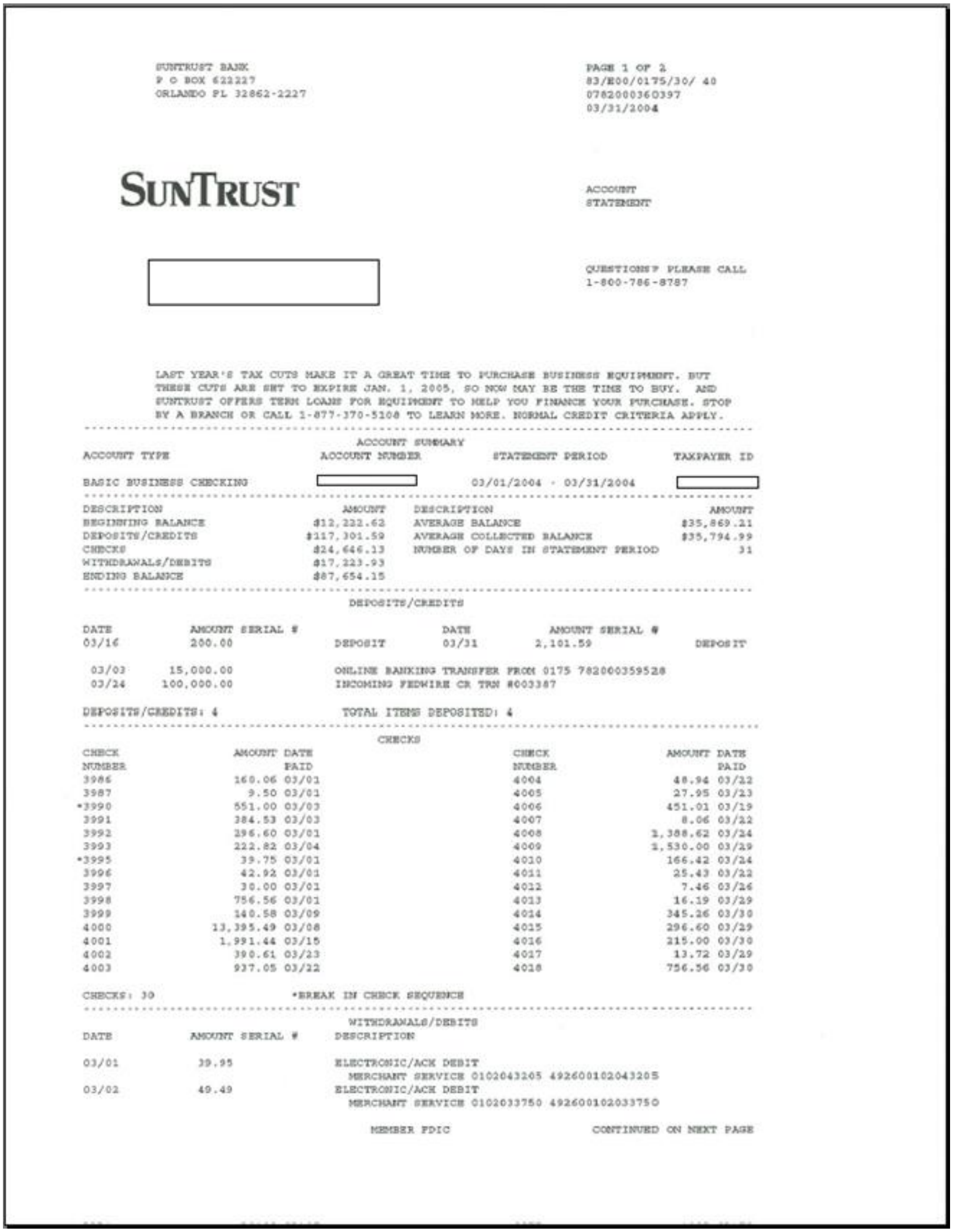
The BankScan Program

If you have to work with financial documents obtained by outside sources you probably understand the difficulty involved in turning such documents into an electronic form suitable for analysis. Certainly investigators and forensic accountants working money laundering and fraud cases deal with a large amount of bank statements obtained through subpoenas. Banks mostly provide these statements as paper or PDF files. Usually the PDF files are just scanned images of the statements, but they can also contain the underlying textual data as well. The task of the investigator is to organize and analyze this financial data, a task best done if the data is normalized into a common electronic format like a spreadsheet or database.

How does one do this? With the advent of inexpensive scanners and accurate optical character recognition (OCR) programs such as OmniPage or Abbyy FineReader, one can quickly convert paper or PDF images into electronic text files. But these text files are raw and unstructured. OCR programs can attempt to tabularize what is recognized, but given the great variety of formats that banks present their statements in, the results normally require extensive manual corrections.

What is desired is an automated method to extract the specific financial transaction data from these documents. BankScan does just that; it is an expert system that uses a library of templates that encapsulate the knowledge about how different banks format their statements. Using BankScan is straightforward. First an off the shelf OCR program is used to create text files from paper or PDF files. These files are input into BankScan, along with the template to use for extracting the data. BankScan then creates a normalized output in several possible formats, such as Excel, CSV, or QIF.

If an OCR program's text recognition accuracy were 100%, BankScan would literally be a "One-Click" operation. Unfortunately banks often provide sub-par quality documents. For example, the figures below show what kind of documents banks can provide. Figure 1 is very clean and legible, and can be expected to translate very accurately into text. Figure 2, because of the small font and poor quality reproduction, will have much less accuracy. This means that the resulting text file will likely contain character errors and other "garbage" text. It may be that an amount like $100.87 is recognized as S100.37. BankScan has extensive error checking capabilities and a means for the operator to make corrections to the files it processes. The key to getting the most success from BankScan is to provide it with the best possible input, i.e. getting the most accurate recognition from whatever OCR program is being used.



SUNTRUST BANK
P O BOX 622227
ORLANDO FL 32862-2227

PAGE 1 OF 2
83/E00/0175/30/ 40
0782000360397
03/31/2004

SUNTRUST

ACCOUNT
STATEMENT

[redacted]

QUESTIONS? PLEASE CALL
1-800-786-8787

LAST YEAR'S TAX CUTS MAKE IT A GREAT TIME TO PURCHASE BUSINESS EQUIPMENT. BUT THESE CUTS ARE SET TO EXPIRE JAN. 1, 2005, SO NOW MAY BE THE TIME TO BUY. AND SUNTRUST OFFERS TERM LOANS FOR EQUIPMENT TO HELP YOU FINANCE YOUR PURCHASE. STOP BY A BRANCH OR CALL 1-877-370-5108 TO LEARN MORE. NORMAL CREDIT CRITERIA APPLY.

ACCOUNT SUMMARY

ACCOUNT TYPE	ACCOUNT NUMBER	STATEMENT PERIOD	TAXPAYER ID
BASIC BUSINESS CHECKING	[redacted]	03/01/2004 - 03/31/2004	[redacted]

DESCRIPTION	AMOUNT	DESCRIPTION	AMOUNT
BEGINNING BALANCE	$12,222.62	AVERAGE BALANCE	$35,869.21
DEPOSITS/CREDITS	$117,301.59	AVERAGE COLLECTED BALANCE	$35,794.99
CHECKS	$24,646.13	NUMBER OF DAYS IN STATEMENT PERIOD	31
WITHDRAWALS/DEBITS	$17,223.93		
ENDING BALANCE	$87,654.15		

DEPOSITS/CREDITS

DATE	AMOUNT	SERIAL #		DATE	AMOUNT	SERIAL #	
03/16	200.00		DEPOSIT	03/31	2,101.59		DEPOSIT

03/03	15,000.00	ONLINE BANKING TRANSFER FROM 0175 782000359528
03/24	100,000.00	INCOMING FEDWIRE CR TRN #003387

DEPOSITS/CREDITS: 4 TOTAL ITEMS DEPOSITED: 4

CHECKS

CHECK NUMBER	AMOUNT	DATE PAID	CHECK NUMBER	AMOUNT	DATE PAID
3986	160.06	03/01	4004	48.94	03/22
3987	9.50	03/01	4005	27.95	03/23
*3990	551.00	03/03	4006	451.01	03/19
3991	384.53	03/03	4007	8.06	03/22
3992	296.60	03/01	4008	1,388.62	03/24
3993	222.82	03/04	4009	1,530.00	03/29
*3995	39.75	03/01	4010	166.42	03/24
3996	42.92	03/01	4011	25.43	03/22
3997	30.00	03/01	4012	7.46	03/26
3998	756.56	03/01	4013	16.19	03/29
3999	140.58	03/09	4014	345.26	03/30
4000	13,395.49	03/08	4015	296.60	03/29
4001	1,991.44	03/15	4016	215.00	03/30
4002	390.61	03/23	4017	13.72	03/29
4003	937.05	03/22	4018	756.56	03/30

CHECKS: 30 *BREAK IN CHECK SEQUENCE

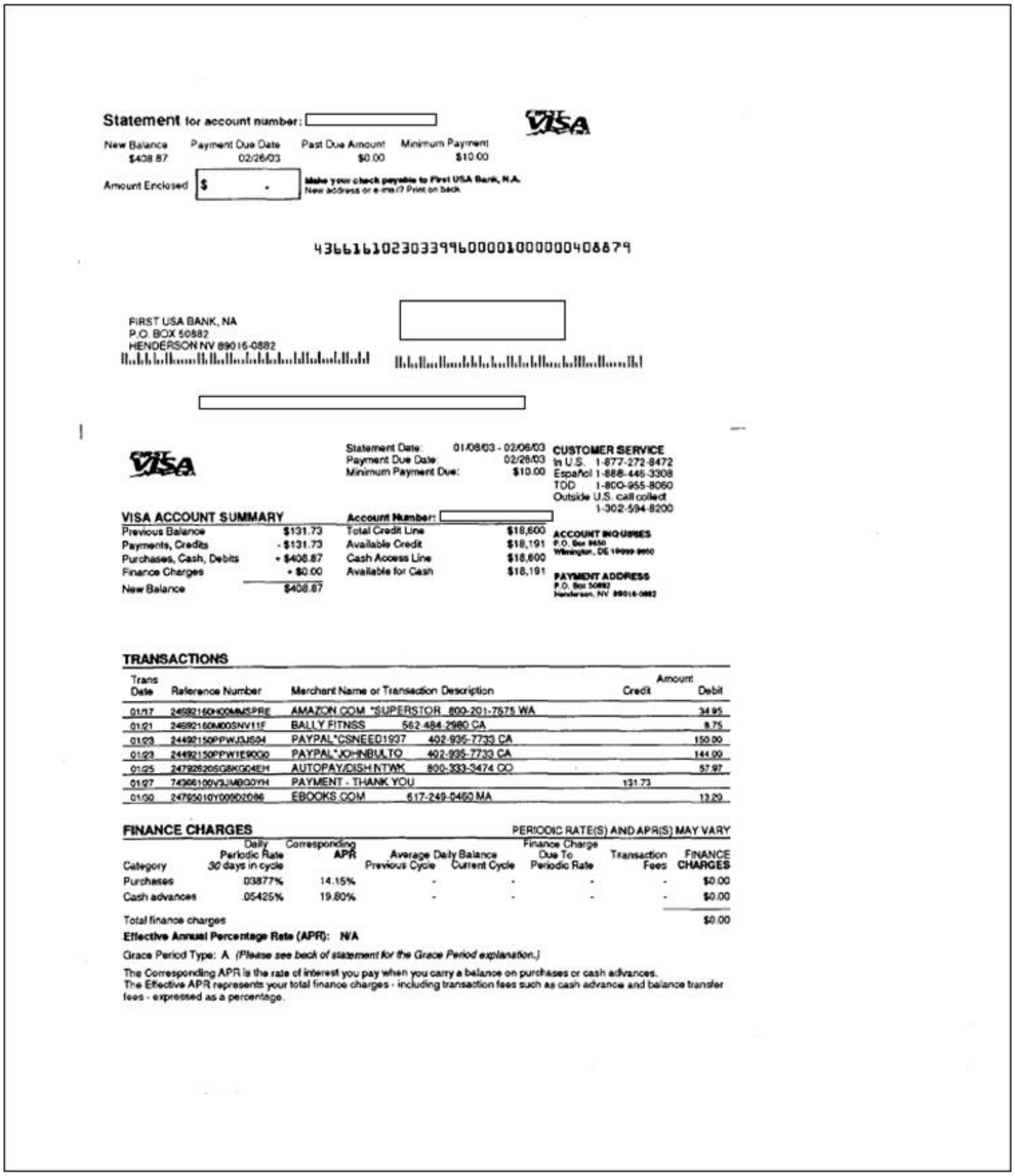
WITHDRAWALS/DEBITS

DATE	AMOUNT	SERIAL #	DESCRIPTION
03/01	39.95		ELECTRONIC/ACH DEBIT MERCHANT SERVICE 0102043205 492600102043205
03/02	49.49		ELECTRONIC/ACH DEBIT MERCHANT SERVICE 0102033750 492600102033750

MEMBER FDIC

CONTINUED ON NEXT PAGE

Figure 1. Ideal statement – uniform legible font, minimal graphics, clean background



Statement for account number:

New Balance	Payment Due Date	Past Due Amount	Minimum Payment
$408 87	02/26/03	$0.00	$10.00

Amount Enclosed $.

Make your check payable to First USA Bank, N.A.
New address or e-mail? Print on back

436616102303399600001000000408879

FIRST USA BANK, NA
P.O. BOX 50882
HENDERSON NV 89016-0882

Statement Date:	01/08/03 - 02/06/03
Payment Due Date:	02/26/03
Minimum Payment Due:	$10.00

CUSTOMER SERVICE

In U.S.	1-877-272-8472
Español	1-888-445-3308
TDD	1-800-955-8060
Outside U.S. call collect	1-302-594-8200

VISA ACCOUNT SUMMARY

Previous Balance	$131.73
Payments, Credits	- $131.73
Purchases, Cash, Debits	+ $408.87
Finance Charges	+ $0.00
New Balance	$408.87

Account Number:

Total Credit Line	$18,600
Available Credit	$18,191
Cash Access Line	$18,600
Available for Cash	$18,191

ACCOUNT INQUIRIES
P.O. Box 8650
Wilmington, DE 19899-8650

PAYMENT ADDRESS
P.O. Box 50882
Henderson, NV 89016-0882

TRANSACTIONS

Trans Date	Reference Number	Merchant Name or Transaction Description	Amount Credit	Amount Debit
01/17	24692160H00MMSPRE	AMAZON.COM *SUPERSTOR 800-201-7575 WA		34.95
01/21	24692160M00SNV11F	BALLY FITNSS 562-484-2980 CA		8.75
01/23	24492150PPWJ3JS04	PAYPAL*CSNEED1937 402-935-7733 CA		150.00
01/23	24492150PPW1E9GG0	PAYPAL*JOHNBULTO 402-935-7733 CA		144.00
01/25	24792620SG8KG04EH	AUTOPAY/DISH NTWK 800-333-3474 CO		57.97
01/27	74966100V3JMBQ0YH	PAYMENT - THANK YOU	131.73	
01/30	24795010Y00SD2066	EBOOKS.COM 617-249-0460 MA		13.20

FINANCE CHARGES PERIODIC RATE(S) AND APR(S) MAY VARY

Category	Daily Periodic Rate *30* days in cycle	Corresponding **APR**	Average Daily Balance Previous Cycle	Average Daily Balance Current Cycle	Finance Charge Due To Periodic Rate	Transaction Fees	FINANCE **CHARGES**
Purchases	03877%	14.15%	-	-	-	-	$0.00
Cash advances	.05425%	19.80%	-	-	-	-	$0.00
Total finance charges							$0.00

Effective Annual Percentage Rate (APR): N/A

Grace Period Type: A *(Please see back of statement for the Grace Period explanation.)*

The Corresponding APR is the rate of interest you pay when you carry a balance on purchases or cash advances.
The Effective APR represents your total finance charges - including transaction fees such as cash advance and balance transfer fees - expressed as a percentage.

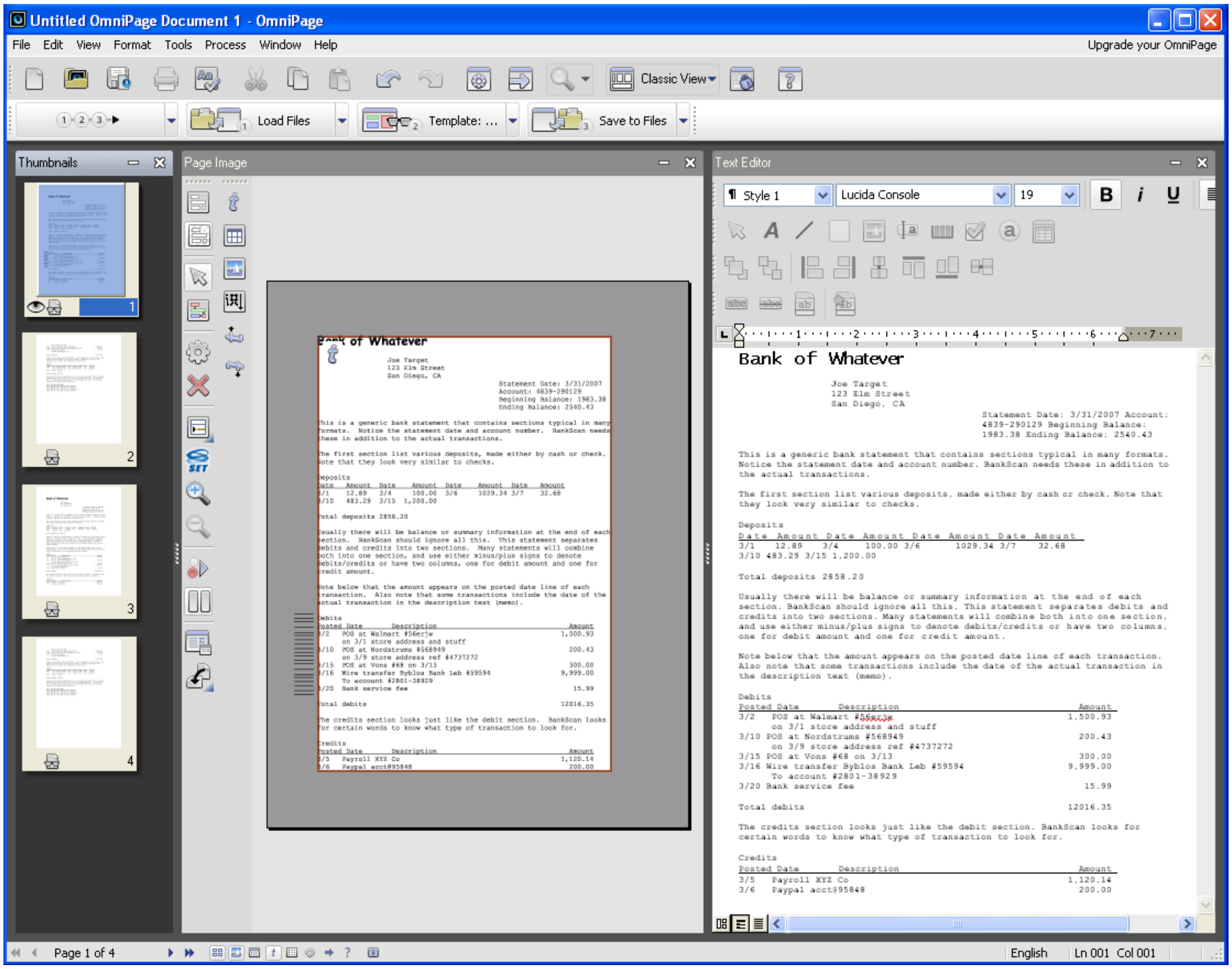
Figure 2. Poor statement – tiny illegible font

BankScan does not attempt to re-invent the wheel with OCR, a field that has been extensively researched for decades. Millions of lines of source code have been written to create commercial products reaching a very high level of accuracy. Our experience from scanning hundreds of different bank statement formats has determined that the OmniPage program sold by Nuance combines the most accurate recognition with a very easy to learn user interface. Features in OmniPage such as the ability to zone specific areas for recognition and a simple means of training to improve accuracy make this the recommended program to use with BankScan.

OmniPage and BankScan are not integrated together, they are standalone applications. For example, if an office has only a few scanners, then several scanning stations can be set up with a copy of OmniPage at each one. An operator can scan and recognize their statements then take the text files to their desk and use BankScan there. In this sense BankScan does not attempt to be a “systems solution”. It is not an evidence management database or analysis tool. It does one thing very well, convert unusable information into usable information.

BankScan Walkthrough

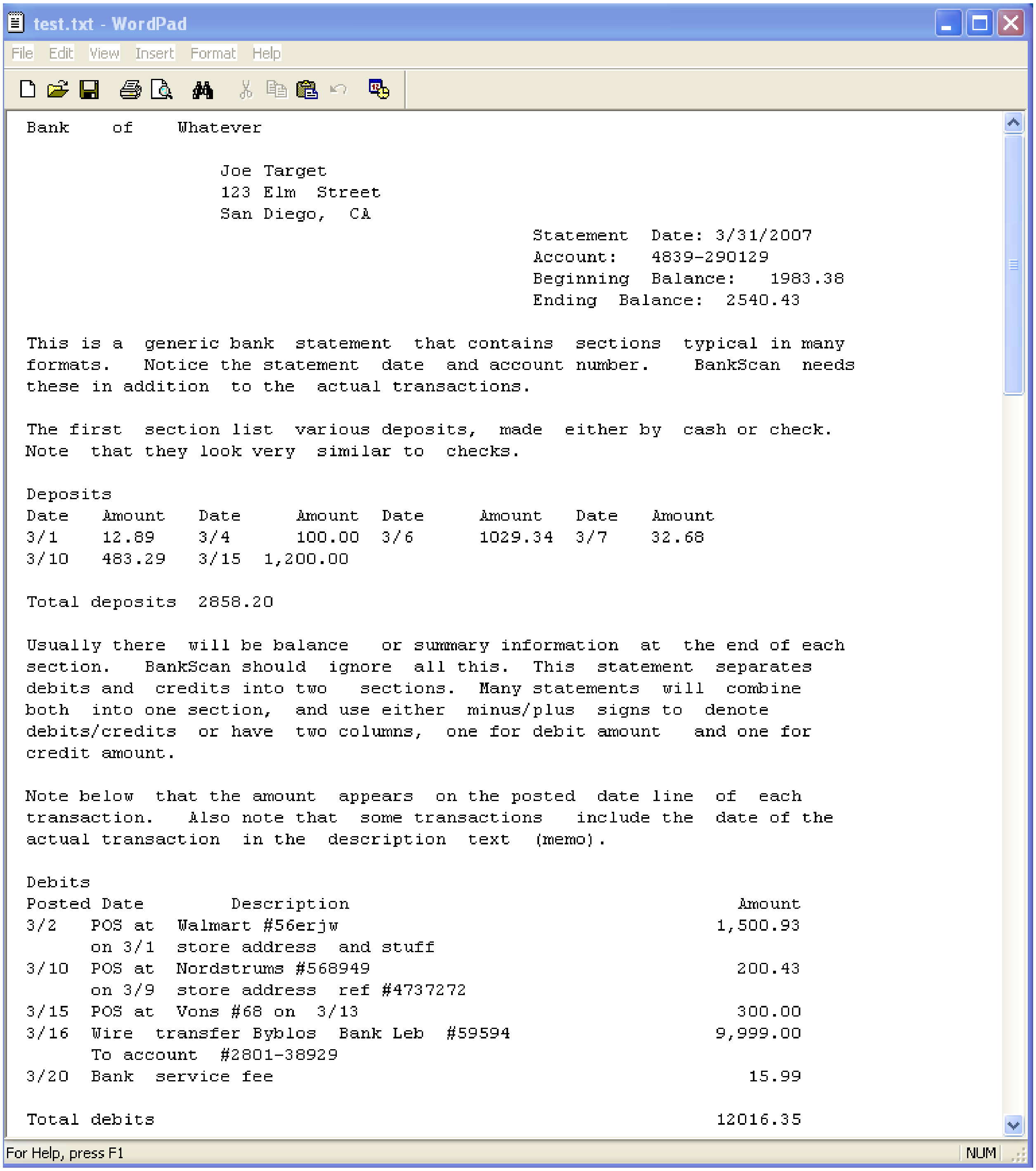
In this section we will walk through the operation of BankScan on a typical generic bank statement. The first step is to check the statements to be scanned. Statements should be in date order and checked for missing pages, duplicates; any issues that would complicate processing further down the line. Then the statements are scanned in OmniPage and converted into text files. Figure 3 shows a view of the OmniPage program in operation. The recognized text can be saved in many different formats, but for BankScan we just need a simple text (.TXT) file.



```
Bank of Whatever

                    Joe Target
                    123 Elm Street
                    San Diego, CA
                                                  Statement Date: 3/31/2007 Account:
                                                  4839-290129 Beginning Balance:
                                                  1983.38 Ending Balance: 2540.43

This is a generic bank statement that contains sections typical in many formats.
Notice the statement date and account number. BankScan needs these in addition to
the actual transactions.

The first section list various deposits, made either by cash or check. Note that
they look very similar to checks.

Deposits
Date  Amount  Date  Amount  Date  Amount  Date  Amount
3/1    12.89   3/4     100.00 3/6      1029.34 3/7    32.68
3/10 483.29 3/15 1,200.00

Total deposits 2858.20

Usually there will be balance or summary information at the end of each
section. BankScan should ignore all this. This statement separates debits and
credits into two sections. Many statements will combine both into one section,
and use either minus/plus signs to denote debits/credits or have two columns,
one for debit amount and one for credit amount.

Note below that the amount appears on the posted date line of each transaction.
Also note that some transactions include the date of the actual transaction in
the description text (memo).

Debits
Posted Date        Description                                        Amount
3/2   POS at Walmart #56erjw                                        1,500.93
      on 3/1 store address and stuff
3/10 POS at Nordstrums #568949                                        200.43
      on 3/9 store address ref #4737272
3/15 POS at Vons #68 on 3/13                                          300.00
3/16 Wire transfer Byblos Bank Leb #59594                           9,999.00
      To account #2801-38929
3/20 Bank service fee                                                  15.99

Total debits                                                        12016.35

The credits section looks just like the debit section. BankScan looks for
certain words to know what type of transaction to look for.

Credits
Posted Date        Description                                        Amount
3/5   Payroll XYZ Co                                                1,120.14
3/6   Paypal acct@95848                                               200.00
```

Figure 3. OmniPage graphical user interface

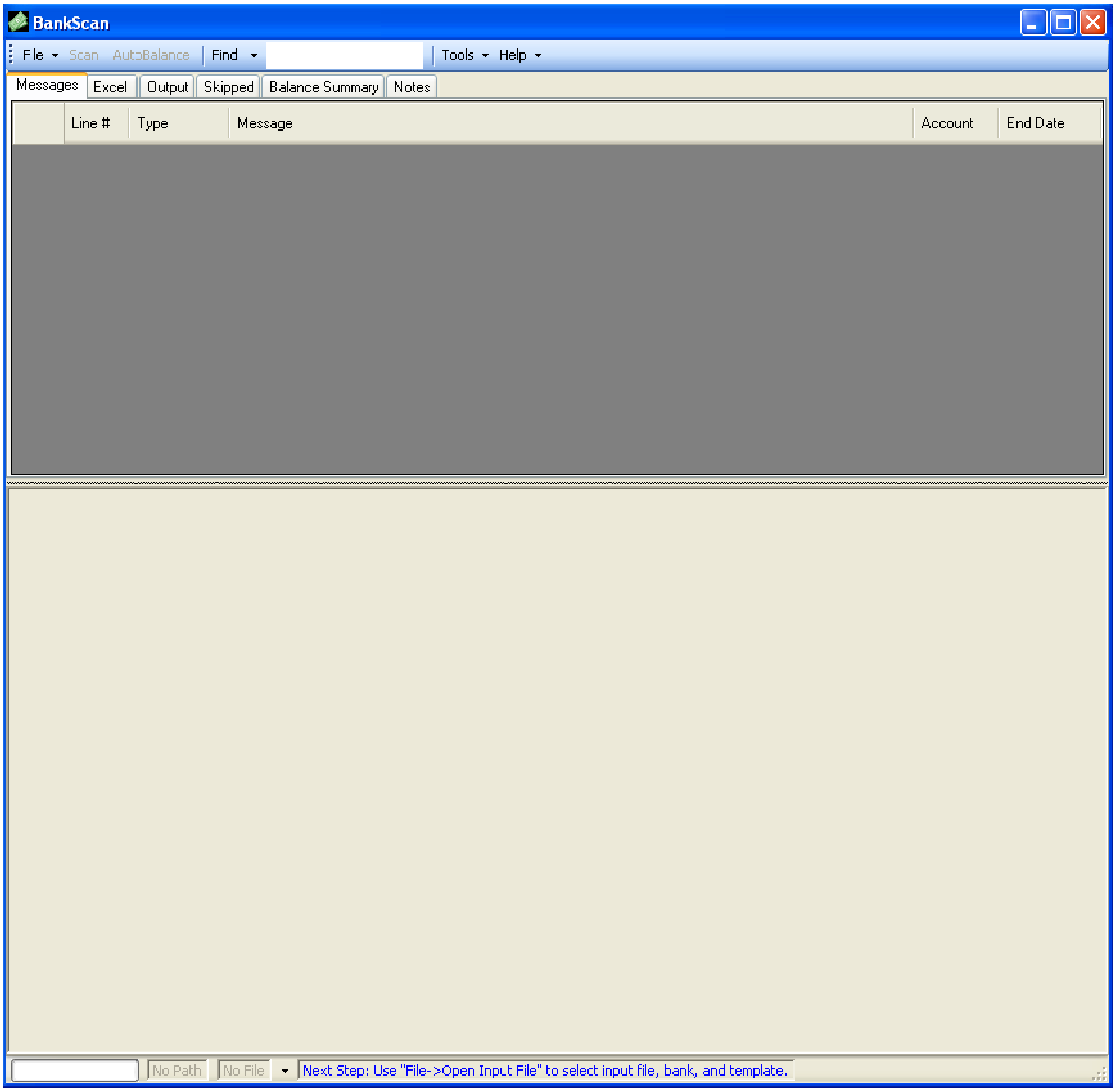
The resulting .TXT file is shown below in WordPad. Notice that OmniPage can preserve much of the original formatting of the original image. This is important when transactions (debits or credits) can only be distinguished by what columns their amounts fall under. The next step is to start BankScan and read in this file for processing.



```
Bank    of    Whatever

                  Joe Target
                  123 Elm  Street
                  San Diego,  CA
                                               Statement  Date: 3/31/2007
                                               Account:   4839-290129
                                               Beginning  Balance:   1983.38
                                               Ending  Balance:  2540.43

This is a  generic bank  statement  that contains  sections  typical in many
formats.   Notice the statement  date  and account number.    BankScan  needs
these in addition  to the  actual transactions.

The first  section list  various deposits,  made  either by  cash or check.
Note  that they look very  similar to  checks.

Deposits
Date   Amount    Date     Amount  Date     Amount   Date   Amount
3/1    12.89     3/4      100.00  3/6      1029.34  3/7    32.68
3/10   483.29    3/15  1,200.00

Total deposits  2858.20

Usually there  will be balance   or summary information  at  the end of each
section.   BankScan should  ignore  all this.   This  statement  separates
debits and  credits into two   sections.  Many statements  will  combine
both  into one section,  and use either  minus/plus  signs to  denote
debits/credits  or have  two columns,  one for debit amount   and one for
credit amount.

Note below  that the amount  appears  on the posted  date line  of  each
transaction.   Also note that  some transactions   include the  date of the
actual transaction  in the  description  text  (memo).

Debits
Posted Date        Description                                    Amount
3/2   POS at  Walmart #56erjw                                   1,500.93
      on 3/1  store address  and stuff
3/10  POS at  Nordstrums #568949                                  200.43
      on 3/9  store address  ref #4737272
3/15  POS at  Vons #68 on  3/13                                   300.00
3/16  Wire  transfer Byblos  Bank Leb  #59594                   9,999.00
      To account  #2801-38929
3/20  Bank  service fee                                            15.99

Total debits                                                    12016.35
```

Figure 4. The recognized text – notice formatting is preserved

The BankScan user interface consists of two main areas. The top half has several tabs that provide information about its operation and results, the lower half is a specialized text editor window for making corrections in less than 100% accurate files.

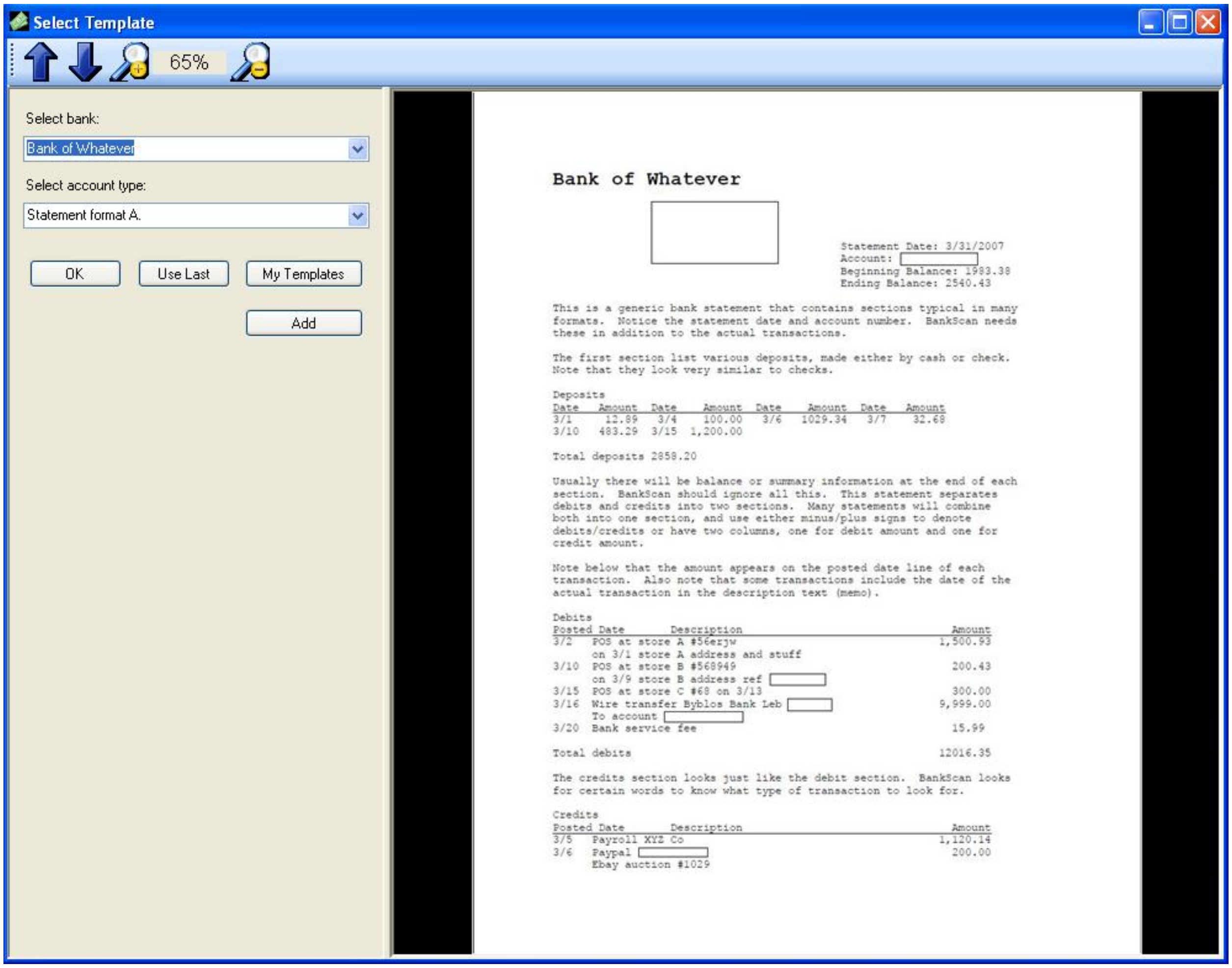


Next Step: Use "File->Open Input File" to select input file, bank, and template.

Figure 5. BankScan graphical user interface

The operator opens the text file to be processed, then selects the appropriate template from the BankScan library. This template tells BankScan everything it needs to know about how to extract the transactions out of a particular statement format. Each template has an associated image representing that format. Selecting the correct template is done by making a visual comparison of the statements to be processed against each template image for that bank. Early attempts to try and detect the correct template automatically proved infeasible, it is much faster to use the pattern matching capabilities of a human! Banks can have MANY different formats, and they are constantly changing them.

Figure 6 shows how the template is selected in BankScan. First the bank is selected from a drop-down list, and then the template images for that bank are checked against the statements being scanned. The closest matching image is selected. If none of the template images match the statement then a new template must be created and added to the library.



Select bank: Bank of Whatever

Select account type: Statement format A.

Bank of Whatever

Statement Date: 3/31/2007
Account:
Beginning Balance: 1983.38
Ending Balance: 2540.43

This is a generic bank statement that contains sections typical in many formats. Notice the statement date and account number. BankScan needs these in addition to the actual transactions.

The first section list various deposits, made either by cash or check. Note that they look very similar to checks.

Deposits

Date	Amount	Date	Amount	Date	Amount	Date	Amount
3/1	12.89	3/4	100.00	3/6	1029.34	3/7	32.68
3/10	483.29	3/15	1,200.00				

Total deposits 2858.20

Usually there will be balance or summary information at the end of each section. BankScan should ignore all this. This statement separates debits and credits into two sections. Many statements will combine both into one section, and use either minus/plus signs to denote debits/credits or have two columns, one for debit amount and one for credit amount.

Note below that the amount appears on the posted date line of each transaction. Also note that some transactions include the date of the actual transaction in the description text (memo).

Debits

Posted Date	Description	Amount
3/2	POS at store A #56erjw on 3/1 store A address and stuff	1,500.93
3/10	POS at store B #568949 on 3/9 store B address ref	200.43
3/15	POS at store C #68 on 3/13	300.00
3/16	Wire transfer Byblos Bank Leb To account	9,999.00
3/20	Bank service fee	15.99

Total debits 12016.35

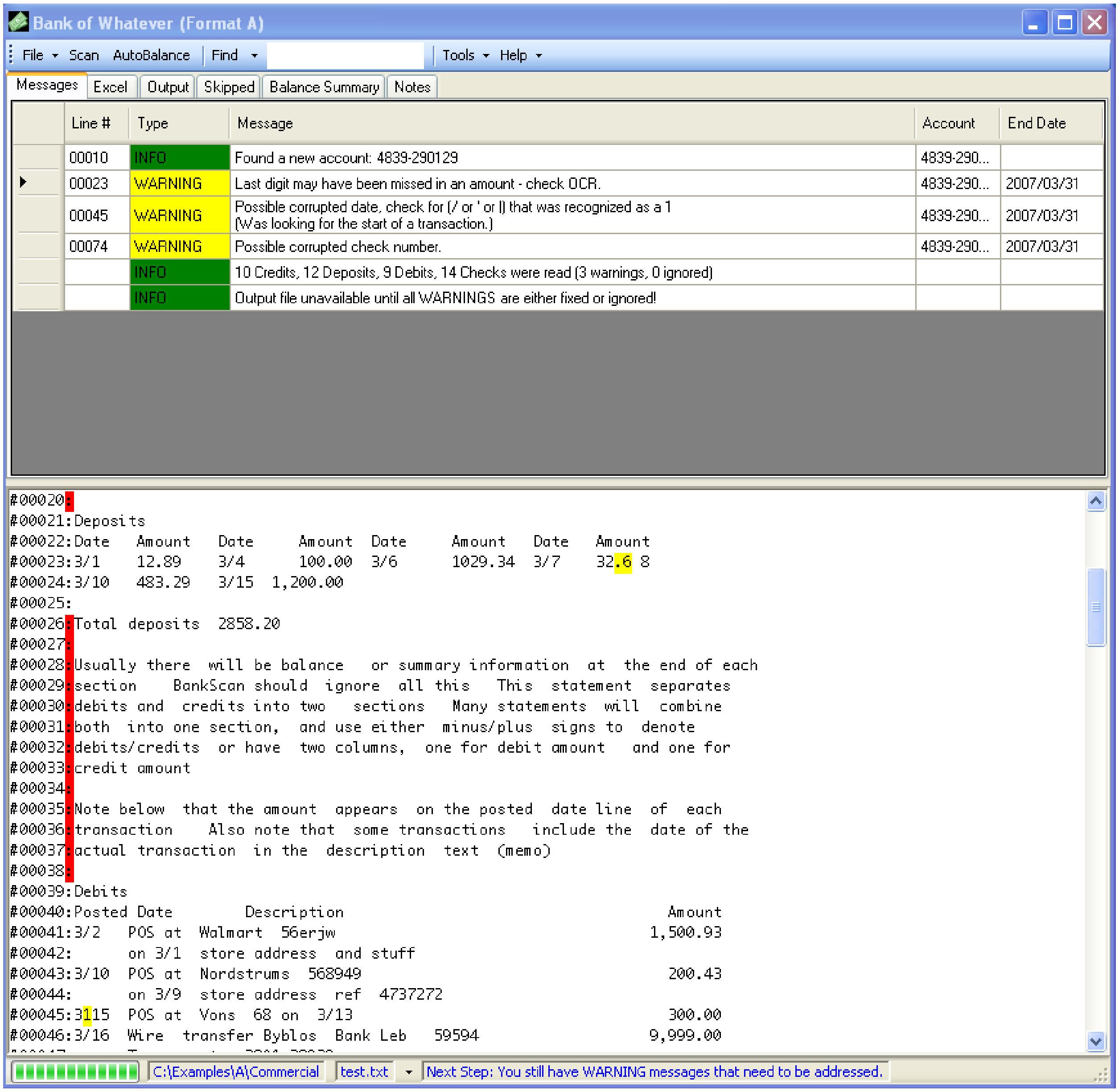
The credits section looks just like the debit section. BankScan looks for certain words to know what type of transaction to look for.

Credits

Posted Date	Description	Amount
3/5	Payroll XYZ Co	1,120.14
3/6	Paypal Ebay auction #1029	200.00

Figure 6. Template chooser

After the template is selected BankScan processes the text file and reports any issues it might have had extracting the transactions. It does this by displaying yellow warning messages in the Messages tab and marking the suspect area in the lower editing window. In Figure 7 we see that BankScan has found three issues that need operator attention. For example, at line #00045 a date 3/15 has been misrecognized to be 3115. In many cases BankScan knows what the problem is, but errs on the side of caution and requires operator verification. For recognition errors that are common to a particular format, the template can be built with automatic corrections. The task of the operator is to either make corrections in the editor window or tell BankScan to ignore the warning (again erring on the side of caution. BankScan can flag non-issues).



Bank of Whatever (Format A)
File ▾ Scan AutoBalance Find ▾ Tools ▾ Help ▾
Messages Excel Output Skipped Balance Summary Notes

	Line #	Type	Message	Account	End Date
	00010	INFO	Found a new account: 4839-290129	4839-290...	
▶	00023	WARNING	Last digit may have been missed in an amount - check OCR.	4839-290...	2007/03/31
	00045	WARNING	Possible corrupted date, check for (/ or ' or |) that was recognized as a 1 (Was looking for the start of a transaction.)	4839-290...	2007/03/31
	00074	WARNING	Possible corrupted check number.	4839-290...	2007/03/31
		INFO	10 Credits, 12 Deposits, 9 Debits, 14 Checks were read (3 warnings, 0 ignored)		
		INFO	Output file unavailable until all WARNINGS are either fixed or ignored!		

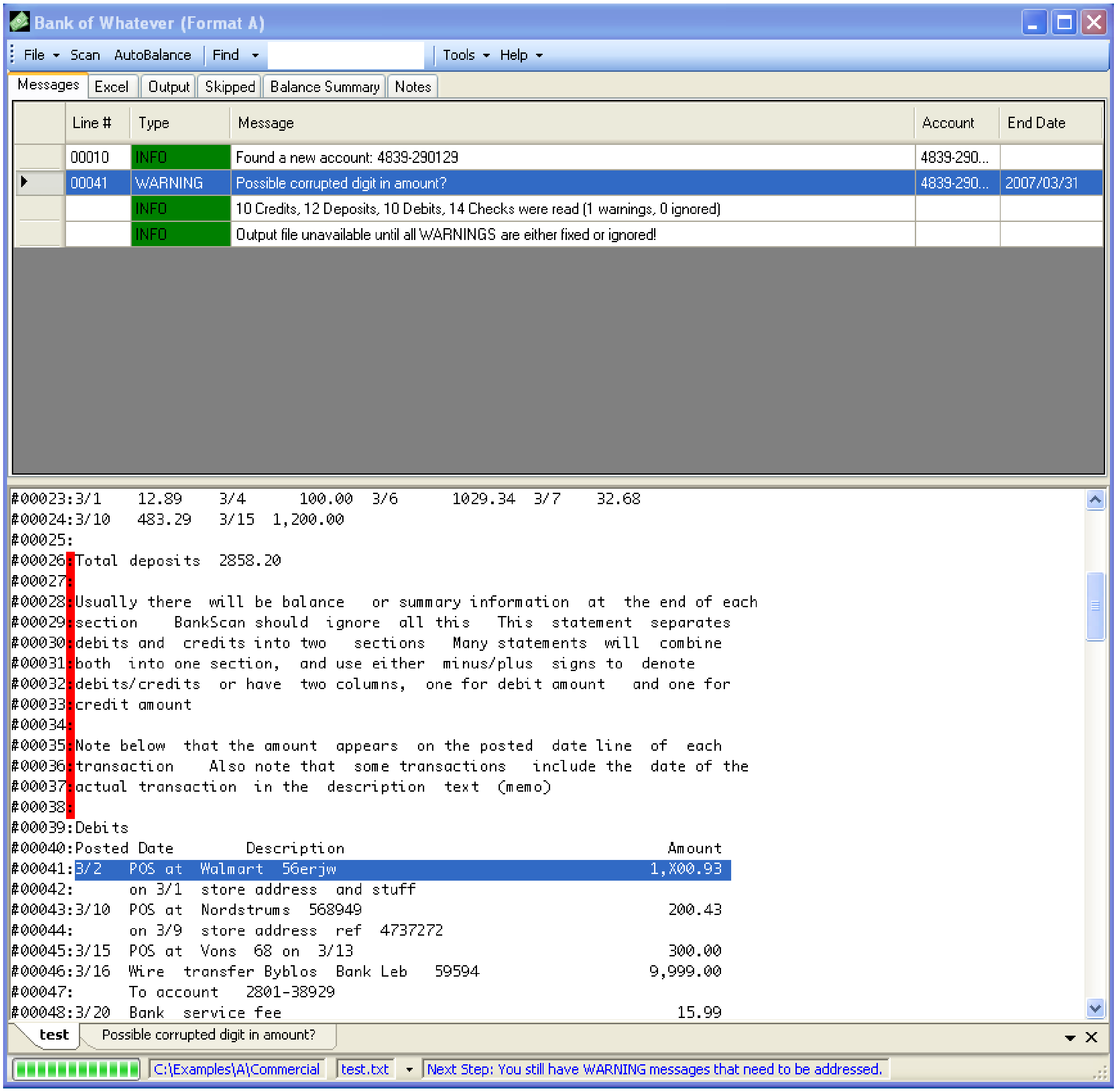
```
#00020:
#00021:Deposits
#00022:Date   Amount   Date     Amount  Date     Amount   Date    Amount
#00023:3/1    12.89    3/4      100.00  3/6      1029.34  3/7     32.6 8
#00024:3/10   483.29   3/15  1,200.00
#00025:
#00026:Total deposits  2858.20
#00027:
#00028:Usually there  will be balance   or summary information  at  the end of each
#00029:section    BankScan should  ignore  all this   This  statement  separates
#00030:debits and  credits into two   sections   Many statements  will  combine
#00031:both  into one section,  and use either  minus/plus  signs to  denote
#00032:debits/credits  or have  two columns,  one for debit amount   and one for
#00033:credit amount
#00034:
#00035:Note below  that the amount  appears  on the posted  date line  of  each
#00036:transaction    Also note that  some transactions   include the  date of the
#00037:actual transaction  in the  description  text  (memo)
#00038:
#00039:Debits
#00040:Posted Date        Description                                 Amount
#00041:3/2   POS at  Walmart  56erjw                                1,500.93
#00042:      on 3/1  store address  and stuff
#00043:3/10  POS at  Nordstrums  568949                               200.43
#00044:      on 3/9  store address  ref  4737272
#00045:3115  POS at  Vons  68 on  3/13                                300.00
#00046:3/16  Wire  transfer Byblos  Bank Leb   59594                9,999.00
```

Next Step: You still have WARNING messages that need to be addressed.

Figure 7. Messages tab

The number of warning messages the operator may have to clear depend on the accuracy of the OCR results. Poor quality statements (those with small or illegible fonts, low contrast, artifacts such as speckling, etc…) will require more corrections. For example, some fonts certain banks use make it very difficult to distinguish 6 from 8, which can cause balance checks to fail when an amount of $606.86 gets turned into $806.68!

In order to avoid the operator from having to flip through stacks of paper statements for making corrections, OmniPage can create a searchable PDF of statement images that BankScan can link to. For example, suppose in Figure 8 that the amount 1,X00.93 needs to be corrected for the bad digit X. The operator can quickly locate the correct amount by double-clicking on the warning message. This causes the PDF to be displayed and the area in question to be highlighted. Then, checking the image it can be seen what the digit X should actually be. This feature is most convenient when the operator has two monitors, one for the PDF display window, and the other for the BankScan program.



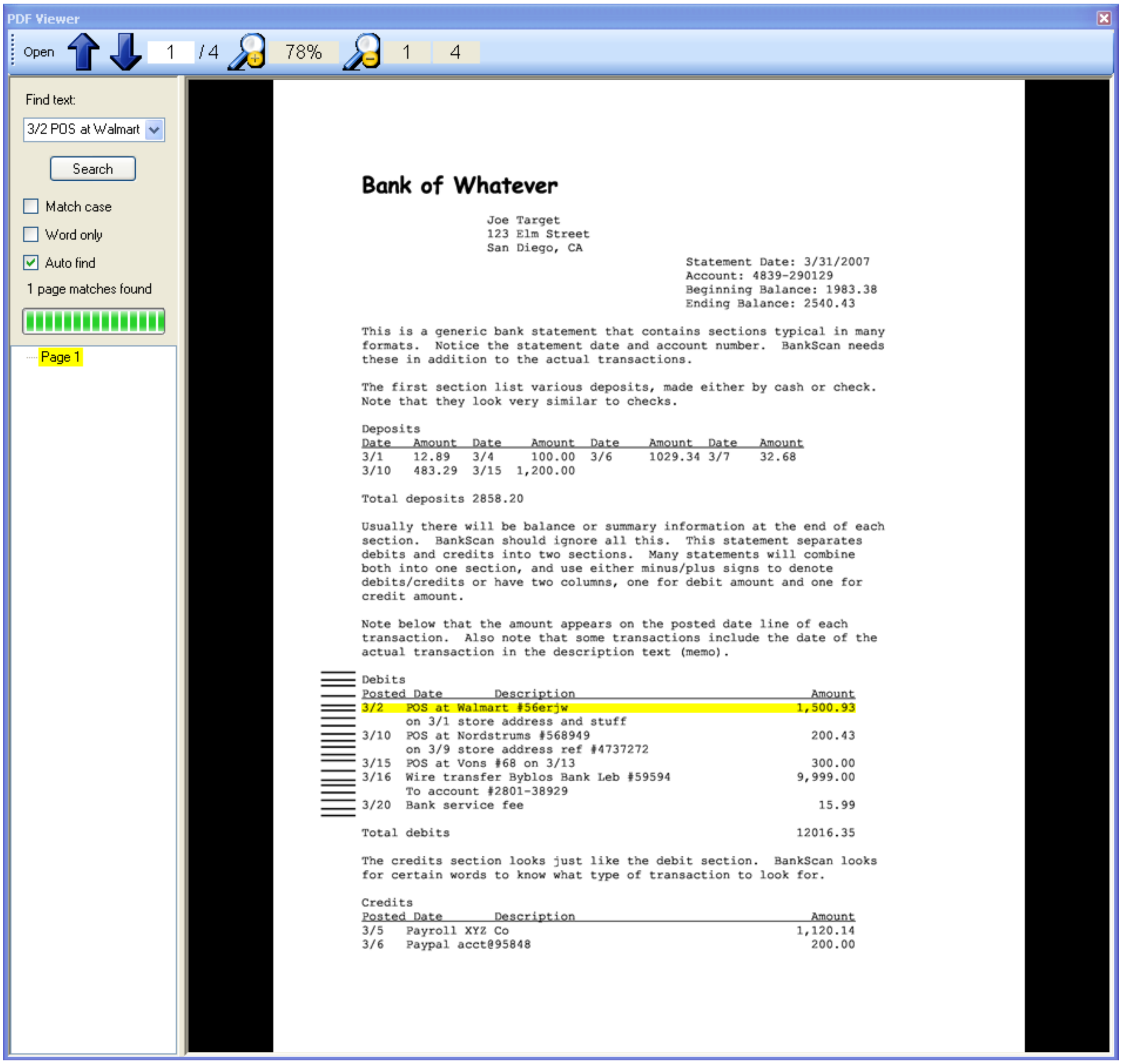
	Line #	Type	Message	Account	End Date
	00010	INFO	Found a new account: 4839-290129	4839-290...	
▶	00041	WARNING	Possible corrupted digit in amount?	4839-290...	2007/03/31
		INFO	10 Credits, 12 Deposits, 10 Debits, 14 Checks were read (1 warnings, 0 ignored)		
		INFO	Output file unavailable until all WARNINGS are either fixed or ignored!		

```
#00023:3/1    12.89    3/4      100.00  3/6       1029.34  3/7     32.68
#00024:3/10   483.29   3/15  1,200.00
#00025:
#00026:Total deposits  2858.20
#00027:
#00028:Usually there  will be balance   or summary information  at  the end of each
#00029:section    BankScan should  ignore  all this   This  statement separates
#00030:debits and  credits into two   sections   Many statements  will  combine
#00031:both  into one section,  and use either  minus/plus  signs to  denote
#00032:debits/credits  or have  two columns,  one for debit amount   and one for
#00033:credit amount
#00034:
#00035:Note below  that the amount  appears  on the posted  date line  of  each
#00036:transaction    Also note that  some transactions  include the  date of the
#00037:actual transaction  in the  description  text  (memo)
#00038:
#00039:Debits
#00040:Posted Date         Description                                   Amount
#00041:3/2   POS at  Walmart  56erjw                                  1,X00.93
#00042:      on 3/1  store address  and stuff
#00043:3/10  POS at  Nordstrums  568949                                 200.43
#00044:      on 3/9  store address  ref  4737272
#00045:3/15  POS at  Vons  68 on  3/13                                  300.00
#00046:3/16  Wire  transfer Byblos  Bank Leb   59594                  9,999.00
#00047:      To account   2801-38929
#00048:3/20  Bank  service fee                                           15.99
```

test   Possible corrupted digit in amount?

C:\Examples\A\Commercial   test.txt   Next Step: You still have WARNING messages that need to be addressed.

Figure 8. Warning message about a corrupted digit



Find text: 3/2 POS at Walmart

□ Match case

□ Word only

☑ Auto find

1 page matches found

Page 1

Bank of Whatever

Joe Target
123 Elm Street
San Diego, CA

Statement Date: 3/31/2007
Account: 4839-290129
Beginning Balance: 1983.38
Ending Balance: 2540.43

This is a generic bank statement that contains sections typical in many formats. Notice the statement date and account number. BankScan needs these in addition to the actual transactions.

The first section list various deposits, made either by cash or check. Note that they look very similar to checks.

Deposits

Date	Amount	Date	Amount	Date	Amount	Date	Amount
3/1	12.89	3/4	100.00	3/6	1029.34	3/7	32.68
3/10	483.29	3/15	1,200.00				

Total deposits 2858.20

Usually there will be balance or summary information at the end of each section. BankScan should ignore all this. This statement separates debits and credits into two sections. Many statements will combine both into one section, and use either minus/plus signs to denote debits/credits or have two columns, one for debit amount and one for credit amount.

Note below that the amount appears on the posted date line of each transaction. Also note that some transactions include the date of the actual transaction in the description text (memo).

Debits

Posted Date	Description	Amount
3/2	POS at Walmart #56erjw	1,500.93
	on 3/1 store address and stuff	
3/10	POS at Nordstrums #568949	200.43
	on 3/9 store address ref #4737272	
3/15	POS at Vons #68 on 3/13	300.00
3/16	Wire transfer Byblos Bank Leb #59594	9,999.00
	To account #2801-38929	
3/20	Bank service fee	15.99

Total debits 12016.35

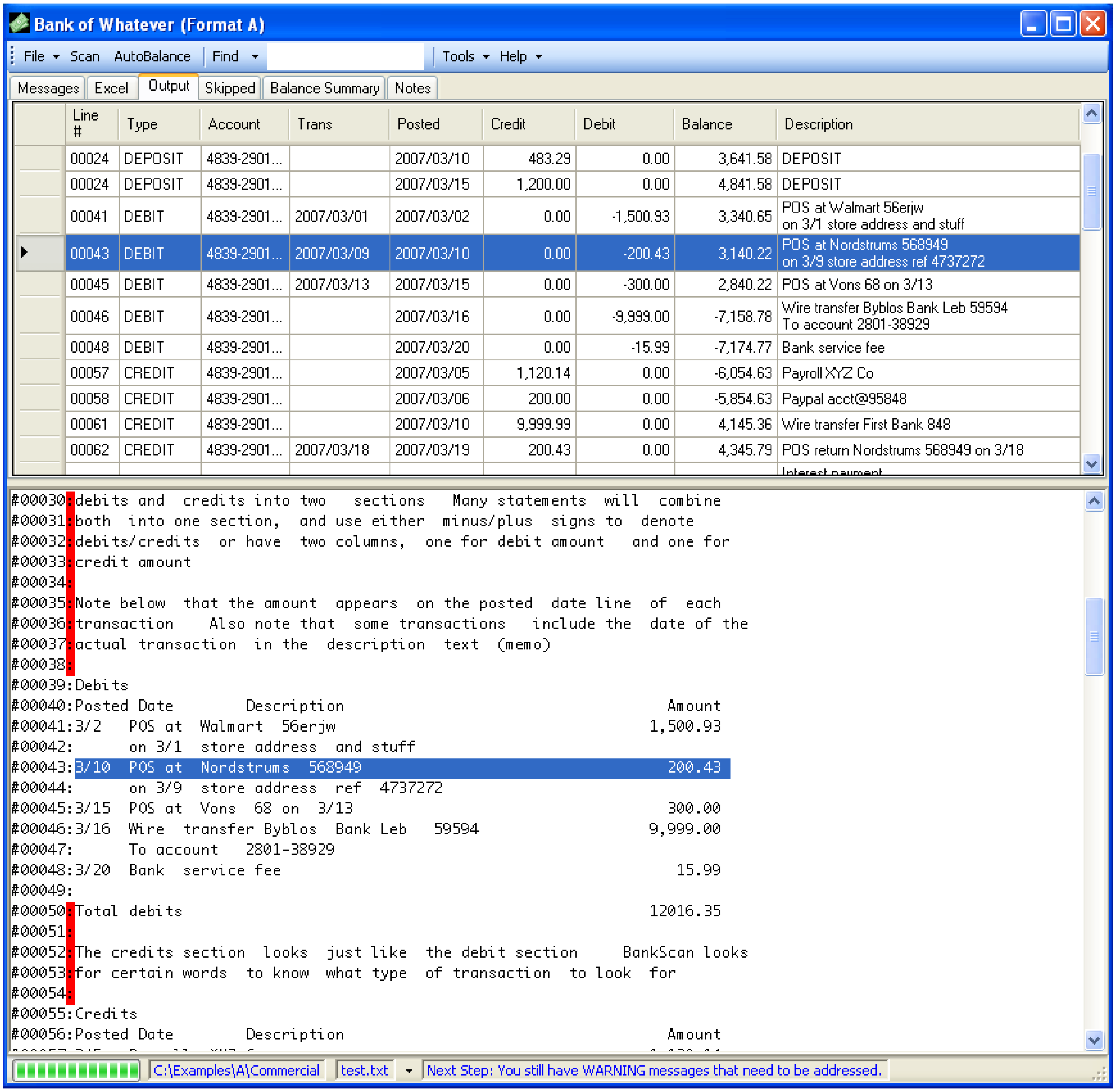
The credits section looks just like the debit section. BankScan looks for certain words to know what type of transaction to look for.

Credits

Posted Date	Description	Amount
3/5	Payroll XYZ Co	1,120.14
3/6	Paypal acct@95848	200.00

Figure 9. Locating a correction in the PDF of the statement

The other tabs in the top half of BankScan show the transactions that have been extracted, lines that have been skipped over, and the results of the AutoBalance function.

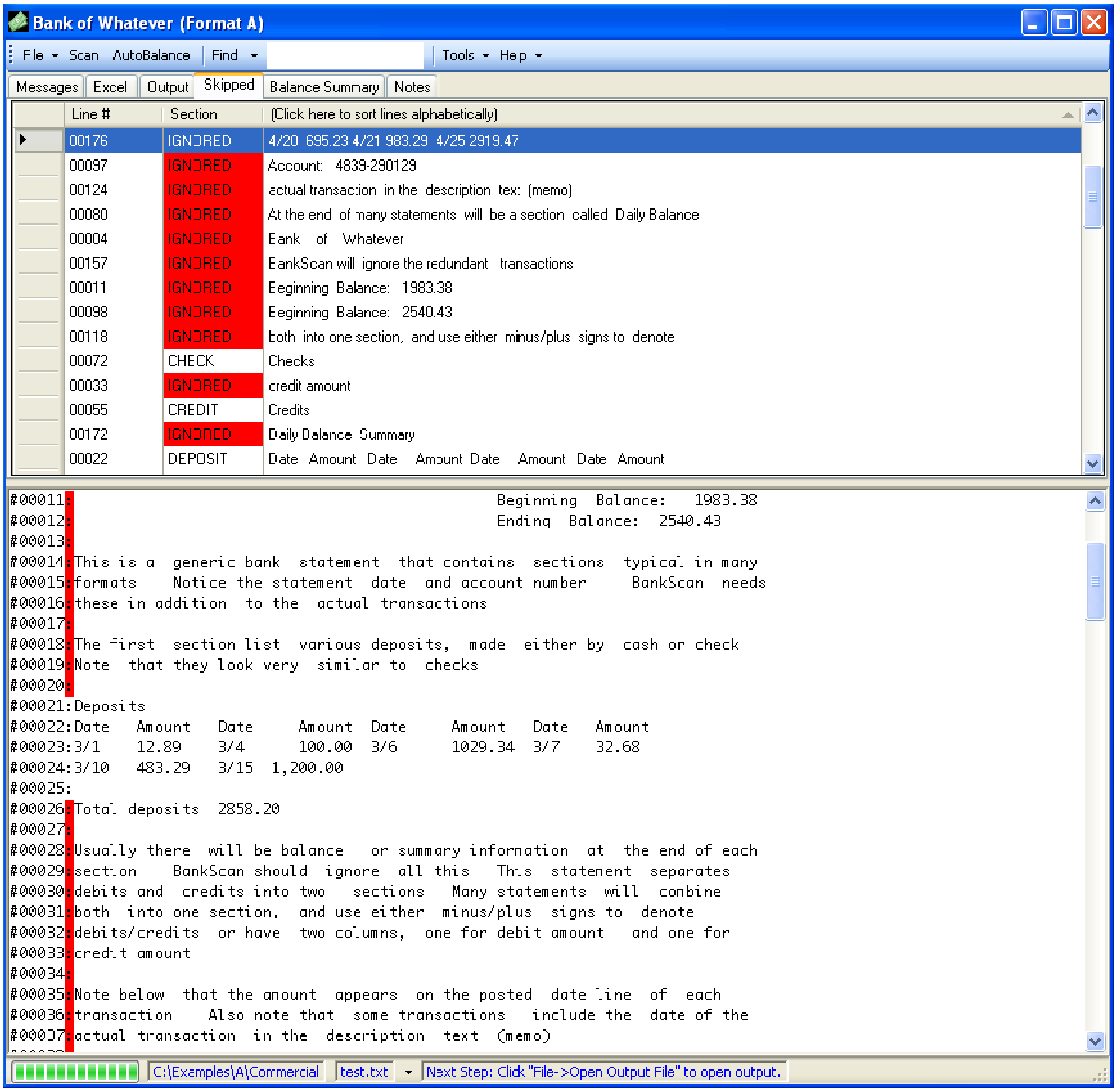


Line #	Type	Account	Trans	Posted	Credit	Debit	Balance	Description
00024	DEPOSIT	4839-2901...		2007/03/10	483.29	0.00	3,641.58	DEPOSIT
00024	DEPOSIT	4839-2901...		2007/03/15	1,200.00	0.00	4,841.58	DEPOSIT
00041	DEBIT	4839-2901...	2007/03/01	2007/03/02	0.00	-1,500.93	3,340.65	POS at Walmart 56erjw on 3/1 store address and stuff
00043	DEBIT	4839-2901...	2007/03/09	2007/03/10	0.00	-200.43	3,140.22	POS at Nordstrums 568949 on 3/9 store address ref 4737272
00045	DEBIT	4839-2901...	2007/03/13	2007/03/15	0.00	-300.00	2,840.22	POS at Vons 68 on 3/13
00046	DEBIT	4839-2901...		2007/03/16	0.00	-9,999.00	-7,158.78	Wire transfer Byblos Bank Leb 59594 To account 2801-38929
00048	DEBIT	4839-2901...		2007/03/20	0.00	-15.99	-7,174.77	Bank service fee
00057	CREDIT	4839-2901...		2007/03/05	1,120.14	0.00	-6,054.63	Payroll XYZ Co
00058	CREDIT	4839-2901...		2007/03/06	200.00	0.00	-5,854.63	Paypal acct@95848
00061	CREDIT	4839-2901...		2007/03/10	9,999.99	0.00	4,145.36	Wire transfer First Bank 848
00062	CREDIT	4839-2901...	2007/03/18	2007/03/19	200.43	0.00	4,345.79	POS return Nordstrums 568949 on 3/18

```
#00030:debits and  credits into two   sections   Many statements  will  combine
#00031:both  into one section,  and use either  minus/plus  signs to  denote
#00032:debits/credits  or have  two columns,  one for debit amount   and one for
#00033:credit amount
#00034:
#00035:Note below  that the amount  appears  on the posted  date line  of  each
#00036:transaction    Also note that  some transactions   include the  date of the
#00037:actual transaction  in the  description  text  (memo)
#00038:
#00039:Debits
#00040:Posted Date         Description                                  Amount
#00041:3/2   POS at  Walmart  56erjw                                  1,500.93
#00042:      on 3/1  store address  and stuff
#00043:3/10  POS at  Nordstrums  568949                                 200.43
#00044:      on 3/9  store address  ref  4737272
#00045:3/15  POS at  Vons  68 on  3/13                                  300.00
#00046:3/16  Wire  transfer Byblos  Bank Leb   59594                  9,999.00
#00047:      To account   2801-38929
#00048:3/20  Bank  service fee                                           15.99
#00049:
#00050:Total debits                                                   12016.35
#00051:
#00052:The credits section  looks  just like  the debit section     BankScan looks
#00053:for certain words  to know  what type  of transaction  to look  for
#00054:
#00055:Credits
#00056:Posted Date         Description                                  Amount
```

C:\Examples\A\Commercial | test.txt | Next Step: You still have WARNING messages that need to be addressed.

Figure 10. Output tab, transactions that have been found



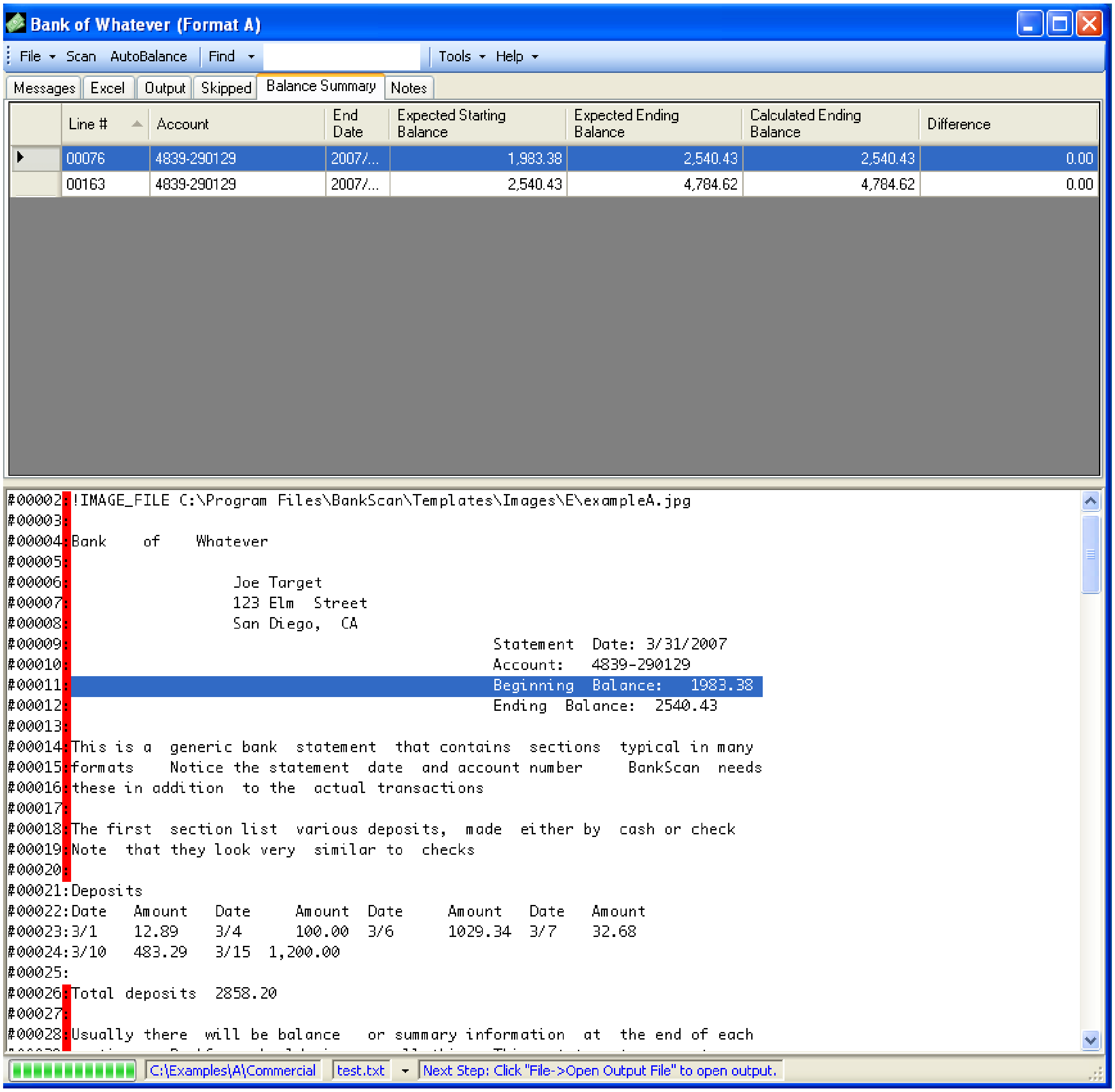
Line #	Section	(Click here to sort lines alphabetically)
00176	IGNORED	4/20 695.23 4/21 983.29 4/25 2919.47
00097	IGNORED	Account: 4839-290129
00124	IGNORED	actual transaction in the description text (memo)
00080	IGNORED	At the end of many statements will be a section called Daily Balance
00004	IGNORED	Bank of Whatever
00157	IGNORED	BankScan will ignore the redundant transactions
00011	IGNORED	Beginning Balance: 1983.38
00098	IGNORED	Beginning Balance: 2540.43
00118	IGNORED	both into one section, and use either minus/plus signs to denote
00072	CHECK	Checks
00033	IGNORED	credit amount
00055	CREDIT	Credits
00172	IGNORED	Daily Balance Summary
00022	DEPOSIT	Date Amount Date Amount Date Amount Date Amount

```
#00011:                                                Beginning  Balance:   1983.38
#00012:                                                Ending  Balance:  2540.43
#00013:
#00014:This is a  generic bank  statement  that contains  sections  typical in many
#00015:formats    Notice the statement  date  and account number     BankScan  needs
#00016:these in addition  to the  actual transactions
#00017:
#00018:The first  section list  various deposits,  made  either by  cash or check
#00019:Note  that they look very  similar to  checks
#00020:
#00021:Deposits
#00022:Date   Amount   Date     Amount  Date     Amount   Date   Amount
#00023:3/1    12.89    3/4      100.00  3/6      1029.34  3/7    32.68
#00024:3/10   483.29   3/15  1,200.00
#00025:
#00026:Total deposits  2858.20
#00027:
#00028:Usually there  will be balance   or summary information  at  the end of each
#00029:section    BankScan should  ignore  all this   This  statement  separates
#00030:debits and  credits into two    sections   Many statements  will  combine
#00031:both  into one section,  and use either  minus/plus  signs to  denote
#00032:debits/credits  or have  two columns,  one for debit amount   and one for
#00033:credit amount
#00034:
#00035:Note below  that the amount  appears  on the posted  date line  of  each
#00036:transaction    Also note that  some transactions  include the  date of the
#00037:actual transaction  in the  description  text  (memo)
```

C:\Examples\A\Commercial | test.txt | Next Step: Click "File->Open Output File" to open output.

Figure 11. Skipped tab, all of the “left over” lines

AutoBalancing compares calculated statement balances for each month and account number to the expected balances pulled from the statement summaries. It provides an audit check to help make sure that the data has been accurately extracted. All of these tabs are cross indexed to the editor window, making it easy to navigate around the file being processed. BankScan has a number of tools, like daily balance summary marking and balance column checking, to help the user more quickly locate the bad amount digits that throw off the AutoBalance.



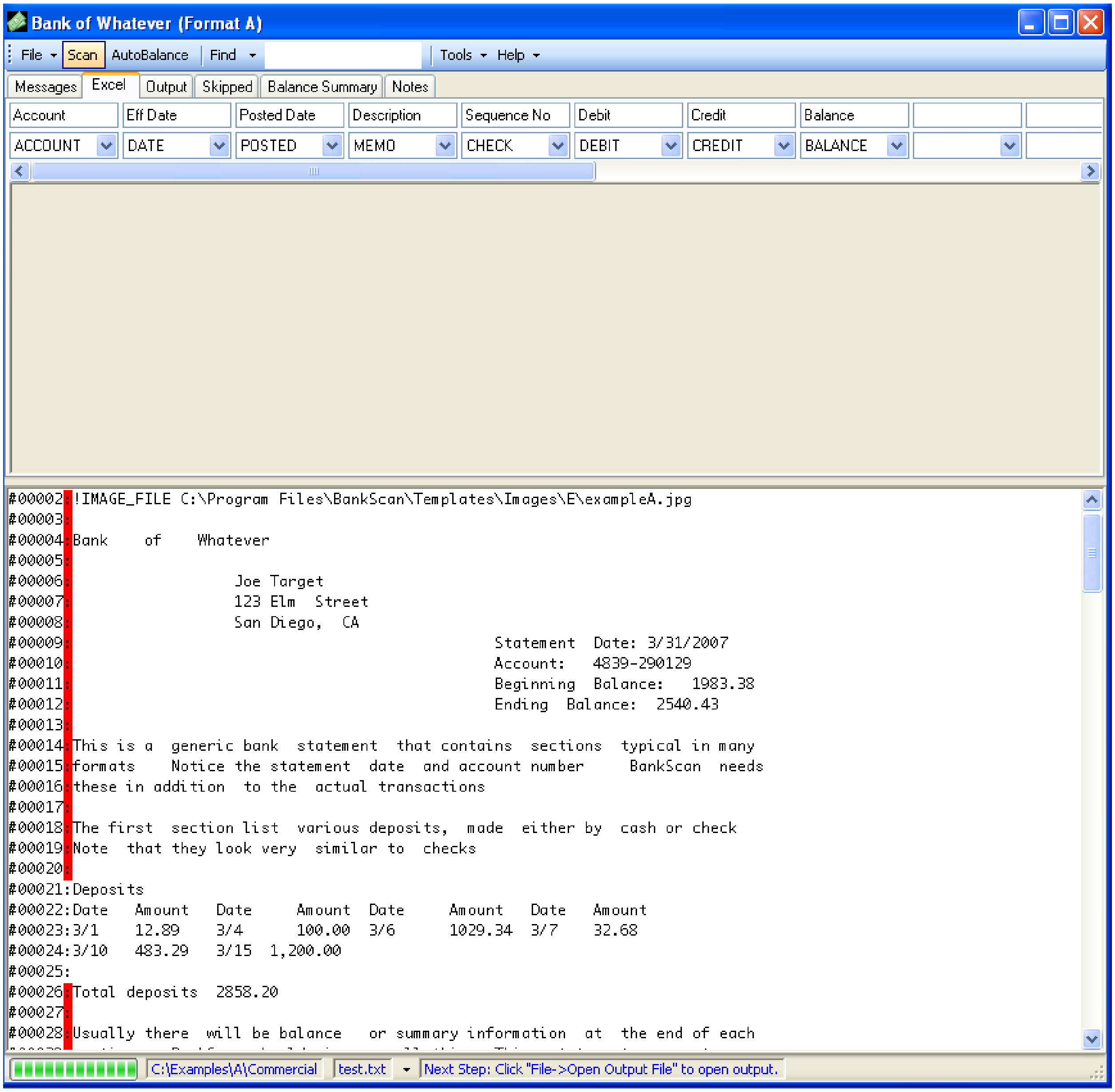
	Line #	Account	End Date	Expected Starting Balance	Expected Ending Balance	Calculated Ending Balance	Difference
▶	00076	4839-290129	2007/...	1,983.38	2,540.43	2,540.43	0.00
	00163	4839-290129	2007/...	2,540.43	4,784.62	4,784.62	0.00

```
#00002:!IMAGE_FILE C:\Program Files\BankScan\Templates\Images\E\exampleA.jpg
#00003:
#00004:Bank    of    Whatever
#00005:
#00006:                  Joe Target
#00007:                  123 Elm  Street
#00008:                  San Diego,  CA
#00009:                                            Statement  Date: 3/31/2007
#00010:                                            Account:   4839-290129
#00011:                                            Beginning  Balance:   1983.38
#00012:                                            Ending  Balance:  2540.43
#00013:
#00014:This is a  generic bank  statement  that contains  sections  typical in many
#00015:formats    Notice the statement  date  and account number     BankScan  needs
#00016:these in addition  to the  actual transactions
#00017:
#00018:The first  section list  various deposits,  made  either by  cash or check
#00019:Note  that they look very  similar to  checks
#00020:
#00021:Deposits
#00022:Date   Amount   Date     Amount  Date     Amount   Date   Amount
#00023:3/1    12.89    3/4      100.00  3/6      1029.34  3/7    32.68
#00024:3/10   483.29   3/15  1,200.00
#00025:
#00026:Total  deposits  2858.20
#00027:
#00028:Usually there  will be balance   or summary information  at  the end of each
```

Next Step: Click "File->Open Output File" to open output.

Figure 12. Balance summary tab

The Excel tab is used to select the desired data columns and their names and positions in the output spreadsheet.

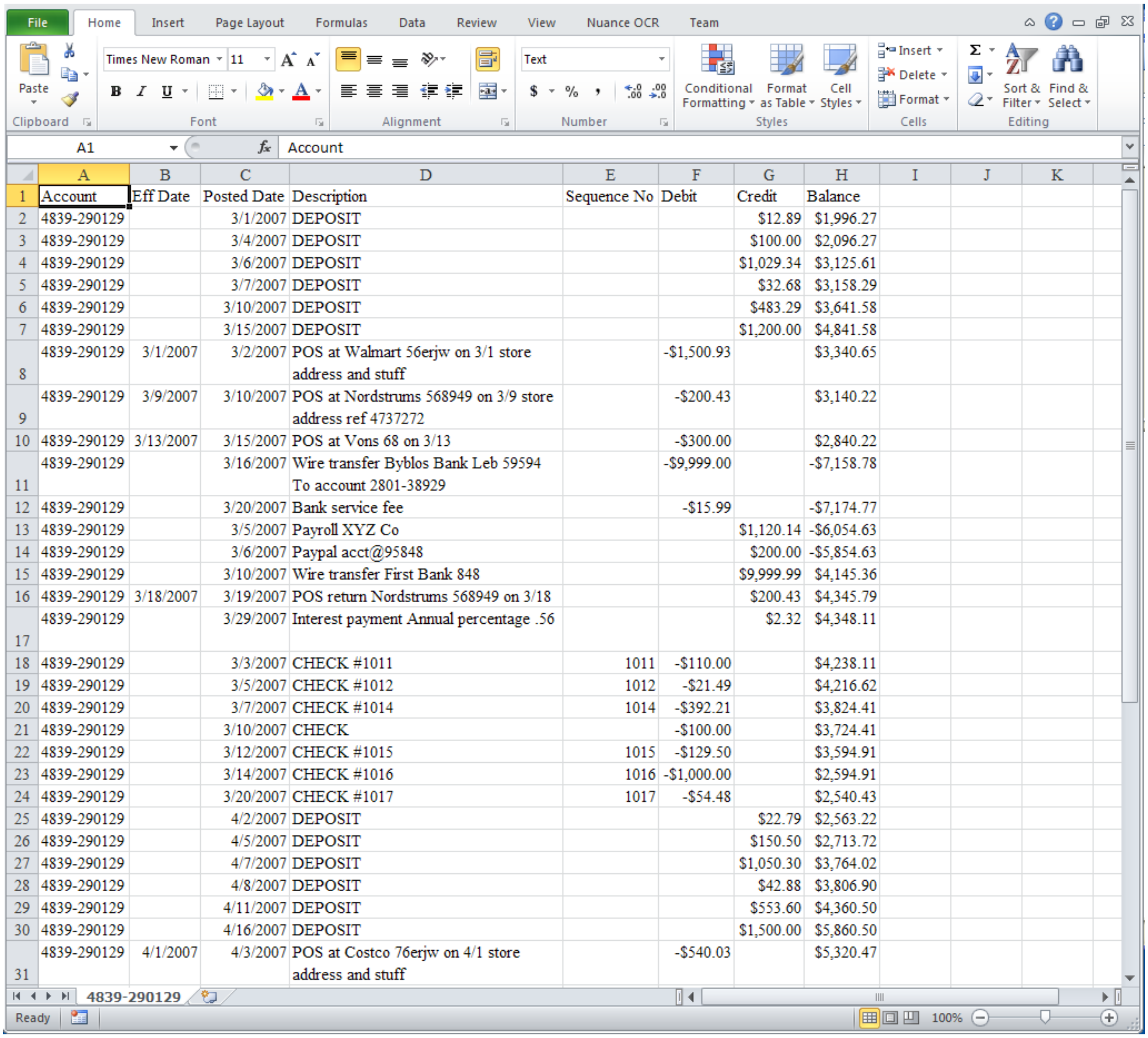


Account	Eff Date	Posted Date	Description	Sequence No	Debit	Credit	Balance		
ACCOUNT	DATE	POSTED	MEMO	CHECK	DEBIT	CREDIT	BALANCE		

```
#00002:!IMAGE_FILE C:\Program Files\BankScan\Templates\Images\E\exampleA.jpg
#00003:
#00004:Bank    of    Whatever
#00005:
#00006:                    Joe Target
#00007:                    123 Elm  Street
#00008:                    San Diego,  CA
#00009:                                                  Statement  Date: 3/31/2007
#00010:                                                  Account:   4839-290129
#00011:                                                  Beginning  Balance:   1983.38
#00012:                                                  Ending  Balance:  2540.43
#00013:
#00014:This is a  generic bank  statement  that contains  sections  typical in many
#00015:formats    Notice the statement  date  and account number     BankScan  needs
#00016:these in addition  to the  actual transactions
#00017:
#00018:The first  section list  various deposits,  made  either by  cash or check
#00019:Note  that they look very  similar to  checks
#00020:
#00021:Deposits
#00022:Date    Amount    Date     Amount  Date     Amount    Date    Amount
#00023:3/1     12.89     3/4      100.00  3/6      1029.34   3/7     32.68
#00024:3/10    483.29    3/15   1,200.00
#00025:
#00026:Total  deposits  2858.20
#00027:
#00028:Usually there  will be balance   or summary information  at  the end of each
```

Figure 13. Excel tab

After the operator has cleared any warnings and checked that calculated and expected balances match, BankScan writes the output to an .XML or .XLMS file that can be opened in Excel. Once the data has been imported into Excel it is now in the hands of the analyst. The job of BankScan is finished.



	A	B	C	D	E	F	G	H
1	Account	Eff Date	Posted Date	Description	Sequence No	Debit	Credit	Balance
2	4839-290129		3/1/2007	DEPOSIT			$12.89	$1,996.27
3	4839-290129		3/4/2007	DEPOSIT			$100.00	$2,096.27
4	4839-290129		3/6/2007	DEPOSIT			$1,029.34	$3,125.61
5	4839-290129		3/7/2007	DEPOSIT			$32.68	$3,158.29
6	4839-290129		3/10/2007	DEPOSIT			$483.29	$3,641.58
7	4839-290129		3/15/2007	DEPOSIT			$1,200.00	$4,841.58
8	4839-290129	3/1/2007	3/2/2007	POS at Walmart 56erjw on 3/1 store address and stuff		-$1,500.93		$3,340.65
9	4839-290129	3/9/2007	3/10/2007	POS at Nordstrums 568949 on 3/9 store address ref 4737272		-$200.43		$3,140.22
10	4839-290129	3/13/2007	3/15/2007	POS at Vons 68 on 3/13		-$300.00		$2,840.22
11	4839-290129		3/16/2007	Wire transfer Byblos Bank Leb 59594 To account 2801-38929		-$9,999.00		-$7,158.78
12	4839-290129		3/20/2007	Bank service fee		-$15.99		-$7,174.77
13	4839-290129		3/5/2007	Payroll XYZ Co			$1,120.14	-$6,054.63
14	4839-290129		3/6/2007	Paypal acct@95848			$200.00	-$5,854.63
15	4839-290129		3/10/2007	Wire transfer First Bank 848			$9,999.99	$4,145.36
16	4839-290129	3/18/2007	3/19/2007	POS return Nordstrums 568949 on 3/18			$200.43	$4,345.79
17	4839-290129		3/29/2007	Interest payment Annual percentage .56			$2.32	$4,348.11
18	4839-290129		3/3/2007	CHECK #1011	1011	-$110.00		$4,238.11
19	4839-290129		3/5/2007	CHECK #1012	1012	-$21.49		$4,216.62
20	4839-290129		3/7/2007	CHECK #1014	1014	-$392.21		$3,824.41
21	4839-290129		3/10/2007	CHECK		-$100.00		$3,724.41
22	4839-290129		3/12/2007	CHECK #1015	1015	-$129.50		$3,594.91
23	4839-290129		3/14/2007	CHECK #1016	1016	-$1,000.00		$2,594.91
24	4839-290129		3/20/2007	CHECK #1017	1017	-$54.48		$2,540.43
25	4839-290129		4/2/2007	DEPOSIT			$22.79	$2,563.22
26	4839-290129		4/5/2007	DEPOSIT			$150.50	$2,713.72
27	4839-290129		4/7/2007	DEPOSIT			$1,050.30	$3,764.02
28	4839-290129		4/8/2007	DEPOSIT			$42.88	$3,806.90
29	4839-290129		4/11/2007	DEPOSIT			$553.60	$4,360.50
30	4839-290129		4/16/2007	DEPOSIT			$1,500.00	$5,860.50
31	4839-290129	4/1/2007	4/3/2007	POS at Costco 76erjw on 4/1 store address and stuff		-$540.03		$5,320.47

Figure 14. Output in Excel spreadsheet

The BankScan Template Library

To date the BankScan template library contains over 3800 templates covering over 2500 financial institutions. This is by no means complete, new templates are constantly being added. When a new template is needed, a sufficient sample of statement data is provided as both a PDF image file and the recognized text file. The sample should cover at least a year and include all possible account types (checking, savings, loans, etc…) and transaction types (checks, deposits, electronic, etc…) Using this sample a new template is built. A simple statement template can take as little as 15 minutes to create.

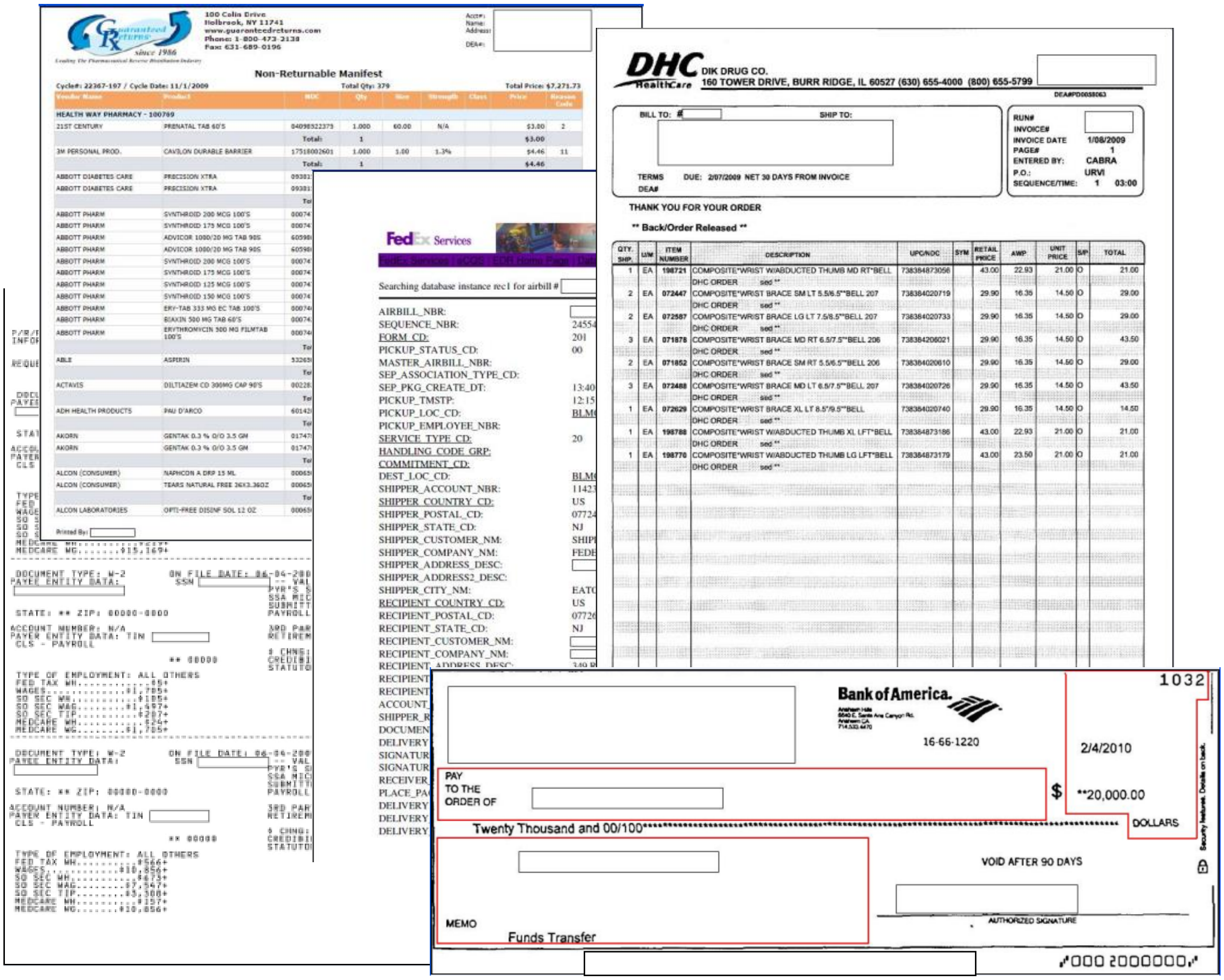
If a sample used to create a template does not contain a particular account type or transaction type, those types may be skipped over in subsequent statements that have them. The BankScan editor window contains tools for pulling in skipped transaction sections and inserting information such as account numbers, statement ending dates, and starting/ending balances as a temporary work around until the existing template can be updated with the new information.

Currently a BankScan licensee does not have the ability to create templates; it is done as a support service for the program. Not only are new templates added to the library, but updates to existing ones also occur on a regular basis. Keeping the program and library updated is done through a simple web based download. First BankScan downloads a signed list of template files along with a hash for each file. It compares these hashes with those calculated from its local templates. If they match then the files are up to date, if not the remote template is downloaded and its hash verified with that in the signed list. If it matches then the local file is replaced.

For installations running BankScan on several machines, a central library location can be defined so that only one library needs to be kept updated. For installations where an internet connection is not allowed for security reasons, an update file can be created on one internet connected machine and then installed on the isolated ones.

Extending BankScan – FileScan

BankScan is a specialized subset of a much more general built in tool called FileScan. FileScan uses templates to extract desired data fields from almost any type of document – shipping invoices, medical records, FedWire reports, etc… Figure 15 shows some of the types of documents that can have data fields pulled out of them. Because of the more generic nature of FileScan there is far less error checking involved.



Guaranteed Returns since 1986
100 Colin Drive
Holbrook, NY 11741
www.guaranteedreturns.com
Phone: 1-800-473-2138
Fax: 631-689-0196

Acct#:
Name:
Address:
DEA#:

Non-Returnable Manifest

Cycle#: 22367-197 / Cycle Date: 11/1/2009 Total Qty: 379 Total Price: $7,271.73

Vendor Name	Product	NDC	Qty	Size	Strength	Class	Price	Reason Code
HEALTH WAY PHARMACY - 100769								
21ST CENTURY	PRENATAL TAB 60'S	04098322373	1.000	60.00	N/A		$3.00	2
		Total:	1				$3.00	
3M PERSONAL PROD.	CAVILON DURABLE BARRIER	17518002601	1.000	1.00	1.3%		$4.46	11
		Total:	1				$4.46	

FedEx Services

Searching database instance rec1 for airbill #

AIRBILL_NBR:
SEQUENCE_NBR:
FORM_CD:
PICKUP_STATUS_CD:
MASTER_AIRBILL_NBR:
SEP_ASSOCIATION_TYPE_CD:
SEP_PKG_CREATE_DT:
PICKUP_TMSTP:
PICKUP_LOC_CD:
PICKUP_EMPLOYEE_NBR:
SERVICE_TYPE_CD:
HANDLING_CODE_GRP:
COMMITMENT_CD:
DEST_LOC_CD:
SHIPPER_ACCOUNT_NBR:
SHIPPER_COUNTRY_CD:
SHIPPER_POSTAL_CD:
SHIPPER_STATE_CD:
SHIPPER_CUSTOMER_NM:
SHIPPER_COMPANY_NM:
SHIPPER_ADDRESS_DESC:
SHIPPER_ADDRESS2_DESC:
SHIPPER_CITY_NM:
RECIPIENT_COUNTRY_CD: US
RECIPIENT_POSTAL_CD:
RECIPIENT_STATE_CD: NJ
RECIPIENT_CUSTOMER_NM:
RECIPIENT_COMPANY_NM:

DHC HealthCare
DIK DRUG CO.
160 TOWER DRIVE, BURR RIDGE, IL 60527 (630) 655-4000 (800) 655-5799

BILL TO: # SHIP TO:
TERMS DUE: 2/07/2009 NET 30 DAYS FROM INVOICE
DEA#

RUN#
INVOICE#
INVOICE DATE 1/08/2009
PAGE# 1
ENTERED BY: CABRA
P.O.: URVI
SEQUENCE/TIME: 1 03:00

THANK YOU FOR YOUR ORDER

** Back/Order Released **

QTY. SHP.	U/M	ITEM NUMBER	DESCRIPTION	UPC/NDC	SYM	RETAIL PRICE	AWP	UNIT PRICE	S/P	TOTAL
1	EA	198721	COMPOSITE*WRIST W/ABDUCTED THUMB MD RT*BELL DHC ORDER sed **	738384873056		43.00	22.93	21.00	O	21.00
2	EA	072447	COMPOSITE*WRIST BRACE SM LT 5.5/6.5"*BELL 207 DHC ORDER sed **	738384020719		29.90	16.35	14.50	O	29.00
2	EA	072587	COMPOSITE*WRIST BRACE LG LT 7.5/8.5"*BELL 207 DHC ORDER sed **	738384020733		29.90	16.35	14.50	O	29.00
3	EA	071878	COMPOSITE*WRIST BRACE MD RT 6.5/7.5"*BELL 206 DHC ORDER sed **	738384206021		29.90	16.35	14.50	O	43.50
2	EA	071852	COMPOSITE*WRIST BRACE SM RT 5.5/6.5"*BELL 206 DHC ORDER sed **	738384020610		29.90	16.35	14.50	O	29.00
3	EA	072488	COMPOSITE*WRIST BRACE MD LT 6.5/7.5"*BELL 207 DHC ORDER sed **	738384020726		29.90	16.35	14.50	O	43.50
1	EA	072629	COMPOSITE*WRIST BRACE XL LT 8.5"/9.5"*BELL DHC ORDER sed **	738384020740		29.90	16.35	14.50	O	14.50
1	EA	198788	COMPOSITE*WRIST W/ABDUCTED THUMB XL LFT*BELL DHC ORDER sed **	738384873186		43.00	22.93	21.00	O	21.00
1	EA	198770	COMPOSITE*WRIST W/ABDUCTED THUMB LG LFT*BELL DHC ORDER sed **	738384873179		43.00	23.50	21.00	O	21.00

Bank of America
16-66-1220
1032
2/4/2010
PAY TO THE ORDER OF $ **20,000.00
Twenty Thousand and 00/100** DOLLARS
VOID AFTER 90 DAYS
MEMO Funds Transfer
AUTHORIZED SIGNATURE
⑈000 2000000⑈

Figure 15. Sample of documents read by FileScan

To illustrate the usefulness of FileScan, consider that banks often provide images of printed checks that have been issued on an account. These images contain important items such as the payee, address, and “memo” line, which do not appear with the associated transactions in the bank statement (which will just show date, sequence number, and amount). These check images can be converted by OmniPage to text, and FileScan used to extract the additional data items. A special merge tool can also be used to match up and combine this data with the overall bank statement spreadsheet.