

The background of the slide features a man in a blue striped shirt looking thoughtfully to the side, with his hand on his chin. Overlaid on the image is a complex, hand-drawn diagram. The diagram includes various geometric shapes like cubes and circles, connected by arrows and lines. Some of the shapes contain mathematical symbols or formulas, such as $z = \frac{1}{x} + \frac{1}{y}$ and $z = \frac{1}{x} + \frac{1}{y} + \frac{1}{z}$. The diagram is set against a background of a light gray diamond pattern.

Первый проект– Извлечение по форматам и ключевым словам

Первый проект– Извлечение по форматам и ключевым словам

- ◆ Мы создали основной класс, называемый счетами и некоторыми подклассами
- ◆ Счета-фактуры могут быть отправлены несколькими тысячами поставщиков, и, возможно, вы не хотите иметь подкласс и локатор Advanced Zone Locator для каждого из этих счетов-фактур ...
- ◆ Предположим, что мы хотим настроить проект более общим образом и извлечь значения для всех этих документов таким же образом
- ◆ Помните, что мы не можем получить идеальные результаты

Первый проект– Извлечение по форматам и ключевым словам

- ◆ Чтобы выполнить извлечение более «общим» способом (во многих разных формах), вам необходимо знать:
- ◆ Какие значения вы хотите извлечь?
- ◆ Какие ключевые слова необходимы?
- ◆ Существует ли связь между значениями?
- ◆ Вы уже знаете значения? (Можете ли вы использовать базу данных?)
- ◆ Мы сделали несколько заметок ранее ...
- ◆ Давайте рассмотрим наши текущие знания документов

Первый проект– Извлечение по форматам и ключевым словам

PO Number
всегда содержит
8 цифр
Ключевое слово:
Purchase Order

Purchase Order: 20026306

Invoice Number
всегда содержит
9 цифр
Ключевое слово
Invoice Number

Invoice Number: 196328182

Invoice Date
всегда в
формате:
MM/DD/YY

Ключевое слово:
Invoice Date

Invoice Date: 12/26/02

Общая сумма
всегда в
формате **US**
Ключевое слово:
Invoice Total

Invoice Total \$ 2,038.66

This is your **INVOICE** *Bo* Page: 1 of 1

FID Number: 74-2616905 Sales Rep: JENNIFER ROCHNER
Customer Number: 000318193 Purchase Order: 20026306
Order Number: 196328182 Order Date: 12/23/02
Invoice Date: 12/26/02
Payment Terms: NET DUE 30 DAYS
Shipped Via: STANDARD GROUND
Waybill Number: V5410005034218

SOLD TO:
JIM McDONALD
KOFAX IMAGE PRODUCTS
PO 20026306
16245 LAGUNA CANYON RD
IRVINE, CA 92618-3903

SHIP TO:
JIM McDONALD
KOFAX IMAGE PRODUCTS
PO 20026306
16245 LAGUNA CANYON RD
IRVINE, CA 92618-3903

PLEASE REVIEW IMPORTANT TERMS & CONDITIONS ON THE REVERSE SIDE OF THIS INVOICE

Qty	Shipped	Item Number	Description	Unit	Unit Price	Amount
1		221-0440	Latitude C840, 2.0GHz to Pentium 4 14.1 XGA, NIC, English	EA	1,857.00	1,857.00
1		311-1369	256MB, Double Data Rate SDRAM, 1 Dimm, for Dell Latitude C840/	EA	0.00	0.00
1		340-3017	C840 Notebooks, Factory Install			
1		340-3017	30GB Hard Drive, 9.5MM for Dell Latitude C800/C500/C610/C510/	EA	0.00	0.00
1		420-0656	C840 Notebooks, Factory Install			
1		420-0656	Windows XP Professional, English for Dell Latitude C840	EA	0.00	0.00
1		313-5205	Notebooks, Factory Install			
1		313-5205	Internal 56K Modem for Dell Latitude C810/C510/C840	EA	0.00	0.00
1		313-1174	Notebooks, Factory Install			
1		313-1174	8-24-10-24X DVD w/Software CD Read-Write Combo for Dell	EA	0.00	0.00
1		312-0042	Latitude C810/C810/C400/C510/C840 Notebooks, Factory Inst	EA	0.00	0.00
1		312-0042	8-Cell Primary Battery for Dell Latitude C840 Notebooks,			
1		420-0153	Factory Install			
1		900-6550	Windows NT/2000/XP, NTFS File System, Factory Install	EA	0.00	0.00
1		900-6550	*Type 3 Contract - Next Business Day Parts and Labor	EA	0.00	0.00
1		900-6550	On-Site Response, Initial Year			
1		900-6550	*Type 3 Contract - Next Business Day Parts and Labor			
1		900-6550	On-Site Response, 2YR Extended	EA	0.00	0.00
1		900-6550	On-Site Response, 2YR Extended			
			System Service Tag: B01M721			

RECEIVED
DEC 30 2002

36738-00 \$ 1857.00
01-1540-500 \$ 181.66
62038-66

ENTERED JAN 9 2003

DELL
DETACH AT PERP AND RETURN WITH PAYMENT

Invoice Number: 196328182
Purchase Order: 20026306
Order Number: 196328182

MAKE CHECK PAYABLE REMIT TO:
DELL MARKETING L.P.
CO DELL USA L.P.
DEPT. LA21205
PASADENA, CA 91366

Ship. &/or Handling	\$	Tax	\$
Subtotal	1,857.00		
Taxable	1,857.00	140.66	
Invoice Total			2,038.66

Balance Due \$ 2,038.66
Amt. Enclosed \$

000196328182000000002038660500003181935

Первый проект– Извлечение по форматам и ключевым словам

Invoice Number	Keyword	Invoice Date	Keyword	PONumber	Keyword	Total amount	Keyword
\d{9}	Invoice Number	US DATE	Invoice Date	\d{8}	Purchase Order	US Amount	Invoice Total

Notes:

Первый проект– Извлечение по форматам и ключевым словам

PONumber
иногда пустое, 6,
8 or 12 цифр или
типа A55611

Ключевое слово
P.O. NUMBER

P.O. NUMBER

Общая сумма
всегда в
формате US

Ключевое слово:
Please pay this
amount

PLEASE PAY THIS AMOUNT
62 76

ORIGINAL #002591 INVOICE PAGE 01 OF 01

FedEx Freight (FXFE)

Freight Bill Number: 632 325 439

DATE: 07/22/04

ORIGIN: AVL DEST: TW

TRACKING NUMBER: 846167

SHIPPER: 21344610
COATS NORTH AMERICA
P O BOX 1208
MAULDIN NC 28752

BILL TO: COATS AMERICAN
% TRENDSET INC
P O BOX 1208
MAULDIN SC 29662

CONSIGNEE: 58413913
FORT WORTH TX 7614

PO NUMBER: 58413913

BL NUMBER: 213

PAY: \$ 62.76

ROUTING: 213

RECEIVED: \$ 62.76

KEY: \$ 62.76

DESCRIPTION: 18 CARTON, THREAD, SEWING,
CON ON
CONES, CORES OR SPOOLS
11 SKIDS
11 SMS STC 10 CTNS
458018
BL
000476 FUEL SURCHG LTL SHPT8 207
DISCOUNT ARFW 6000 0129433 0014
REFXF 1000 06/14/04 LS 31277
** FAK RATES APPLIED **

WT(LBS): 213

N M F C: 182220-01

CLASS: 060

RATE: FLAT

TOTAL CHARGES: \$ 58.00

TERMS: PPD

213

\$ 62.76

Reliable, Responsive Regional and Interregional LTL

Remittance Advice
Please Return This Portion With Your Payment
Terms Net 15 Days

PAYING PARTY: 84616794
COATS AMERICAN
% TRENDSET INC
P O BOX 1208
MAULDIN

Freight Bill Number: 632 325 439

DATE: 07/22/04

PLEASE PAY THIS AMOUNT: 62 76

TRACKING NUMBER: 846167

FedEx Freight (FXFE)

Send Payment to: FedEx Freight East, P O Box 406708, Atlanta GA 30384-8708
P O Box 840, Harrison AR 72602-0840
Phone 1 870 741 9000 Fax 1 870 365 4354 Toll-Free 1 866 393 4585

THANK YOU "

Invoice
Number всегда 9
цифр

Ключевое
слово Freight Bill
Number

FREIGHT BILL NUMBER
632 325 439

Invoice Date
всегда в
формате
MM/DD/YY

Ключевое слово:
Date

Можно найти под
invoice number

DATE
07/22/04

Первый проект– Извлечение по форматам и ключевым словам

Invoice Number	Keyword	Invoice Date	Keyword	PONumber	Keyword	Total amount	Keyword
\d{9}	Invoice Number	US DATE	Invoice Date	\d{8}	Purchase Order	US Amount	Invoice Total
\d{9}	Freight Bill Number	US DATE	Date	\d{6} \d{8} \d{12} A\d{4}I	P.O. Number	US Amount	Please pay this amount

Notes:

Fedex: Poor Keyword for Invoice Date, but it is below Invoice number.
 Keywords ORIGIN and DEST are always below Invoice Date.
 PO Number can be empty

Первый проект– Извлечение по форматам и ключевым словам

Invoice Date is
always
MM/DD/YY
Keyword is
Invoice Date

INVOICE DATE
01/29/03

PO Number is 8
digits

Keyword is
Customer PO NO

CUSTOMER P.O. NO.
20020280

785 NORTH MARY AVENUE • SUNNYVALE, CA 94085-2909 • PHONE (408) 720-1900

INVOICE

3685
PO#

INVOICE NO. 307243 RI

INVOICE DATE 01/29/03

SHIP TO: 202180 KOFAX IMAGE PRODUCTS, INC. 16245 LAGUNA CANYON ROAD IRVINE CA 92618-3603

SHIP TO: 202180 KOFAX IMAGE PRODUCTS, INC. 16245 LAGUNA CANYON ROAD IRVINE CA 92618-3603

CUSTOMER P.O. NO. 20020280

TERMS NET THIRTY DAYS

F.O.B. FOB-SHIPPING PO

REF NO. 1031

REP NAME KOFAX IMAGE PRODUCTS

REFERENCE (SPSSL NO.)

ITEM NO.	PART NUMBER	CUSTOMER PART NO.	U/M	QUANTITY	UNIT PRICE	EXTENSION
1	MSM99Q029-001LS	Tony	EA	1200	17.9500	21,540.00 USD

RECEIVED
FEB 03 2003

ENTERED FEB - 6 2003

Please Remit To. OKI SEMI-SF LOCKBOX
P.O. BOX 7060
SAN FRANCISCO
CA 94120
CUSTOMER

SUB TOTAL 21,540.00 USD

SALES TAX 3683.36 USD

AMOUNT DUE 21,540.00 USD

AMOUNT DUE 21,540.00 USD

Invoice Number is
always
6 digits
and 2 chars
Keyword is
Invoice Number

INVOICE NO.
307243 RI

Total amount
is always in
US Format
Keyword is
AMOUNT
DUE

AMOUNT DUE 21,540.00 USD

KOFAX

Первый проект– Извлечение по форматам и ключевым словам

Invoice Number	Keyword	Invoice Date	Keyword	PONumber	Keyword	Total amount	Keyword
\d{9}	Invoice Number	US DATE	Invoice Date	\d{8}	Purchase Order	US Amount	Invoice Total
\d{9}	Freight Bill Number	US DATE	Date	\d{6} \d{8} \d{12} A\d{4}l	P.O. Number	US Amount	Please pay this amount
\d{6}[A-Z]{2}	Invoice NO	US DATE	Invoice Date	\d{8}	Customer PO NO	US Amount	Amount Due

Notes:

Fedex: Poor Keyword for Invoice Date, but it is below Invoice number
PO Number can be empty

Oki: Poor Keyword for Total Amount, because it has white text on a black background
Total Amount is always below Sales Taxes

Первый проект– Извлечение по форматам и ключевым словам

- ◆ Основываясь на этих знаниях, мы можем создать некоторые «общие» локаторы для нашего проекта
- ◆ Поскольку это общие знания, мы хотим создать их, привязанные к базовому классу «Счета-фактуры»
- ◆ Каждый подкласс ниже Invoices наследует локаторы и поля из этого базового класса
- ◆ Иногда базовые классы называются родительскими классами
- ◆ Иногда подклассы называются дочерними классами

Первый проект– Извлечение по форматам и ключевым словам

1) Добавьте новый локатор в свою систему и назовите FL_InvoiceNumber

Filename	Classification Result	Confidence	Assigned Class
FedEx3.xdc*	FedEx	99.8%	FedEx
FedEx30.xdc	FedEx	83.5%	
FedEx29.xdc	FedEx	84.1%	
FedEx27.xdc	FedEx	84.0%	
FedEx28.xdc	FedEx	83.8%	
FedEx34.xdc	FedEx	84.3%	
FedEx35.xdc	FedEx	84.1%	
FedEx33.xdc	FedEx	83.9%	
FedEx31.xdc	FedEx	84.1%	
FedEx32.xdc	FedEx	84.3%	
FedEx20.xdc	FedEx	82.9%	
FedEx21.xdc	FedEx	81.9%	
FedEx22.xdc	FedEx	84.0%	

Первый проект– Извлечение по форматам и ключевым словам

3) Клик на свойства локатора

2) Используйте Format Locator как Locator Method

Details: Locator "FL_InvoiceNumber" in Class "Invoices"

Settings

Definition Class: Invoices

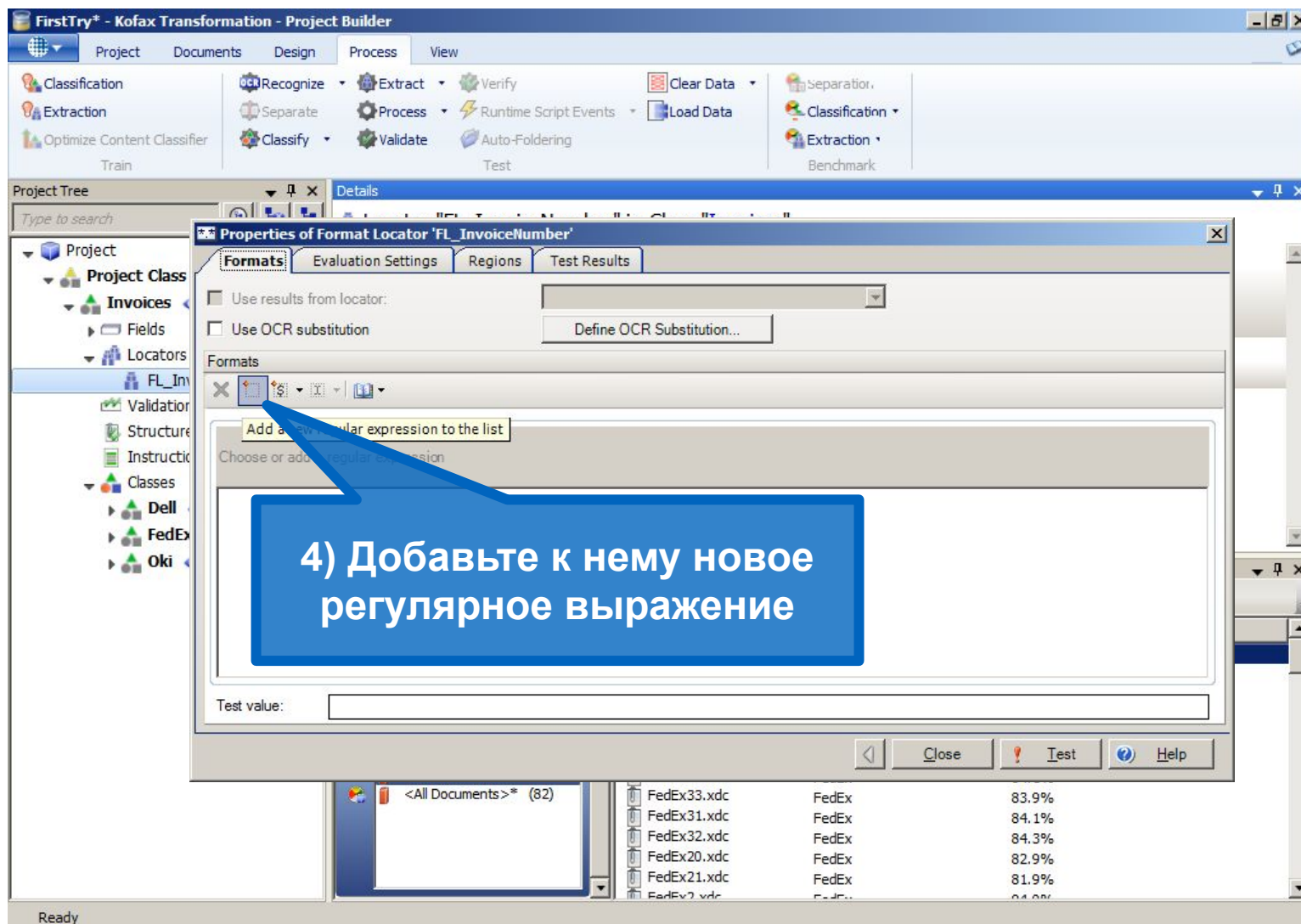
Locator Method: <none> [Copy Method]

The locator is as <none>

Documents - Benchmark Set *

File Name	Classification Result	Confidence	Assigned Class
FedEx30.xdc	FedEx	99.8%	FedEx
FedEx29.xdc	FedEx	83.5%	FedEx
FedEx27.xdc	FedEx	84.1%	FedEx
FedEx28.xdc	FedEx	84.0%	FedEx
FedEx34.xdc	FedEx	83.8%	FedEx
FedEx35.xdc	FedEx	84.3%	FedEx
FedEx33.xdc	FedEx	84.1%	FedEx
FedEx31.xdc	FedEx	83.9%	FedEx
FedEx32.xdc	FedEx	84.1%	FedEx
FedEx20.xdc	FedEx	84.3%	FedEx
FedEx21.xdc	FedEx	82.9%	FedEx
FedEx22.xdc	FedEx	81.9%	FedEx
FedEx23.xdc	FedEx	84.0%	FedEx

Первый проект– Извлечение по форматам и ключевым словам



Первый проект– Извлечение по форматам и ключевым словам

5) Переключиться на Evaluation Settings

4) Введите регулярное выражение

Format	Description	Ignore Characters
<input checked="" type="checkbox"/>	\d{9}	<Description>

Test value:

Первый проект– Извлечение по форматам и ключевым словам

6) Введите ключевое слово

8) Добавьте в список

7) Выберите “Match all words as a phrase” если у вас есть пробел между вашими ключевыми словами, и если вы ожидаете, что они будут одной строкой

Первый проект– Извлечение по форматам и ключевым словам

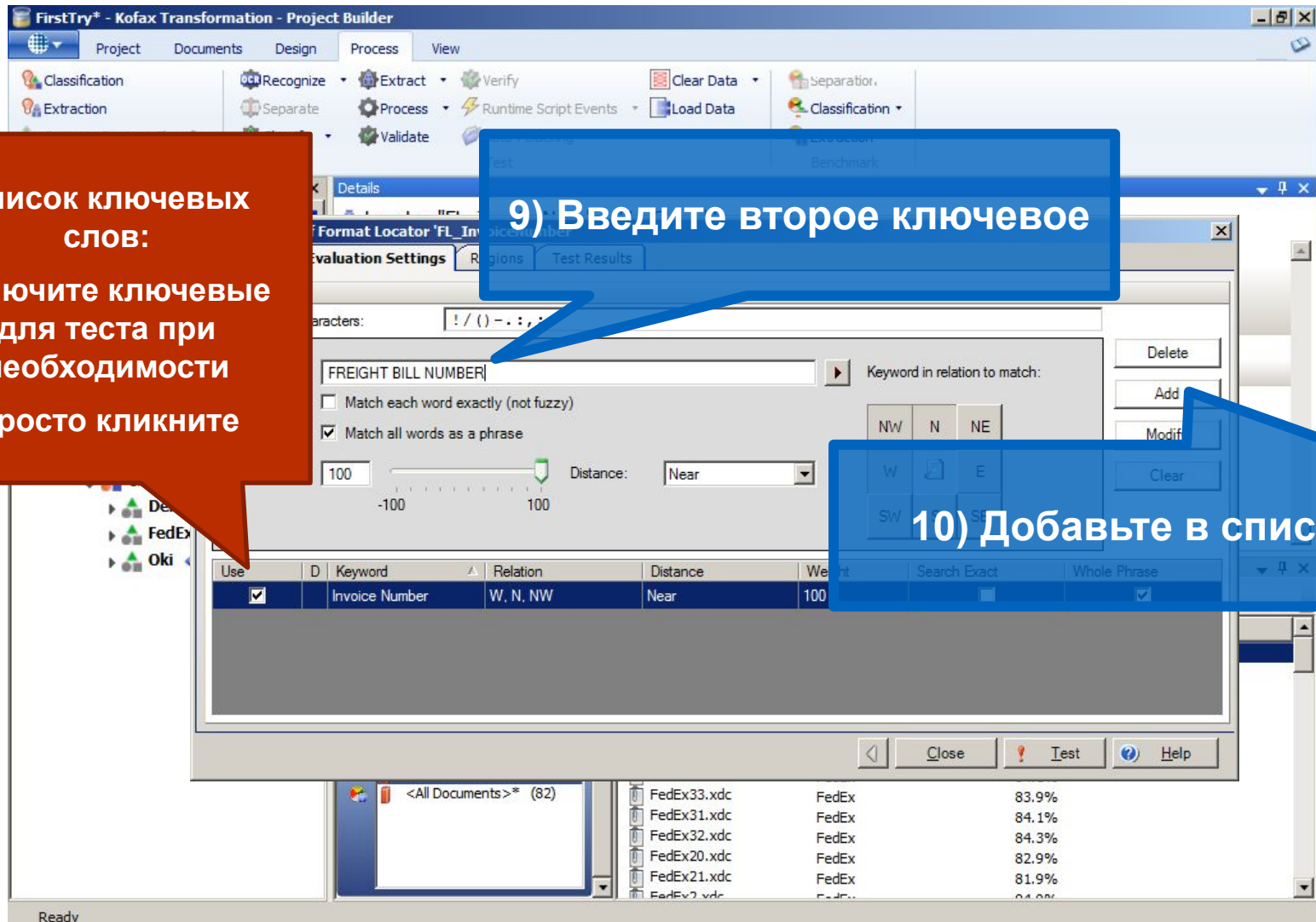
Список ключевых слов:

Включите ключевые для теста при необходимости

Просто кликните

9) Введите второе ключевое

10) Добавьте в список



Первый проект– Извлечение по форматам и ключевым словам

Теперь все ключевые в списке

11) Проверьте свои настройки на документе, который выбран в тестовых документах

Use	D	Keyword	Relation	Distance	Weight	Search Back	Find Phrase
<input checked="" type="checkbox"/>		FREIGHT BILL NUM...	W, N, NW	Near			
<input checked="" type="checkbox"/>		INVOICE NO	W, N, NW	Near	100		
<input checked="" type="checkbox"/>		Invoice Number	W, N, NW	Near	100		<input checked="" type="checkbox"/>

Document	Format	Score
FedEx33.xdc	FedEx	83.9%
FedEx31.xdc	FedEx	84.1%
FedEx32.xdc	FedEx	84.3%
FedEx20.xdc	FedEx	82.9%
FedEx21.xdc	FedEx	81.9%
FedEx22.xdc	FedEx	84.0%

Первый проект– Извлечение по форматам и ключевым словам

The screenshot displays the Kofax Transformation Project Builder interface. The main window shows a document viewer with a FedEx invoice. The 'Classification result' is 'FedEx'. The 'Detected Alternatives' table is visible at the bottom left.

Results of the test in the viewer

Results of the test in the locator properties

Confidence	Text
100.00 %	632325413
100.00 %	632325413
0.00 %	187074190
0.00 %	187036543
0.00 %	186639345
0.00 %	600001294

12) Выберите другой документ из тестового набора и повторите попытку

Первый проект– Извлечение по форматам и ключевым словам

FirstTry+ - Project Builder

Project Documents Design Process View

Classification Extraction Optimize Content Classifier Train

Recognize Separate Classify

Extract Process Validate

Verify Runtime Script Events Auto-Foldering Test

Clear Data Load Data

Project Tree

Type to search

- Project
 - Project Class
 - Invoices
 - Fields
 - Locators
 - FL_InvoiceNumber

Details

Locator "FL_InvoiceNumber" in Class "Invoices"

Settings

Definition Class: Invoices

Locator Method: Format Locator

The locator is assigned to the following fields:

<none>

Document Viewer

Classification result: Dell

Dell

Invoice Number (100%)

100.00%

100.00%

Properties of Format Locator 'FL_InvoiceNumber'

Formats Evaluation Settings Regions

Detected Alternatives

	Confidence	Text
1	100.00 %	242825868
2	96.95 %	333323807
3	96.95 %	242825868
4	0.00 %	000318193
5	0.00 %	333323807
6	0.00 %	333323807
7	0.00 %	000318193
8	0.00 %	333323807
9	0.00 %	000318193
10	0.00 %	333323807

Слишком много результатов с высокой степенью уверенности в документах Dell
Давайте посмотрим ближе

800dpi 44 / 82 2 / 2 TIFF

Первый проект– Извлечение по форматам и ключевым словам

The screenshot displays the Kofax Transformation Project Builder interface. A central window shows extracted invoice data:

Invoice Number: 333323807
Invoice Number (100%): 333323807
Invoice Date: 04/27/03
Payment Terms: NET DUE 30 DAYS
Shipped Via: 2ND DAY
Waybill Number: 24282586860
100.00%

Below this, it states: "The locator is assigned to the following fields: <none>".

A separate window titled "Properties of Format Locator 'FL_InvoiceNumber'" shows a table of detected alternatives:

	Confidence	Text
1	100.00 %	242825868
2	96.95 %	333323807
3	96.95 %	242825868
4	0.00 %	000318193
5	0.00 %	333323807
6	0.00 %	333323807
7	0.00 %	000318193
8	0.00 %	333323807
9	0.00 %	000318193
10	0.00 %	333323807

A red callout box points to the first row of the table, stating: "Первый результат исходит со страницы 2, и это номер накладной" (The first result comes from page 2, and this is the invoice number).

The background shows a document viewer with a Dell invoice. A blue arrow points from the "Invoice Number (100%)" field to the "Invoice Number" field in the document viewer, which also shows "100.00%".

Первый проект– Извлечение по форматам и ключевым словам

The screenshot displays the Kofax Transformation Project Builder interface. The main window shows a document with extracted fields: Invoice Number (333323807), Invoice Date (04/27/03), Payment Terms (NET DUE 30 DAYS), Shipped Via (2ND DAY), and Waybill Number (2428258660). A blue callout box highlights the first document viewer. A red callout box points to the 'Properties of Format Locator' window, which shows a table of detected alternatives for the 'FL_InvoiceNumber' locator.

Properties of Format Locator 'FL_InvoiceNumber'

	Confidence	Text
1	100.00 %	242825868
2	96.95 %	333323807
3	96.95 %	242825868
4	0.00 %	000318193
5	0.00 %	333323807
6	0.00 %	333323807
7	0.00 %	000318193
8	0.00 %	333323807
9	0.00 %	000318193
10	0.00 %	333323807

Второй результат исходит со страницы 1, и номер счета-фактуры

Первый проект– Извлечение по форматам и ключевым словам

The screenshot displays the Kofax Transformation Project Builder interface. The main window shows a document with extracted fields: Invoice Number (333323807), Invoice Date (04/27/03), Payment Terms (NET DUE 30 DAYS), Shipped Via (2ND DAY), and Waybill Number (2428258660). A blue box highlights the first two fields. A red box highlights the 'Properties of Format Locator' window, which shows a table of detected alternatives for the 'FL_InvoiceNumber' locator. The table lists 10 alternatives, with the third one (242825868) selected. A red speech bubble points to this row, containing Russian text. The bottom of the interface shows a 'Ready' status and buttons for 'Close', 'Test', and 'Help'.

Page: 1 of 2

Invoice Number: 333323807
Invoice Number (98.95%) 98.95%

Invoice Date: 04/27/03
Payment Terms: NET DUE 30 DAYS
Shipped Via: 2ND DAY
Waybill Number: 2428258660
98.95%

The locator is assigned to the following fields:
<none>

SHIP TO:

Properties of Format Locator 'FL_InvoiceNumber'

	Confidence	Text
1	100.00 %	242825868
2	96.95 %	333323807
3	96.95 %	242825868
4	0.00 %	000318193
5	0.00 %	333323807
6	0.00 %	333323807
7	0.00 %	000318193
8	0.00 %	333323807
9	0.00 %	000318193
10	0.00 %	333323807

SHIP TO:

Amount: 2,282.00

100dpi 44 / 82 1 / 2 TIFF

Ready

Close Test Help

Третий результат исходит со
страницы 1, и это еще раз
накладная. УВЕДОМЛЕНИЕ:
ЗЕЛЕНЫЕ ОСНОВНЫЕ
ПОКАЗАТЕЛИ ВСЕГДА
ВЫБРАННЫЕ РЕЗУЛЬТАТЫ

Первый проект– Извлечение по форматам и ключевым словам

The screenshot displays the Kofax Transformation Project Builder interface. The top window shows a document with extracted fields: Invoice Number (333323807), Invoice Date (04/27/03), Payment Terms (NET DUE 30 DAYS), Shipped Via (2ND DAY), and Waybill Number (2428258660). A blue box highlights the 'Invoice Number' field. The bottom window shows the 'Properties of Format Locator 'FL_InvoiceNumber'' dialog, specifically the 'Evaluation Settings' tab. It includes a 'Keywords' section with 'Ignore these characters' set to '! / () - . : , ;'. The 'Keyword' is 'Invoice Number', and the 'Weight' is 100. The 'Distance' is set to 'Near'. A directional grid shows 'NW', 'N', 'NE', 'W', 'E', 'SW', 'S', 'SE'. The 'Use' column is checked for 'FREIGHT BILL NUM...', 'INVOICE NO', and 'Invoice Number'. The 'Relation' column is set to 'W, N, NW' for all three. The 'Distance' is 'Near' and 'Weight' is '100' for all three. The 'Search Exact' column is checked for 'INVOICE NO' and 'Invoice Number'.

Page: 1 of 2

1496 *

Invoice Number: 333323807

Invoice Number (98.95%) 98.95%

Invoice Date: 04/27/03

Payment Terms: NET DUE 30 DAYS

Shipped Via: 2ND DAY

Waybill Number: 2428258660

Waybill Number (98.95%) 98.95%

The locator is assigned to the following fields:

<none>

SHIP TO:

Page: 1 of 2

1496 *

Invoice Number: 333323807

Invoice Number (98.95%) 98.95%

Invoice Date: 04/27/03

Payment Terms: NET DUE 30 DAYS

Shipped Via: 2ND DAY

Waybill Number: 2428258660

Waybill Number (98.95%) 98.95%

Properties of Format Locator 'FL_InvoiceNumber'

Formats Evaluation Settings Regions Test Results

Keywords

Ignore these characters: ! / () - . : , ;

Keyword: Invoice Number

Match each word exactly (not fuzzy)

Match all words as a phrase

Weight: 100

Distance: Near

NW N NE

W E

SW S SE

Use	D	Keyword	Relation	Distance	Weight	Search Exact
<input checked="" type="checkbox"/>		FREIGHT BILL NUM...	W, N, NW	Near	100	<input type="checkbox"/>
<input checked="" type="checkbox"/>		INVOICE NO	W, N, NW	Near	100	<input type="checkbox"/>
<input checked="" type="checkbox"/>		Invoice Number	W, N, NW	Near	100	<input type="checkbox"/>

Close Test Help

Отношение ключевого слова сейчас важно. Ключевое "Invoice Number" на Западе (W) от цели. Отмените выбор NW и N и нажмите «Изменить».

Первый проект– Извлечение по форматам и ключевым словам

The screenshot displays the Kofax Transformation Project Builder interface. The main window shows a document with extracted fields: Invoice Number (333323807), Invoice Date (04/27/03), Payment Terms (NET DUE 30 DAYS), Shipped Via (2ND DAY), and Waybill Number (2428258660). The 'Properties of Format Locator 'FL_InvoiceNumber'' dialog is open, showing the 'Evaluation Settings' tab. The 'Keywords' section lists 'Invoice Number' with a weight of 100 and a distance of 'Near'. The 'Relation' is set to 'W' (West). The 'Distance' is set to 'Near'. The 'Weight' is set to 100. The 'Match all words as a phrase' checkbox is checked. The 'Ignore these characters' field contains '! / () - . : , ;'. The 'Relation' grid shows 'W' selected. The 'Distance' dropdown is set to 'Near'. The 'Weight' slider is at 100. The 'Match each word exactly (not fuzzy)' checkbox is unchecked. The 'Match all words as a phrase' checkbox is checked. The 'Keywords' table shows the following data:

Use	D	Keyword	Relation	Distance	Weight	Search Exact	Whole Phrase
<input checked="" type="checkbox"/>		FREIGHT BILL NUM...	W, N, NW	Near	100	<input type="checkbox"/>	<input checked="" type="checkbox"/>
<input checked="" type="checkbox"/>		INVOICE NO	W, N, NW	Near	100	<input type="checkbox"/>	<input checked="" type="checkbox"/>
<input checked="" type="checkbox"/>		Invoice Number	W	Near	100	<input type="checkbox"/>	<input checked="" type="checkbox"/>

Связь только западная (W)
Если вы проверите его снова, вы увидите ...

Первый проект– Извлечение по форматам и ключевым словам

The screenshot displays the Kofax Transformation Project Builder interface. The 'Project Tree' on the left shows the 'Invoices' class with a 'Format Locator' assigned to the 'FL_InvoiceNumber' field. The 'Details' pane shows the 'Settings' for this locator, including the 'Definition Class' and 'Locator Method'. The 'Test Results' tab is active, showing a table of detected alternatives. The first alternative, '333323807', is highlighted with a 96.95% confidence level. A blue callout box points to this result, containing the text: 'У вас есть только один результат с высокой степенью уверенности/ Отлично!'. The 'Document Viewer' on the right shows a sample invoice with the 'Invoice Number' field highlighted in green, matching the result in the test results table.

	Confidence	Text
1	96.95 %	333323807
2	0.00 %	000318193
3	0.00 %	333323807
4	0.00 %	242825868
5	0.00 %	333323807
6	0.00 %	000318193
7	0.00 %	333323807
8	0.00 %	000318193
9	0.00 %	333323807
10	0.00 %	242825868

Первый проект– Извлечение по форматам и ключевым словам

Назначьте свое поле локатору

1) Дважды щелкните по полям ниже Счета-фактуры

2) Выберите графу “Field populated by” to FL_InvoiceNumber

Field	Language	Display name	Field populated by	Fixed field value
(4 selected)				
InvoiceNumber (Inv	<Default>	InvoiceNumber	<none>	
InvoiceDate (Invoice	<Default>	InvoiceDate	<none>	
OrderNumber (Invoik	<Default>	OrderNumber	<fixed value>	
Totalamount	<Default>	Totalamount	<none>	

Filename Classification

FedEx40.xdc	FedEx	85.2%
FedEx39.xdc	FedEx	83.8%
FedEx37.xdc	FedEx	83.8%
FedEx38.xdc	FedEx	84.2%
FedEx44.xdc	FedEx	84.1%
FedEx45.xdc	FedEx	86.2%
FedEx43.xdc	FedEx	83.0%
FedEx41.xdc	FedEx	83.8%
FedEx42.xdc	FedEx	83.7%
00000255.xdc	Dell	100.0%
00000256.xdc	Dell	91.3%
00000254.xdc*	Dell	75.5%
00000252.xdc	Dell	80.4%

Первый проект– Извлечение по форматам и ключевым словам

3) Контролируйте новые настройки извлечения

The screenshot shows the 'FirstTry* - Kofax Transformation - Project Builder' window. The 'Project Tree' on the left shows the 'Invoices' class with 'Fields' selected. The main area displays the 'Fields' configuration table for 'Invoices'.

Field	Language	Display name	Field populated by	Fixed field value
(4 selected)				
InvoiceNumber (Inv	<Default>	InvoiceNumber	<none>	
InvoiceDate (Invoice	<Default>	InvoiceDate	<none>	
OrderNumber (Invoik	<Default>	OrderNumber	<fixed value> FL_InvoiceNumber	
Totalamount (Invoic	<Default>	Totalamount	<none>	

The bottom panel shows the 'Documents - Benchmark Set: "C:\Documents\unsorted" *' with a list of documents and their classification results.

Filename	Classification Res	Confidence	Assigned Class
FedEx40.xdc	FedEx	85.2%	
FedEx39.xdc	FedEx	83.8%	
FedEx37.xdc	FedEx	85.1%	
FedEx38.xdc	FedEx	84.2%	
FedEx44.xdc	FedEx	84.1%	
FedEx45.xdc	FedEx	98.2%	
FedEx43.xdc	FedEx	83.0%	
FedEx41.xdc	FedEx	83.8%	
FedEx42.xdc	FedEx	83.7%	
00000255.xdc	Dell	100.0%	
00000256.xdc	Dell	91.3%	
00000254.xdc*	Dell	75.5%	
00000252.xdc	Dell	80.4%	

Первый проект– Извлечение по форматам и ключевым словам

The screenshot displays the Kofax Transformation Project Builder interface. The 'Process' tab is active, showing a workflow with steps like Recognize, Extract, Verify, Separate, Process, Runtime Script Events, Load Data, Classify, Validate, and Auto-Folding. The 'Project Tree' on the left shows the project structure, including 'Invoices' and 'Fields'. A blue callout box with the text 'Значительное улучшение' (Significant improvement) points to the 'Invoices' class in the Project Tree and the 'Summary' table in the 'Extraction Benchmark' window.

Extraction Benchmark - Extraction Benchmark (All Classes)

Legend

- A Correct valid fields
- B Correct invalid fields
- C Incorrect invalid fields
- D Incorrect valid fields

Summary

	InvoiceNumber	InvoiceDate	OrderNumber	Totalamount
All Fields				
A	23.44%	62.50%	18.75%	6.25%
B	17.19%	6.25%	6.25%	25.00%
C	0.00%	6.25%	75.00%	87.50%
D	0.00%	0.00%	0.00%	68.75%

Details - Sorted by Column Content

Filename	Class	A	B	C	D	InvoiceNumber	InvoiceDate	OrderNumber	Totalamount
000001FC.xd	Oki	0.00%	0.00%	100.00%	0.00%	AGEPRODUCTS\JNC.y7\ [310264 RI]	-/- [05/21/2003]	05 / 21 / 03 [20030032]	-/- [21553.42]
000001FD.xd	Oki	0.00%	50.00%	50.00%	0.00%	307707 RI	// uj [02/17/2003]	2UUUUU21 [20030021]	21590.66
000001FE.xd	Oki	0.00%	0.00%	0.00%	0.00%	307243 RI	01/29/2003	20020280	21540.00
000001FF.xd	Oki	25.00%	50.00%	25.00%	0.00%	306091 RI	12/09/2002	20026193 [20026193]	21540.00
00000200.xd	Oki	0.00%	50.00%	50.00%	0.00%	303947 RI	u? / ? z z uz: [09/27/2002]	/SUU?UUZVJ [20020020]	21540.00
00000201.xd	Oki	75.00%	0.00%	0.00%	0.00%	301814 RI	07/25/2002	20020020	USD [0.00]
00000202.xd	Oki	0.00%	75.00%	25.00%	0.00%	300603 RI	06/19/2002	2###2###2 [20020020]	21555.74
0000024F.xd	Dell	25.00%	0.00%	75.00%	0.00%	196328182	-/- [12/26/2002]	-/- [20026306]	-/- [2038.66]
00000250.xd	Dell	25.00%	0.00%	75.00%	0.00%	347784804	-/- [05/18/2003]	-/- [20036121]	-/- [1702.46]
00000251.xd	Dell	25.00%	0.00%	75.00%	0.00%	345004270	-/- [05/08/2003]	-/- [20036114]	-/- [1404.28]
00000252.xd	Dell	25.00%	0.00%	75.00%	0.00%	344964623	-/- [05/09/2003]	-/- [20036113]	-/- [1318.86]
00000253.xd	Dell	25.00%	0.00%	75.00%	0.00%	341350479	-/- [05/08/2003]	-/- [20036106]	-/- [1053.80]

Progress

Extraction complete

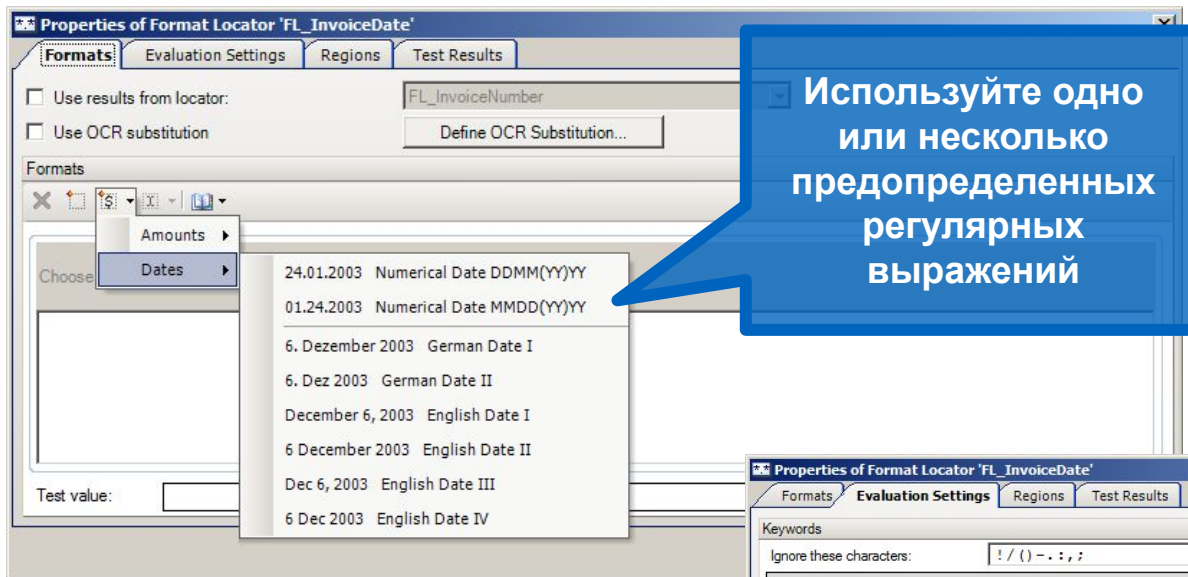
Processed 16 of 82 documents

Elapsed Time: 5 seconds

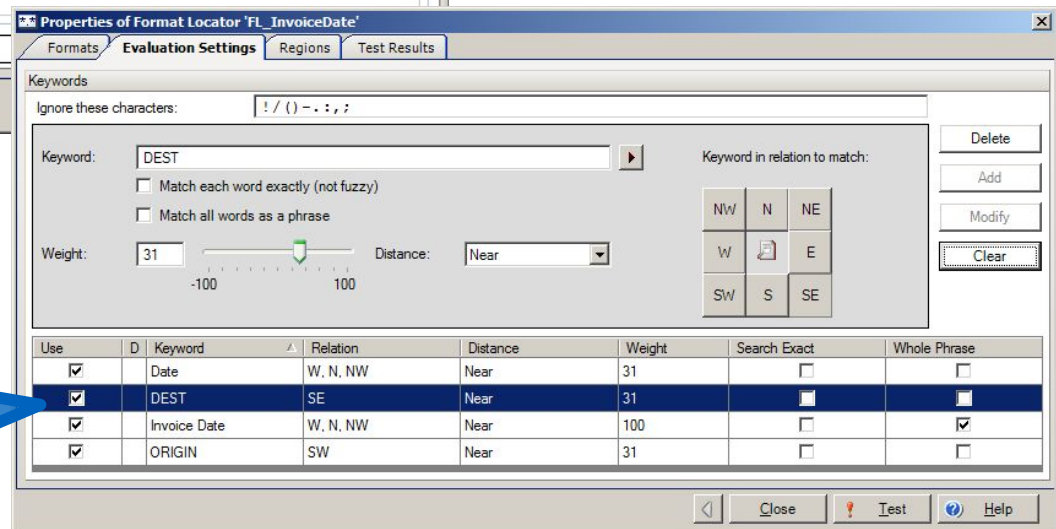
Buttons: Start, Stop, Save..., Export..., Help

Первый проект– Извлечение по форматам и ключевым словам

◆ Настройки для FL_InvoiceDate:

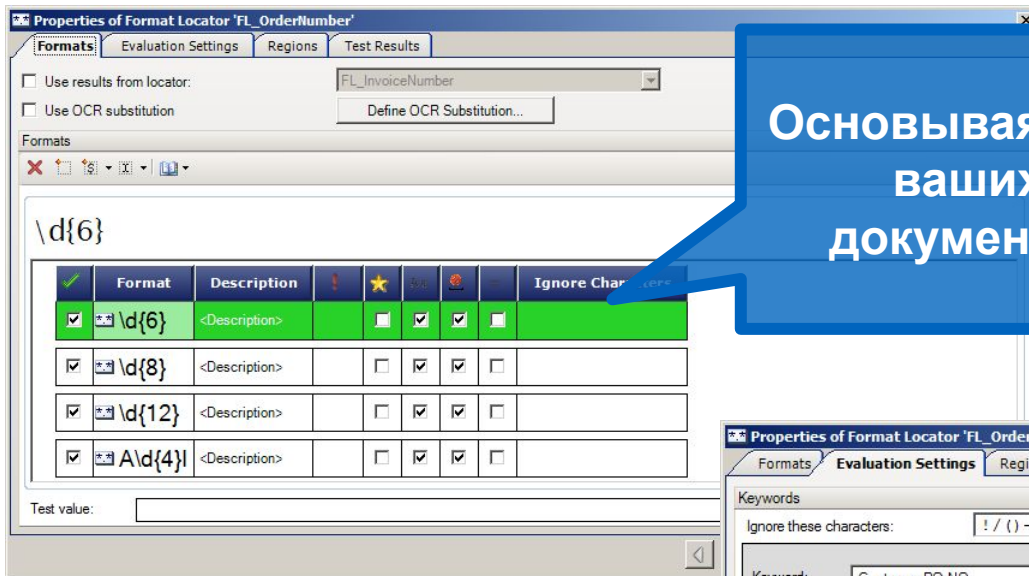


Используйте
ключевые слова



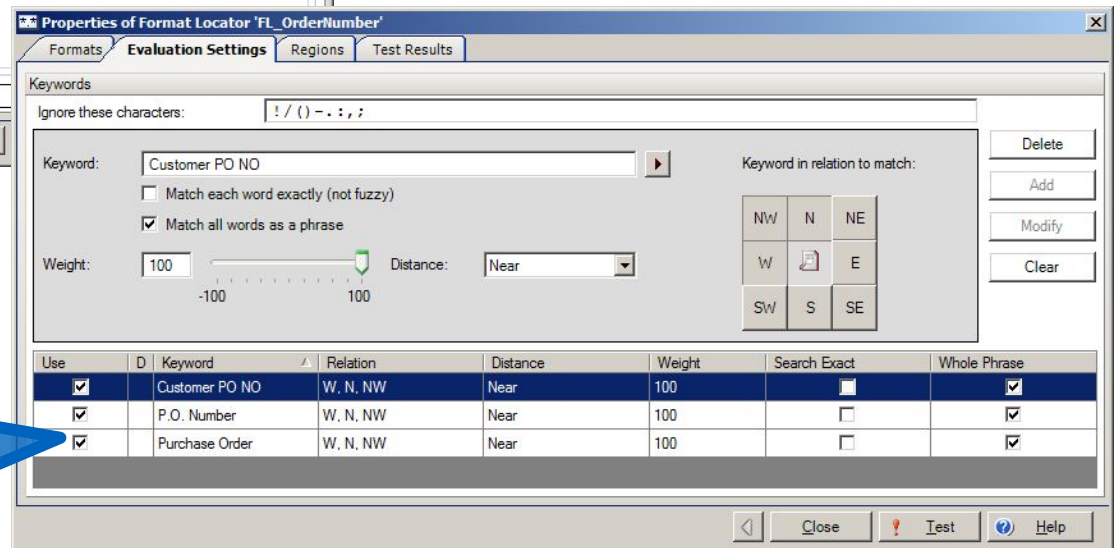
Первый проект– Извлечение по форматам и ключевым словам

◆ Настройки для FL_OrderNumber:



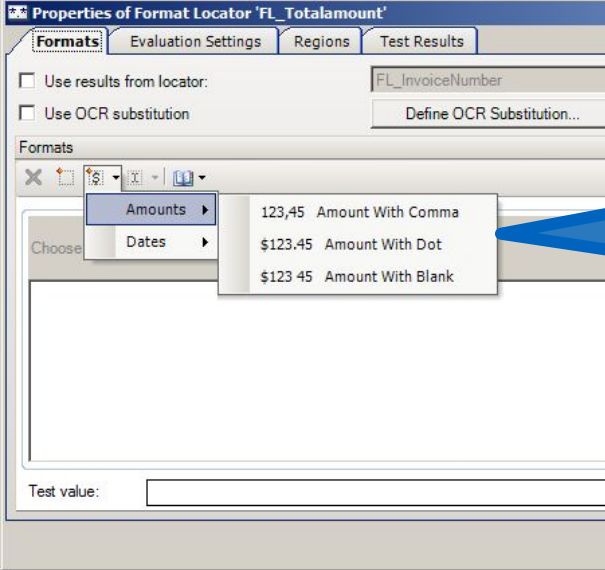
Основываясь на
ваших
документах

Используйте
ключевые слова

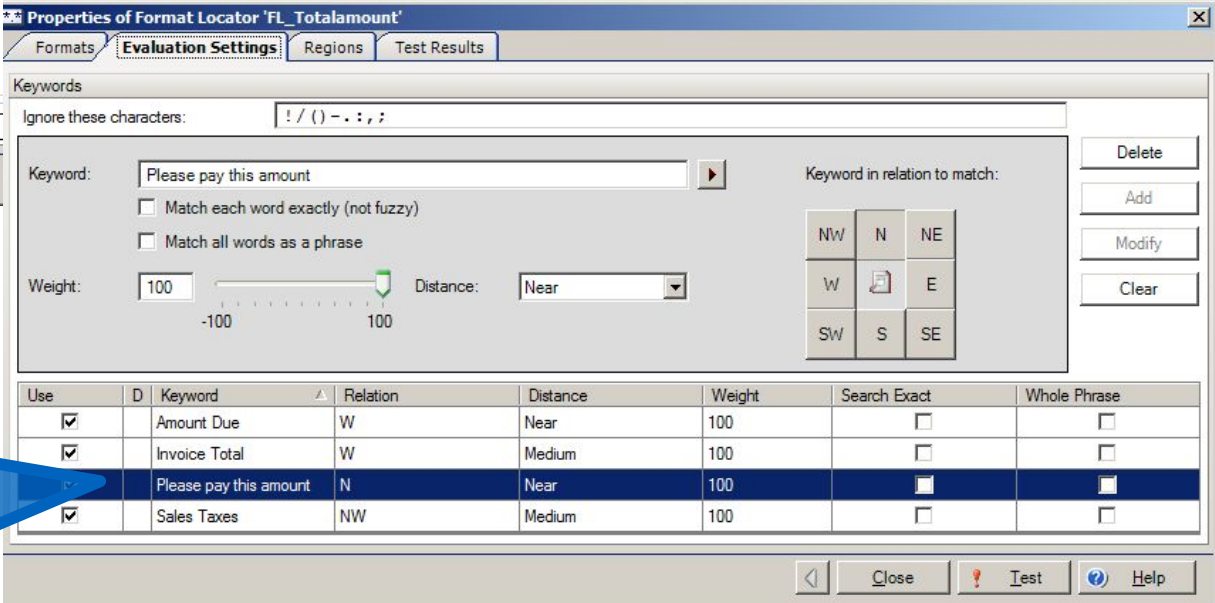


Первый проект– Извлечение по форматам и ключевым словам

◆ Настройки для FL_OrderNumber:



Предопределенные регулярные выражения для доступных сумм



Используйте ключевые слова

Use	D	Keyword	Relation	Distance	Weight	Search Exact	Whole Phrase
<input checked="" type="checkbox"/>		Amount Due	W	Near	100	<input type="checkbox"/>	<input type="checkbox"/>
<input checked="" type="checkbox"/>		Invoice Total	W	Medium	100	<input type="checkbox"/>	<input type="checkbox"/>
<input checked="" type="checkbox"/>		Please pay this amount	N	Near	100	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
<input checked="" type="checkbox"/>		Sales Taxes	NW	Medium	100	<input type="checkbox"/>	<input type="checkbox"/>

Первый проект– Извлечение по форматам и ключевым словам

FirstTry - Kofax Transformation - Project Builder

Project Documents Design Process View

Classification Recognize Extract Verify Clear Data Separation
Extraction Separate Process Runtime Script Events Load Data Classification
Optimize Content Classifier Classify Validate Auto-Folding Test Extraction Benchmark

Project Tree

Type to search

Project

Project Class

Invoices

Fields

InvoiceNumber
InvoiceDate
OrderNumber
Totalamount

Locators
Validation Rules
Structure Rules
Instructions
Classes

Details

Multiple Selection: Fields

Field	Language	Display name	Field populated by	Fixed field value
(3 selected)	<Default>			
InvoiceNumber (Inv	<Default>	InvoiceNumber	FLInvoiceNumber	
InvoiceDate (Invoice	<Default>	InvoiceDate	FLInvoiceDate	
OrderNumber (Invoic	<Default>	OrderNumber	FLOrderNumber	
Totalamount (Invoic	<Default>	Totalamount	FL_Totalamount	

Dropdown menu options:

- <none>
- <fixed value>
- FL_InvoiceNumber
- FL_InvoiceDate
- FL_OrderNumber
- FL_Totalamount

Test Set

Filename

Classification Res:

Confidence

Assigned Class

Ready

Дважды щелкните по полям ниже
Счета-фактуры класса
Свяжите все свои поля с локаторами

Первый проект– Извлечение по форматам и ключевым словам

Запустите extraction benchmark еще раз

Прекрасный результат
Но это недостаточно
хорошо 😊

Extraction Benchmark - Extraction Benchmark (All Classes)

Legend

- A Correct valid fields
- B Correct invalid fields
- C Incorrect invalid fields
- D Incorrect valid fields

Summary

	All Fields	InvoiceNumber	InvoiceDate	OrderNumber	Totalamount
A	66.16%	95.12%	39.02%	41.46%	89.02%
B	29.88%	3.66%	60.98%	52.44%	2.44%
C	3.96%	1.22%	0.00%	6.10%	8.54%
D	0.00%	0.00%	0.00%	0.00%	0.00%

Details - Sorted by Column Content

Filename	Class	A	B	C	D	InvoiceNumber	InvoiceDate	OrderNumber	Totalamount
000001FF.xd	Oki	100.00%	0.00%	0.00%	0.00%	310264RI	05/21/2003	20030032	21553.42
000001FD.xd	Oki	75.00%	25.00%	0.00%	0.00%	307707RI	02/17/2003	15800051	21590.66
00000200.xd	Oki	100.00%	0.00%	0.00%	0.00%	307243RI	01/29/2003	20020280	21540.00
00000201.xd	Oki	100.00%	0.00%	0.00%	0.00%	306091RI	12/09/2002	15800051	21540.00
00000202.xd	Oki	100.00%	0.00%	0.00%	0.00%	303947RI	09/27/2002	15800051	21540.00
00000202.xd	Oki	100.00%	0.00%	0.00%	0.00%	301814RI	07/25/2002	20020020	21555.74
00000202.xd	Oki	100.00%	0.00%	0.00%	0.00%	300603RI	06/19/2002	15800051	21555.74
00000202.xd	Oki	100.00%	0.00%	0.00%	0.00%	196328182	12/26/2002	20026306	2038.66
00000205.xd	Dell	75.00%	25.00%	0.00%	0.00%	347784804	05/18/2003	20036121	1702.46
00000251.xd	Dell	75.00%	25.00%	0.00%	0.00%	345004270	05/08/2003	20036114	1404.28
00000252.xd	Dell	75.00%	25.00%	0.00%	0.00%	344964623	05/09/2003	20036113	1318.86
00000253.xd	Dell	75.00%	25.00%	0.00%	0.00%	341350479	05/08/2003	20036106	1053.80

Progress

Extraction complete

Processed 82 documents

Elapsed Time: 25 seconds

Первый проект– Извлечение по форматам и ключевым словам

- ◆ Задание
 - ◆ Добавить локаторы формата в класс Invoices
 - ◆ Свяжите свои новые локаторы с полями
 - ◆ Отвяжите поля из AZL в классе Oki (Установите его обратно в унаследованное)
 - ◆ Запустить extraction benchmark
 - ◆ Просмотрите свои результаты

Первый проект - Усовершенствования

Всегда улучшайте свой проект

Всегда следите за своим оператором

Первый проект - Усовершенствования

- ◆ Первый проект имеет хорошие показатели распознавания, но есть еще место для дальнейшего совершенствования
- ◆ Если вы хотите оптимизировать извлечение, хороший способ - сначала проверить тест извлечения.
- ◆ Еще один способ получить дополнительные знания - проверить многие документы и сделать несколько заметок
- ◆ Давайте подробнее рассмотрим эталонный тест извлечения

Первый проект - Усовершенствования

Extraction Benchmark - Extraction Benchmark (All Classes)

Benchmark Comparison

Legend

A Correct valid fields **B** Correct invalid fields **C** Incorrect invalid fields

Зеленые поля корректные (будут пропущены на валидации)

	InvoiceDate	OrderNumber	Totalamount
A	65.85%	93.90%	39.02%
B	0.00%	0.00%	0.00%
C	3.96%	1.22%	0.00%

Зеленый знак на валидации означает, что значение правильное.

Filename	OK	A	B	C	D	InvoiceNumber	InvoiceDate	OrderNumber	Totalamount
000001FC.xd	Ok	50.00%	0.00%	50.00%	0.00%	-/- [310264 RI]	05/21/2003	-/- [20030032]	21553.42
000001FD.xd	Ok	50.00%	0.00%	50.00%	25.00%	307707RI [307707 RI]	02/17/2003		
000001FE.xd	Ok	100.00%	0.00%	0.00%	0.00%	307243RI	01/29/2003		
000001FF.xd	Ok	100.00%	0.00%	0.00%	0.00%	306091RI	12/09/2002		
00000200.xd	Ok	75.00%	25.00%	0.00%	0.00%	303947RI	09/27/2002		
00000201.xd	Ok	50.00%	25.00%	25.00%	0.00%	301814RI	07/25/2002		
00000202.xd	Ok	100.00%	0.00%	0.00%	0.00%	300603RI	06/19/2002		
0000024F.xd	Dell	75.00%	25.00%	0.00%	0.00%	196328182	12/26/2002		
00000250.xd	Dell	75.00%	25.00%	0.00%	0.00%	347784804	05/18/2003		
00000251.xd	Dell	75.00%	25.00%	0.00%	0.00%	345004270	05/08/2003		
00000252.xd	Dell	75.00%	25.00%	0.00%	0.00%	344964623	05/09/2003		
00000253.xd	Dell	75.00%	25.00%	0.00%	0.00%	341350479	05/08/2003		

Progress

Extraction complete

Processed 82 documents

Elapsed Time: 24 seconds

Test Validation

Batch Edit View Folder Document Page Field Options Help

Classification Result Best Fit

The document was classified to the following [Ok]

4 fields valid, 0 fields invalid, 0 fields invisible, 0 fields read-only

InvoiceNumber 306091 RI

InvoiceDate 12/09/02

OrderNumber 15800051

Totalamount 21,540.00

Invoice

OKI Semiconductor

INVOICE

DATE: 12/09/02

TO: OKI SEMICONDUCTOR, INC.

FROM: OKI SEMICONDUCTOR, INC.

AMOUNT: 21,540.00

STATUS: PAID

RECEIVED 02-11-03

InvoiceNumber 306091 RI

InvoiceNumber 306091RI

Первый проект - Усовершенствования

Восклицательный знак
Значение поля пустое
Кто-то должен ввести значение поля

Пустые поля - это ошибочные поля (не соответствующие значению в наборе ссылок).
и отображаются в заголовке, и кто-то должен подтвердить свои значения.

Progress
Extraction complete
100 %
Processed 82 documents
Elapsed Time: 24 seconds

InvoiceNumber	InvoiceDate	OrderNumber	Totalamount
307707RI	05/21/2003	21553.42	
307243RI	01/29/2003		
306091RI	12/09/2002		
303947RI	09/27/2002		
301814RI	07/25/2002		
300603RI	06/19/2002		
196328182	12/26/2002		
347784804	05/18/2003		
345004270	05/08/2003		
344964623	05/09/2003		
341350479	05/08/2003		

Test Validation
The document was classified to the following (Ok)
2 fields valid, 2 fields invalid (0 fields invisible, 0 fields read-only)
InvoiceNumber: 307707RI
InvoiceDate: 05/21/03
OrderNumber: 21553.42
Totalamount: 21,553.42

Classification Result
The document was classified to the following (Ok)
2 fields valid, 2 fields invalid (0 fields invisible, 0 fields read-only)
InvoiceNumber: 307707RI
InvoiceDate: 05/21/03
OrderNumber: 21553.42
Totalamount: 21,553.42

Ok! Semiconductor
RECEIVED MAY 27 2003

Первый проект - Усовершенствования

Extraction Benchmark - Extraction Benchmark (All Classes)

Benchmark | Comparison

Legend

- A** Correct valid fields
- B** Correct invalid fields
- C** Incorrect invalid fields
- D** Incorrect valid fields

Progress

Extraction complete

Processed 82 documents

Elapsed Time: 24 seconds

100 %

Filename	Class	A	B	C	D	InvoiceNumber	InvoiceDate	OrderNumber	Totalamount
0000015D.xd	Ok	50.00%	0.00%	25.00%	25.00%	307707RI	02/17/2003	15800051	21553.42
0000015E.xd	Ok	100.00%	0.00%	0.00%	0.00%	307243RI	01/29/2003		
0000015F.xd	Ok	100.00%	0.00%	0.00%	0.00%	306091RI	12/09/2002		
00000201.xd	Ok	50.00%	25.00%	5.00%	0.00%	301814RI	07/25/2002		
00000202.xd	Ok	100.00%	0.00%	0.00%	0.00%	300603RI	06/19/2002		
0000024E.xd	Dell	75.00%	25.00%	0.00%	0.00%	196328182	12/26/2002		
00000250.xd	Dell	75.00%	25.00%	0.00%	0.00%	347784804	05/18/2003		
00000251.xd	Dell	75.00%	25.00%	0.00%	0.00%	345004270	05/08/2003		
00000252.xd	Dell	75.00%	25.00%	0.00%	0.00%	344964623	05/09/2003		
00000253.xd	Dell	75.00%	25.00%	0.00%	0.00%	341350479	05/08/2003		

Test Validation

Batch Edit View Folder Document Page Field Options Help

Classification Result

The document was classified to the following:

3 fields valid, 1 field invalid, 0 fields invisible, 0 fields read-only

InvoiceNumber: 307707 RI

InvoiceDate: 02/17/2003

OrderNumber: 15800051

Totalamount:

Current Error

The amount field does not contain a valid amount.

Invoice

Ok! Semiconductor

INVOICE

DATE: 02/17/2003

AMOUNT: 21553.42

ORDER: 15800051

INVOICED 10/13/2003

Красные поля это поля которые неверны, И никто не будет их рассматривать, потому что КТМ уверен, что их значение правильное.

Зеленый значок спереди

Значение некорректно

Первый проект - Усовершенствования

Extraction Benchmark - Extraction Benchmark (All Classes)

Benchmark | Comparison

Legend

- A** Correct valid fields
- B** Correct invalid fields

All Fields	InvoiceNumber	InvoiceDate	OrderNumber	Totalamount
A 65.85%	93.90%	39.02%	41.46%	89.02%
B 60.98%			52.44%	2.44%
C 3.96%	1.22%	0.00%	6.10%	8.54%
D 0.00%	0.00%	0.00%	0.00%	0.00%

Test Validation

Batch Content

Classification Result

The document was classified to the following (Ok)

2 fields valid, 2 fields invalid (0 fields invisible, 0 fields read-only)

InvoiceNumber 301814 RI

InvoiceDate 07/25/02

OrderNumber 20020020

Totalamount

Callout 1 (Left): Синие поля - это поля, которые являются правильными, НО кто-то должен их подтвердить.

Callout 2 (Right): Красный восклицательный знак Значение неверное

Progress

Extraction complete

100 %

Processed 82 documents

Elapsed Time: 24 seconds

Buttons: Start, Stop, Save..., Export..., Help

Первый проект - Усовершенствования

Extraction Benchmark - Extraction Benchmark (All Classes)

Benchmark Comparison

Legend

A Correct valid fields B Correct invalid fields C Incorrect invalid fields D Incorrect valid fields

Summary

	All Fields	InvoiceNumber	InvoiceDate	OrderNumber	Totalamount
A	65.85%	93.90%	39.02%	41.46%	89.02%
B	29.88%	3.66%	60.98%	52.44%	2.44%
C	3.96%	1.22%	0.00%	6.10%	8.54%
D	0.30%	1.22%	0.00%	0.00%	0.00%

Details - Sorted by Column Content

Filename	Class	A	B	C	D	InvoiceNumber	InvoiceDate	OrderNumber	Totalamount
000001FC.xd	Ok	50.00%	0.00%	50.00%	0.00%	-/- [310264 RI]	05/21/2003	-/- [20030032]	21553.42
000001FD.xd	Ok	50.00%	0.00%	25.00%	25.00%	307707RI [307707 RI]	02/17/2003	15800051	-/- [21590.66]
000001FE.xd	Ok	100.00%	0.00%	0.00%	0.00%	307243RI	01/29/2003	20020280	21540.00
000001FF.xd	Ok	100.00%	0.00%	0.00%	0.00%	306091RI	12/09/2002	15800051	21540.00
00000200.xd	Ok	75.00%	25.00%	0.00%	0.00%	303947RI	09/27/2002	15800051	21540.00
00000201.xd	Ok	50.00%	25.00%	25.00%	0.00%	301814RI	07/25/2002	20020020	-/- [0.00]
00000202.xd	Ok	100.00%	0.00%	0.00%	0.00%	300603RI	06/19/2002	15800051	21555.74
0000024F.xd	Dell	75.00%	25.00%	0.00%	0.00%	196328182	12/26/2002	20026306	2038.66
00000250.xd	Dell	75.00%	25.00%	0.00%	0.00%	347784804	05/18/2003	20036121	1702.46
00000251.xd	Dell	75.00%	25.00%	0.00%	0.00%	345004270	05/08/2003	20036114	1404.28
00000252.xd	Dell	75.00%	25.00%	0.00%	0.00%	344964623	05/09/2003	20036113	1318.86
00000253.xd	Dell	75.00%	25.00%	0.00%	0.00%	341350479	05/08/2003	20036106	1053.80

Progress

Extraction complete

Processed 82 documents

Elapsed Time: 24 seconds

Start Stop

Save... Export... Help

Сохраните результаты, прежде чем пытаться улучшить свой проект ...

Первый проект - Усовершенствования

.. потому что позже вы можете сравнить свои тесты

Extraction Benchmark - Extraction Benchmark (All Classes)

Benchmark **Comparison**

Legend

A Correct valid fields B Correct invalid fields C Incorrect invalid fields D Incorrect valid fields

Summary

	All Fields	InvoiceNumber	InvoiceDate	OrderNumber	Totalamount
A	65.85%	93.90%	39.02%		
B	29.88%	3.66%	60.98%		
C	3.96%	1.22%	0.00%		
D	0.30%	1.22%	0.00%		

Details - Sorted by Column Content

Filename	Class	A	B
000001FC.xd	Ok	50.00%	0.00%
000001FD.xd	Ok	50.00%	0.00%
000001FE.xd	Ok	100.00%	0.00%
000001FF.xd	Ok	100.00%	0.00%
00000200.xd	Ok	75.00%	25.00%
00000201.xd	Ok	50.00%	25.00%
00000202.xd	Ok	100.00%	0.00%
0000024F.xd	Dell	75.00%	25.00%
00000250.xd	Dell	75.00%	25.00%
00000251.xd	Dell	75.00%	25.00%
00000252.xd	Dell	75.00%	25.00%
00000253.xd	Dell	75.00%	25.00%

Progress

Extraction complete

Processed 82 documents

Elapsed Time: 24 seconds

Extraction Benchmark - Extraction Benchmark (All Classes)

Benchmark **Comparison**

Legend

A Correct valid fields B Correct invalid fields C Incorrect invalid fields D Incorrect valid fields

Benchmark 1

Select Benchmark 1: BM_2013_7_19_14_10_2

Benchmark 1 details

Date: 2013-07-19 14:10:02

Comment:

☒ Baseline Delete

Benchmark 2

Select Benchmark 2: BM_2013_7_19_14_11_13

Benchmark 2 details

Date: 2013-07-19 14:11:13

Comment:

☐ Delete

Summary Statistics Comparison

	All Fields	InvoiceNumber	InvoiceDate	OrderNumber	Totalamount
A	65.85% [23.48%]	93.90% [93.90%]	39.02%	41.46%	89.02%
B	29.88% [0.91%]	3.66% [3.66%]	60.98%	52.44%	2.44%
C	3.96% [-24.70%]	1.22% [-98.78%]	0.00%	6.10%	8.54%
D	0.30% [0.30%]	1.22% [1.22%]	0.00%	0.00%	0.00%

Details Statistics Comparison

Filename	Class	A	B	C	D	InvoiceNumber	InvoiceDate	OrderNumber
C:\Documents\unsorted\000001FC.xdc	Ok	50.00%	0.00%	50.00%	0.00%	-/-	05/21/2003	-/-
C:\Documents\unsorted\000001FD.xdc	Ok	50.00%	0.00%	25.00% [-25.00%]	25.00% [25.00%]	307707RI [-/-]	02/17/2003	15800051
C:\Documents\unsorted\000001FE.xdc	Ok	100.00% [25.00%]	0.00%	0.00% [-25.00%]	0.00%	307243RI [-/-]	01/29/2003	20020280
C:\Documents\unsorted\000001FF.xdc	Ok	100.00% [25.00%]	0.00%	0.00% [-25.00%]	0.00%	306091RI [-/-]	12/09/2002	15800051
C:\Documents\unsorted\00000200.xdc	Ok	75.00% [25.00%]	25.00%	0.00% [-25.00%]	0.00%	303947RI [-/-]	09/27/2002	15800051
C:\Documents\unsorted\00000201.xdc	Ok	50.00% [25.00%]	25.00%	25.00% [-25.00%]	0.00%	301814RI [-/-]	07/25/2002	20020020
C:\Documents\unsorted\00000202.xdc	Ok	100.00% [25.00%]	0.00%	0.00% [-25.00%]	0.00%	300603RI [-/-]	06/19/2002	15800051

Save... Export... Help

Elapsed Time: 25 seconds

Первый проект - Усовершенствования

- ◆ Почему поле зеленое или красное?
- ◆ Во-первых проверьте уверенность извлечения
 - ◆ Во-вторых проверьте расстояние между первым и вторым результатом
 - ◆ В-третьих проверьте форматирование
 - ◆ В-четвертых проверьте правило валидации
- ◆ Что такое правило валидации?
 - ◆ Обсудим позже

Первый проект - Усовершенствования

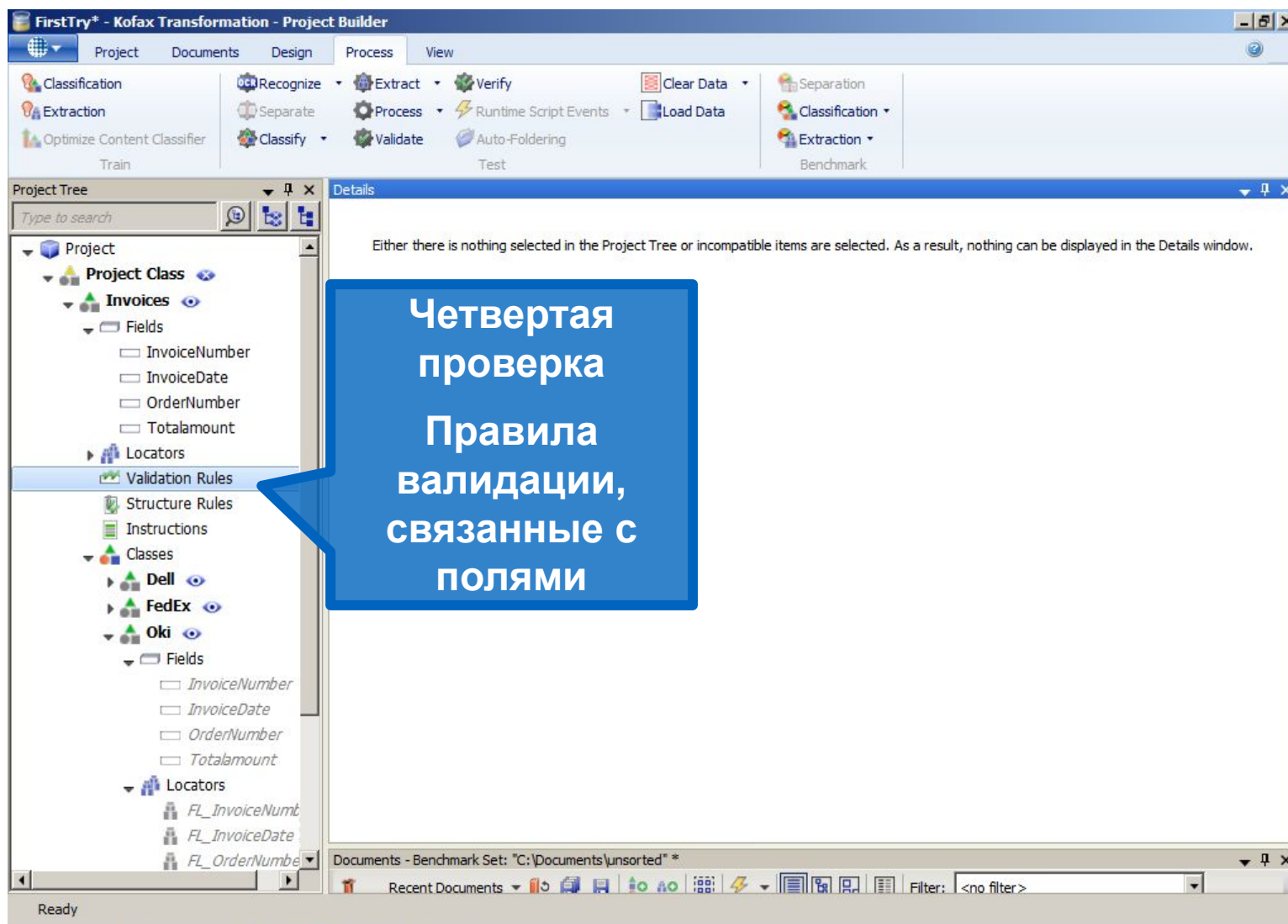
The screenshot displays the 'FirstTry* - Kofax Transformation - Project Builder' window. The 'Project Tree' on the left shows a hierarchy: Project > Project Class > Invoices > Fields > InvoiceNumber. The 'Details' pane on the right is titled 'Field "InvoiceNumber" in Class "Invoices"' and contains the following sections:

- General**
 - Field populated by: FL_InvoiceNumber
 - Display name: InvoiceNumber
 - Field formatting method: <no formatting>
 - ☐ Always CAPITAL letters
 - ☒ Use for statistics
- Training**
 - Group name: [empty]
 - ☒ Show in training dialog
 - ☐ Monitor for Automatic Learning
- Thresholds for Validation**
 - Minimum confidence for locator: 10 %
 - Minimum confidence for valid: 80 %
 - Minimum distance to second best alternative: 10 %
 - ☐ Require manual field confirmation
 - ☐ Always valid
- Reread Options**
 - Recognition profile: <no rereading>
 - Minimum confidence to accept reread result: 70 %

Three blue callout boxes highlight specific features:

- Первая проверка Порог валидации** (First check: Validation threshold) points to the 'Thresholds for Validation' section.
- Третья проверка Форматтер** (Third check: Formatter) points to the 'Field formatting method' dropdown.
- Вторая проверка Расстояние между двумя результатами** (Second check: Distance between two results) points to the 'Minimum distance to second best alternative' slider.

Первый проект - Усовершенствования



Первый проект - Усовершенствования

Extraction Benchmark - Extraction Benchmark (All Classes)

Benchmark Comparison

Legend

A Correct valid fields **B** Correct invalid fields **C** Incorrect invalid fields **D** Incorrect valid fields

Summary

	All Fields	InvoiceNumber	InvoiceDate	OrderNumber	Total amount
A	65.85%	93.90%	39.02%	41.46%	89.02%
B	29.88%	3.66%	60.98%	52.44%	2.44%
C	3.96%	1.22%	0.00%	6.10%	8.54%
D	0.30%	1.22%	0.00%	0.00%	0.00%

Details - Sorted by Column Content

Filename	Class	A	B	C	D	InvoiceNumber	InvoiceDate	OrderNumber	Total amount
000001FC.xd	Ok	50.00%	0.00%	50.00%	0.00%	-/- [310264 RI]	05/21/2003	-/- [20030032]	21553.42
000001FD.xd	Ok	50.00%	0.00%	25.00%	25.00%	307707RI [307707 RI]	02/17/2003	15800051	-/- [21590.66]
000001FE.xd	Ok	100.00%	0.00%	0.00%	0.00%	307243RI	01/29/2003	20020280	21540.00
000001FF.xd	Ok	100.00%	0.00%	0.00%	0.00%	306091RI	12/09/2002	15800051	21540.00
00000200.xd	Ok	75.00%	25.00%	0.00%	0.00%	303947RI	09/27/2002	15800051	21540.00
00000201.xd	Ok	50.00%	25.00%	25.00%	0.00%	301814RI	07/25/2002	20020020	-/- [0.00]
00000202.xd	Ok	100.00%	0.00%	0.00%	0.00%	300603RI	06/19/2002	15800051	21555.74
0000024F.xd	Dell	75.00%	25.00%	0.00%	0.00%	196328182	12/26/2002	20026306	2038.66
00000250.xd	Dell	75.00%	25.00%	0.00%	0.00%	347784804	05/18/2003	20036121	1702.46
00000251.xd	Dell	75.00%	25.00%	0.00%	0.00%	345004270	05/08/2003	20036114	1404.28
00000252.xd	Dell	75.00%	25.00%	0.00%	0.00%	344964623	05/09/2003	20036113	1318.86
00000253.xd	Dell	75.00%	25.00%	0.00%	0.00%	341350479	05/08/2003	20036106	1053.80

Progress

Extraction complete

100 %

Processed 82 documents

Start Stop

Save... Export... ? Help

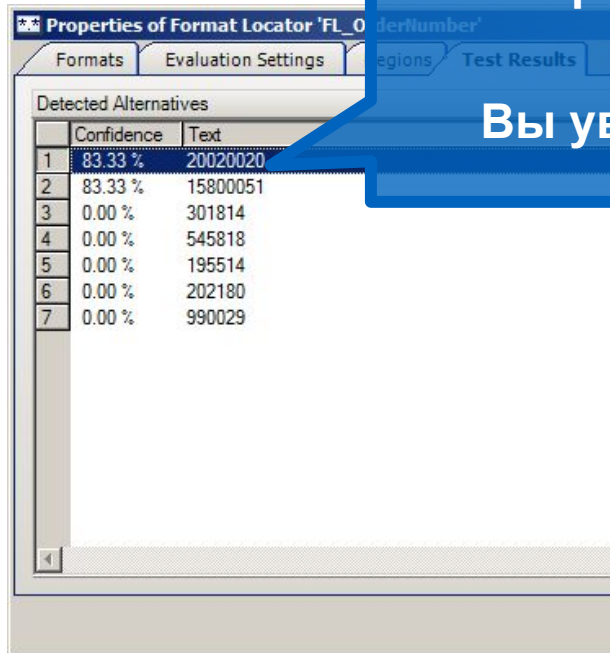
Elapsed Time: 24 seconds

Номер заказа:
Поскольку у нас нет форматов или правил валидации, применяемых к порядковому номеру, состояние поля красное или зеленое на основе уверенности.

Первый проект - Усовершенствования

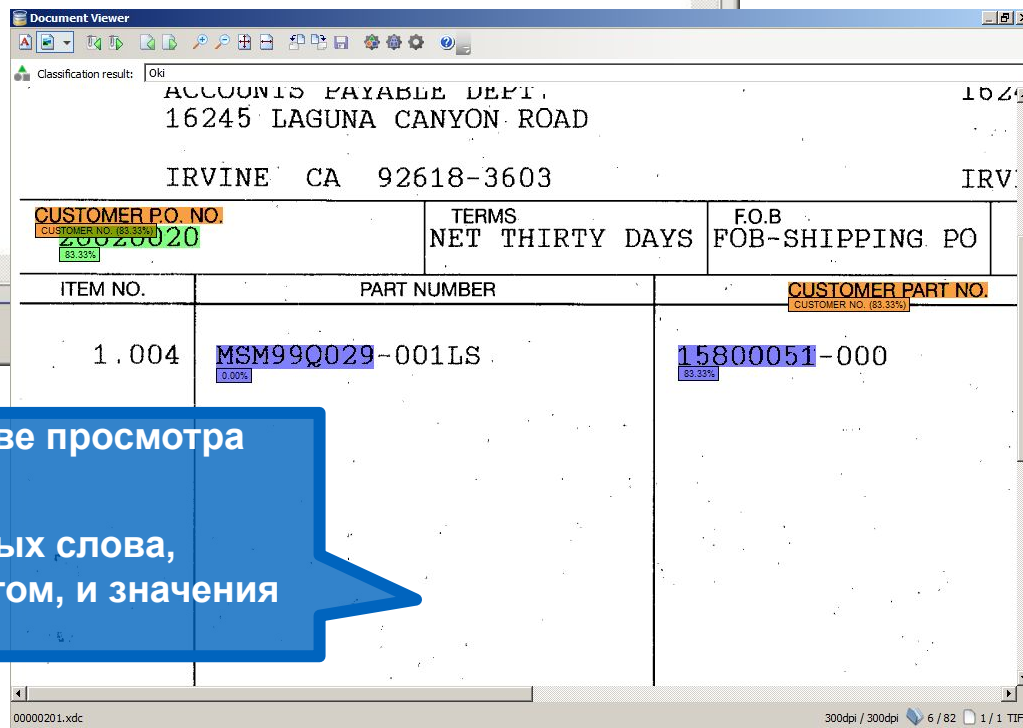
Проверьте свой локатор
FL_OrderNumber

Вы увидите два результата с
доверием 83,33%



The screenshot shows the 'Properties of Format Locator' window with the 'Test Results' tab selected. It displays a table of detected alternatives for the 'FL_OrderNumber' locator.

	Confidence	Text
1	83.33 %	20020020
2	83.33 %	15800051
3	0.00 %	301814
4	0.00 %	545818
5	0.00 %	195514
6	0.00 %	202180
7	0.00 %	990029



The screenshot shows a scanned invoice in a Document Viewer. The invoice contains the following information:

Classification result: Oki

ACCOUNTS PAYABLE DEPT.
16245 LAGUNA CANYON ROAD
IRVINE CA 92618-3603

CUSTOMER P.O. NO. 20020020 (83.33%)
TERMS NET THIRTY DAYS
F.O.B. FOB-SHIPING PO

ITEM NO.	PART NUMBER	CUSTOMER PART NO.
1.004	MSM99Q029-001LS (0.00%)	15800051-000 (83.33%)

Протестируйте его в средстве просмотра
документов

Вы увидите два ключевых слова,
выделенных оранжевым цветом, и значения
рядом с ним

Первый проект - Усовершенствования

Properties of Format Locator 'FL_OrderNumber'

Formats Evaluation Settings Regions Test Results

Detected Alternatives

	Confidence	Text
1	83.33 %	20020020
2	83.33 %	15800051
3	0.00 %	301814
4	0.00 %	545818
5	0.00 %	195514
6	0.00 %	202180
7	0.00 %	990029

PO NO клиента был
найден нечетким
поиском

Система получает
номер клиента, это
83,33% вашего
ключевого слова
Клиент PO NO

Document Viewer

Classification result: Oki

ACCOUNTS PAYABLE DEPT. 102

16245 LAGUNA CANYON ROAD

IRVINE CA 92618-3603 IRV

CUSTOMER P.O. NO. CUSTOMER NO. (83.33%) 20020020 83.33%	TERMS NET THIRTY DAYS	F.O.B FOB-SHIPPING PO
ITEM NO. 1.004	PART NUMBER MSM99Q020-001LS 83.33%	CUSTOMER PART NO. CUSTOMER NO. (83.33%) 15800051-000 83.33%

CUSTOMER P.O. NO.
CUSTOMER NO. (83.33%)
20020020
83.33%

Первый проект - Усовершенствования

Properties of Format Locator 'FL_OrderNumber'

Formats Evaluation Settings Regions **Test Results**

Detected Alternatives

	Confidence	Text
1	83.33 %	20020020
2	83.33 %	15800051
3	0.00 %	301814
4	0.00 %	545818
5	0.00 %	195514
6	0.00 %	202180
7	0.00 %	990029

Customer PO NO клиента
был найден нечетким
поиском

Система получает NO
клиента, это 83.33% всего
ключевого слова

Fir

Document Viewer

Classification result: Oki

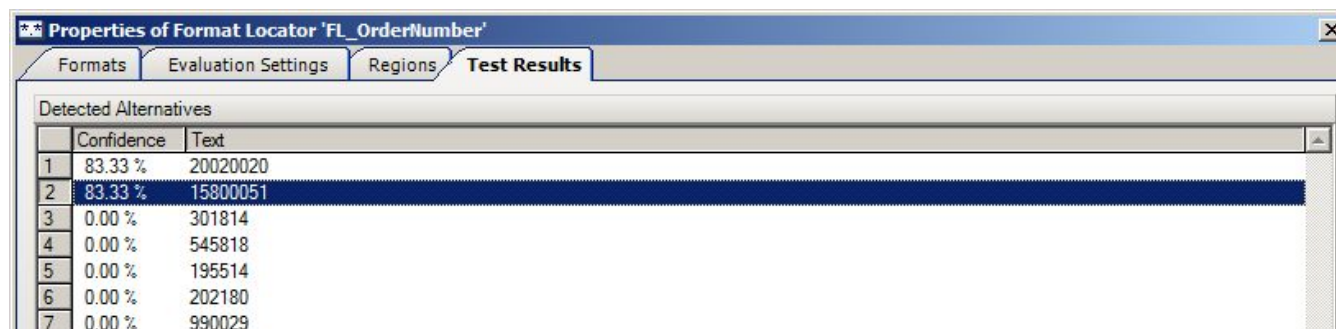
IRVINE CA 92618-3603 IR

CUSTOMER P.O. NO. CUSTOMER NO. (83.33%) 20020020 83.33%		TERMS NET THIRTY DAYS	F.O.B FOB-SHIPING. PO
ITEM NO. 1.004	PART NUMBER MSM99Q029-001LS 0.00%	CUSTOMER PART NO. CUSTOMER NO. (83.33%) 15800051-000 83.33%	

00000201.xdc 300dpi / 300dpi 6 / 82 1 / 1 TIFF

CUSTOMER PART NO.
CUSTOMER NO. (83.33%)
15800051-000
83.33%

Первый проект - Усовершенствования



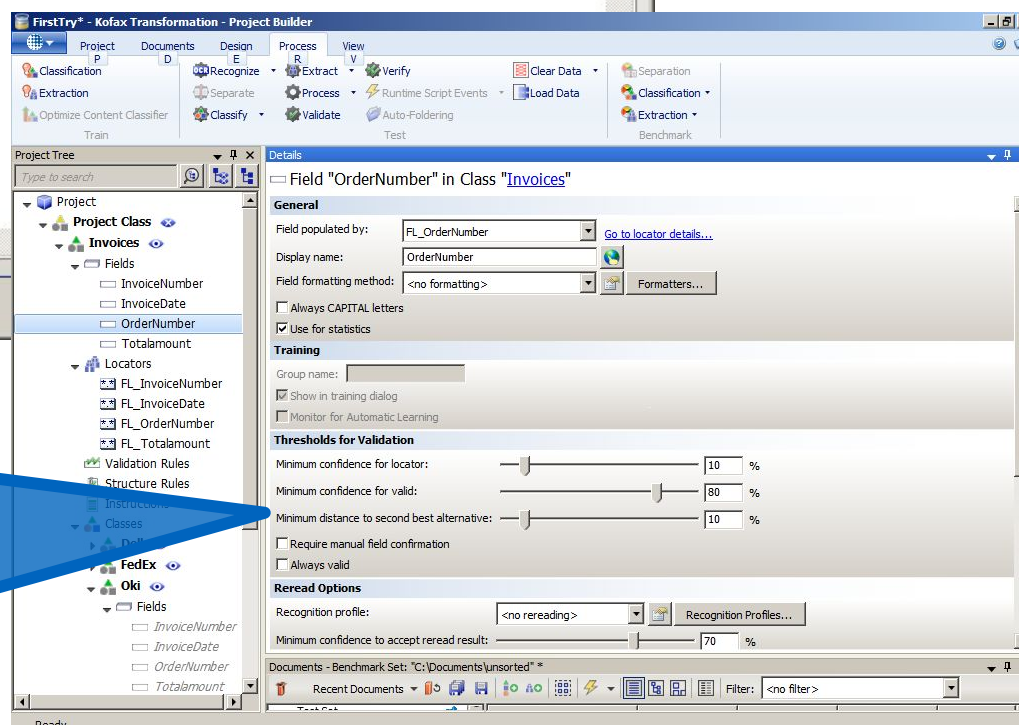
The screenshot shows the 'Properties of Format Locator' window with the 'Test Results' tab selected. It displays a table of detected alternatives for the 'FL_OrderNumber' locator. The table has two columns: 'Confidence' and 'Text'. The first two rows show a confidence of 83.33%, while the remaining rows show 0.00%.

	Confidence	Text
1	83.33 %	20020020
2	83.33 %	15800051
3	0.00 %	301814
4	0.00 %	545818
5	0.00 %	195514
6	0.00 %	202180
7	0.00 %	990029

Минимальное
расстояние - 10%

В нашем случае
расстояние
составляет 83.33 -
83.33 = 0%

Состояние поля
становится красным



Первый проект - Усовершенствования

Extraction Benchmark - Extraction Benchmark (All Classes)

Benchmark Comparison

Legend

A Correct valid fields

Summary

	All Fields	InvoiceNumber
A	65.85%	93.90%
B	29.88%	3.66%
C	3.96%	1.22%
D	0.30%	1.22%

Details - Sorted by Column Content

Filename	Class	A	B	C	D	InvoiceNumber	InvoiceDate	OrderNumber	Totalamount
000001FC.xd	Ok	50.00%	0.00%	50.00%	0.00%	-/- [310264 RI]	05/21/2003	-/- [20030032]	21553.42
000001FD.xd	Ok	50.00%	0.00%	25.00%	25.00%	307707RI [307707 RI]	02/17/2003	1580005	-/- [21590.66]
000001FE.xd	Ok	100.00%	0.00%	0.00%	0.00%	307243RI	01/29/2003	20020280	21540.00
000001FF.xd	Ok	100.00%	0.00%	0.00%	0.00%	306091RI	12/09/2002	15800051	21540.00
00000200.xd	Ok	75.00%	25.00%	0.00%	0.00%	303947RI	09/27/2002	15800051	21540.00
00000201.xd	Ok	50.00%	25.00%	25.00%	0.00%	301814RI	07/25/2002	20020020	-/- [0.00]
00000202.xd	Ok	100.00%	0.00%	0.00%	0.00%	300603RI	06/19/2002	15800051	21555.74
0000024F.xd	Dell	75.00%	25.00%	0.00%	0.00%	196328182	12/26/2002	20026306	2038.66
00000250.xd	Dell	75.00%	25.00%	0.00%	0.00%	347784804	05/18/2003	20036121	1702.46
00000251.xd	Dell	75.00%	25.00%	0.00%	0.00%	345004270	05/08/2003	20036114	1404.28
00000252.xd	Dell	75.00%	25.00%	0.00%	0.00%	344964623	05/09/2003	20036113	1318.86
00000253.xd	Dell	75.00%	25.00%	0.00%	0.00%	341350479	05/08/2003	20036106	1053.80

Progress

Extraction complete

100 %

Processed 82 documents

Start Stop

Save... Export... Help

Elapsed Time: 24 seconds

Давайте начнем с легкого. У нас есть много синих документов для счетов-фактур Dell. Есть возможность для улучшения. Дважды щелкните по первому Dell, чтобы открыть средство просмотра документов и закрыть Эталонный тест извлечения

Первый проект - Усовершенствования

FirstTry* - Kofax Transformation - Project Builder

Project Documents Design Process View

Classification Extraction Optimize Content Classifier Train

Recognize Extract Process Classify Validate

Document Viewer

Classification result: Dell

Project Tree

Type to search

Invoices

Fields

InvoiceNumber

InvoiceDate

OrderNumber

Totalamount

Locators

FL_InvoiceNumber

FL_InvoiceDate

FL_OrderNumber

FL_Totalamount

Validation Rules

Structure Rules

Instructions

Classes

Dell

FedEx

Oki

Fields

InvoiceNu

InvoiceDa

OrderNum

Totalamou

Locators

FL_Invoice

Properties of Format Locator 'FL_OrderNumber'

Formats Evaluation Settings Regions Test Results

Use results from locator: FL_InvoiceNumber

Use OCR substitution

Define OCR Substitution...

Formats

\d{6}

Format	Description	Star	5%	Ignore Characters
✓ \d{6}	<Description>	☐	☑	☑
✓ \d{8}	<Description>	☐	☑	☐

Test value:

Close Test Help

Двойной клик на локатор FL_OrderNumber

Тестируем

Первый проект - Усовершенствования

The screenshot displays the Kofax Transformation Project Builder interface. The main window shows a document titled "This is your INVOICE" with handwritten notes. The "Classification result" is "Dell". The "Purchase Order" field is highlighted, and the "Order Number" field is also highlighted. A blue callout box points to the "Order Number" field, stating: "Вторая альтернатива имеет ключевые слова к северо-западу от ее местоположения".

The "Properties of Format Locator 'FL_OrderNumber'" window is open, showing the "Test Results" tab. It lists "Detected Alternatives" with columns for "Confidence" and "Text".

	Confidence	Text
1	100.00 %	20026306
2	100.00 %	19632818
3	100.00 %	20026306
4	100.00 %	19632818
5	0.00 %	261680
6	0.00 %	00031819
7	0.00 %	19632818
8	0.00 %	42050342
9	0.00 %	00031819
10	0.00 %	00024020
11	0.00 %	20026306
12	0.00 %	20026283
13	0.00 %	19632818
14	0.00 %	00031819

A blue callout box points to the first alternative (row 1), stating: "Альтернатива 1 - правильный номер счета-фактуры".

Первый проект - Усовершенствования

The screenshot displays the Kofax Transformation Project Builder interface. The 'Document Viewer' window shows a scanned invoice with the classification result 'Dell'. The 'Properties of Format Locator 'FL_OrderNumber'' window is open, showing the 'Evaluation Settings' tab. A blue callout box with a pointer highlights the 'Relation' column in the table, specifically pointing to the 'W' relation for the 'Purchase Order' keyword.

Document Viewer:

Classification result: Dell

This is your INVOICE

Customer Number: 000018122
Purchase Order: 20026306
Order Number: 196328182
Order Date: 12/23/02

05 01 0 01 01 N

Properties of Format Locator 'FL_OrderNumber' - Evaluation Settings:

Keywords: Ignore these characters: [! / () - . : , ;]

Keyword: Purchase Order
☐ Match each word exactly (not fuzzy)
☒ Match all words as a phrase
Weight: 100
Distance: Near

Keyword in relation to match:

NW	N	NE
W		E
SW	S	SE

Table:

Use	D	Keyword	Relation	Distance	Weight	Search Exact	Whole Phrase
<input checked="" type="checkbox"/>		Customer PO NO	W, N, NW	Near	100	<input type="checkbox"/>	<input checked="" type="checkbox"/>
<input checked="" type="checkbox"/>		P.O. Number	W, N, NW	Near	100	<input type="checkbox"/>	<input checked="" type="checkbox"/>
<input checked="" type="checkbox"/>		Purchase Order	W	Near	100	<input type="checkbox"/>	<input checked="" type="checkbox"/>

Первый проект - Усовершенствования

The screenshot displays the Kofax Transformation Project Builder interface. The main window shows a document with extracted fields: Customer Number, Purchase Order, Order Number, Order Date, Invoice Number, Invoice Date, Payment Term, Shipped Via, Waybill Number, and SHIP TO. The Purchase Order and Order Number fields are highlighted with orange boxes. The Order Number field is also highlighted with a green box. The Order Date is 12/23/02. The Invoice Number is 1470. The Invoice Date is 05 01 O 01 01 N. The Payment Term is 00031819. The Shipped Via is 20026306. The Waybill Number is 196328182. The SHIP TO field is empty.

The Project Tree on the left shows the structure of the project, including Invoices, Fields, Locators, Validation Rules, Structure Rules, Instructions, Classes, and Locators. The Properties of Format Locator 'FL_OrderNumber' window is open, showing the detected alternatives for the Order Number field. The table below shows the detected alternatives:

	Confidence	Text
1	100.00 %	20026306
2	100.00 %	20026306
3	0.00 %	261680
4	0.00 %	00031819
5	0.00 %	19632818
6	0.00 %	19632818
7	0.00 %	42050342
8	0.00 %	00031819
9	0.00 %	00024020
10	0.00 %	20026306
11	0.00 %	20026283
12	0.00 %	19632818
13	0.00 %	00031819
14	0.00 %	19632818

The Properties of Format Locator 'FL_OrderNumber' window also shows the Settings tab, where the Definition Class is set to 'OrderNumber' and the Locator Method is set to 'The locator is a'.

Новый тест показывает, что только PONumber получает уверенность в 100%.

Поскольку номер заказа печатается дважды в документе, вы видите первые два результата с одинаковым значением

Первый проект - Усовершенствования

Запускаем эталонный тест
извлечения

Номер заказа на поставку для
Dell оптимизирован!

Прокрутите вниз, чтобы найти
следующие проблемы, не
беспокойтесь о
незначительных вещах в
первом запуске

The screenshot displays the Kofax Transformation Project Builder interface. The main window shows the 'Extraction Benchmark - Extraction Benchmark (All Classes)' results. The 'Summary' table provides an overview of the test results for various fields.

	All Fields	InvoiceNumber	InvoiceDate	OrderNumber	Totalamount
A	73.48%	93.90%	39.02%	71.95%	89.02%
B	22.26%	3.66%	60.98%	21.95%	2.44%
C	3.96%	1.22%	0.00%	6.10%	8.54%
D	0.30%	1.22%	0.00%	0.00%	0.00%

The 'Details - Sorted by Column Content' table provides a more granular view of the test results, including the filename, class, and specific field values.

Filename	Class	A	B	C	D	InvoiceNumber	InvoiceDate	OrderNumber	Totalamount
000001FC.xd	Oki	50.00%	0.00%	50.00%	0.00%	310264 RI	05/21/2003	1553.42	
000001FD.xd	Oki	50.00%	0.00%	25.00%	25.00%	307707 RI	02/17/2003	1553.42	
000001FE.xd	Oki	100.00%	0.00%	0.00%	0.00%	307243 RI	01/29/2003	21540.00	
000001FF.xd	Oki	100.00%	0.00%	0.00%	0.00%	306091 RI	12/09/2002	21540.00	
00000200.xd	Oki	75.00%	25.00%	0.00%	0.00%	303947 RI	09/27/2002	21540.00	
00000201.xd	Oki	50.00%	25.00%	25.00%	0.00%	301814 RI	07/25/2002	21555.74	
00000202.xd	Oki	100.00%	0.00%	0.00%	0.00%	300603 RI	06/19/2002	21555.74	
0000024F.xd	Dell	100.00%	0.00%	0.00%	0.00%	196328182	12/26/2002	2038.66	
00000250.xd	Dell	100.00%	0.00%	0.00%	0.00%	347784804	05/18/2003	1702.46	
00000251.xd	Dell	100.00%	0.00%	0.00%	0.00%	345004270	05/08/2003	1404.28	
00000252.xd	Dell	100.00%	0.00%	0.00%	0.00%	344964623	05/09/2003	1318.86	
00000253.xd	Dell	100.00%	0.00%	0.00%	0.00%	341350479	05/08/2003	1053.80	
00000254.xd	Dell	100.00%	0.00%	0.00%	0.00%	333323807	04/27/2003	2523.52	

The 'Progress' section shows that the extraction is complete, with 82 documents processed. The elapsed time is 25 seconds.

Первый проект - Усовершенствования

FirstTry* - Kofax Transformation - Project Builder

Project Documents Design Process View

Classification Extraction Optimize Content Classifier Train

Project Tree

Type to search

Invoices

Fields

InvoiceNumber InvoiceDate OrderNumber Totalamount

Locators

FL_Invoice1 FL_Invoice2 FL_OrderNumber FL_Totalamount

Validation Rule Structure Rule Instructions

Classes

Dell FedEx Oki

Fields

InvoiceNumber InvoiceDate OrderNumber Totalamount

Locators

FL_Invoice1 FL_Invoice2 FL_OrderNumber FL_Totalamount

Extraction Benchmark - Extraction Benchmark (All Classes)

Benchmark Comparison

Legend

A Correct valid fields B Correct invalid fields C Incorrect invalid fields D Incorrect valid fields

Summary

All Fields	InvoiceNumber	InvoiceDate	OrderNumber	Totalamount
A 73.48%	93.90%	39.02%	71.95%	89.02%
B 22.26%	3.66%	60.98%	21.95%	2.44%
C 3.96%	1.22%	0.00%	6.10%	8.54%
D 0.30%	1.22%	0.00%	0.00%	0.00%

Details - Sorted by Column Content

Filename	Class	A	B	C	D	InvoiceNumber	InvoiceDate	OrderNumber
FedEx1.xdc	FedEx	50.00%	50.00%	0.00%	0.00%	632325439	07/22/2004	-/-
FedEx10.xdc	FedEx	50.00%	50.00%	0.00%	0.00%	610129696	07/23/2004	-/-
FedEx11.xdc	FedEx	75.00%	25.00%	0.00%	0.00%	632324101	07/23/2004	024162
FedEx12.xdc	FedEx	50.00%	50.00%	0.00%	0.00%	632324085	07/23/2004	-/-
FedEx13.xdc	FedEx	50.00%	50.00%	0.00%	0.00%	551661301	03/30/2004	810725
FedEx14.xdc	FedEx	75.00%	25.00%	0.00%	0.00%	557638758	03/04/2004	311399
FedEx15.xdc	FedEx	50.00%	25.00%	25.00%	0.00%	476338431	07/02/2004	317936
FedEx16.xdc	FedEx	50.00%	50.00%	0.00%	0.00%	476338456	07/02/2004	-/-
FedEx17.xdc	FedEx	75.00%	25.00%	0.00%	0.00%	476338449	07/02/2004	-/-
FedEx18.xdc	FedEx	50.00%	50.00%	0.00%	0.00%	352782926	03/08/2004	-/-
FedEx19.xdc	FedEx	25.00%	25.00%	50.00%	0.00%	352783411	04/06/2004	313288 [313288212400] -/- [115.40]
FedEx2.xdc	FedEx	75.00%	25.00%	0.00%	0.00%	610129688	07/22/2004	4552 [455225212167] 92.05
FedEx20.xdc	FedEx	50.00%	25.00%	25.00%	0.00%	352783429	04/06/2004	313525 [313525212167] 319

Progress

Extraction complete

Processed 82 documents

Elapsed Time: 25 seconds

Ready

Дата счета на классе FedEx всегда синяя

1. Двойной клик по первому документу (Откроется Document viewer)
2. Сохраните эталонный тест извлечения
3. Закройте эталонный тест извлечения
4. Двойной клик по FL_InvoiceDate (откроются свойства локатора)
5. Тестируйте

Первый проект - Усовершенствования

The screenshot shows the Kofax Transformation Project Builder interface. The main window displays a FedEx invoice. The 'Properties of Format Locator 'FL_InvoiceDate'' window is open, showing the 'Test Results' tab. The 'Detected Alternatives' table lists three alternatives with their confidence levels and extracted dates.

	Confidence	Text
1	31.00 %	07/22/04
2	31.00 %	07/22/04
3	0.00 %	06/14/04

Уверенность в альтернативе слишком низкая
Давайте снова проверять ключевые слова

Первый проект - Усовершенствования

The screenshot shows the Kofax Transformation Project Builder interface. The 'Document Viewer' displays a FedEx invoice. The 'Properties of Format Locator' dialog is open, showing the 'Evaluation Settings' tab. The 'Keywords' section lists 'ORIGIN' with a weight of 31. The 'Relations' table at the bottom shows the following data:

Use	D	Keyword	Relation	Distance
<input checked="" type="checkbox"/>		Date	W, N, NW	Near
<input checked="" type="checkbox"/>		DEST	S, SE	Near
<input checked="" type="checkbox"/>		Invoice Date	W, N, NW	Near
<input checked="" type="checkbox"/>		ORIGIN	S, SW	Near

A blue callout box contains the following text:

Ключевое слово DEST и ORIGIN был установлен на юго-восток и юго-запад. Отношение OCR (xdoc) не всегда совпадает с тем, что вы видите на экране. Добавьте Юг для обоих и снова проверьте его.

Первый проект - Усовершенствования

The screenshot displays the Kofax Transformation Project Builder interface. The main window shows a scanned FedEx invoice. The 'Classification result' is 'FedEx'. The 'Document Viewer' shows the invoice details, including the tracking number 652 325 439 and the date 07/22/04. The 'Properties of Format Locator 'FL_InvoiceDate'' dialog box is open, showing the 'Test Results' tab. The 'Detected Alternatives' table lists three options, with the first one having a 93.00% confidence.

	Confidence	Text
1	93.00 %	07/22/04
2	31.00 %	07/22/04
3	0.00 %	06/14/04

Идеально! 😊

Первый проект - Усовершенствования

Запускаем эталонный тест
извлечения

Посмотрите на
сводку результатов
тестирования:

88.72 %

являются
правильными и
пропущены!

The screenshot displays the 'Extraction Benchmark - Extraction Benchmark (All Classes)' window. It includes a 'Legend' section with four categories: A (Correct valid fields), B (Correct invalid fields), C (Incorrect invalid fields), and D (Incorrect valid fields). Below this is a 'Summary' table showing overall statistics for all fields and individual fields. The 'Details - Sorted by Column Content' table provides a row-by-row breakdown of the extraction results for various FedEx documents. The 'Progress' section at the bottom indicates that the extraction is complete and 82 documents have been processed.

	All Fields	InvoiceNumber	InvoiceDate	OrderNumber	Totalamount
A	88.72%	93.90%	100.00%	71.95%	89.02%
B	7.01%	3.66%	0.00%	21.95%	2.44%
C	3.96%	1.22%	0.00%	6.10%	8.54%
D	0.30%	1.22%	0.00%	0.00%	0.00%

Filename	Class	A	B	C	D	InvoiceNumber	InvoiceDate	OrderNumber	Totalamount
FedEx34.xdc	FedEx	100.00%	0.00%	0.00%	0.00%	425772185	11/24/2003	296175	52.52
FedEx35.xdc	FedEx	75.00%	25.00%	0.00%	0.00%	426120293	11/17/2003	-/-	1762.59
FedEx36.xdc	FedEx	75.00%	25.00%	0.00%	0.00%	426120335	11/17/2003	-/-	721.08
FedEx37.xdc	FedEx	75.00%	25.00%	0.00%	0.00%	426120343	11/17/2003	-/-	1508.54
FedEx38.xdc	FedEx	100.00%	0.00%	0.00%	0.00%	426120384	11/17/2003	299013	436.76
FedEx39.xdc	FedEx	100.00%	0.00%	0.00%	0.00%	426120525	11/19/2003	299207	158.44
FedEx4.xdc	FedEx	100.00%	0.00%	0.00%	0.00%	570874695	07/22/2004	454673	181.66
FedEx40.xdc	FedEx	75.00%	25.00%	0.00%	0.00%	476333010	12/01/2003	-/-	1062.69
FedEx41.xdc	FedEx	75.00%	25.00%	0.00%	0.00%	476334133	11/25/2003	-/-	79.23
FedEx42.xdc	FedEx	75.00%	25.00%	0.00%	0.00%	476334158	11/26/2003	-/-	167.46
FedEx43.xdc	FedEx	75.00%	25.00%	0.00%	0.00%	476334232	11/25/2003	-/-	473.60
FedEx44.xdc	FedEx	100.00%	0.00%	0.00%	0.00%	476335130	12/09/2003	299627	451.74
FedEx45.xdc	FedEx	100.00%	0.00%	0.00%	0.00%	632324093	07/23/2004	033229	81.69

Progress: Extraction complete. 100 %

Processed 82 documents

Elapsed Time: 25 seconds

Первый проект - Усовершенствования

В некоторых документах нет номера заказа (и это правильно).
НО ни один номер заказа не означает уверенности, поэтому он останется красным, пока не будет использовано правило проверки.

Project Tree:

- Invoices
 - Fields
 - InvoiceNumber
 - InvoiceDate
 - OrderNumber
 - Totalamount
 - Locators
 - FL_InvoiceNumber
 - FL_InvoiceDate
 - FL_OrderNumber
 - FL_Totalamount
 - Validation Rules
 - Structure Rules
 - Instructions
 - Classes
 - Dell
 - FedEx
 - Oki
 - Fields
 - InvoiceNumber
 - InvoiceDate
 - OrderNumber
 - Totalamount
 - Locators
 - FL_InvoiceNumber

Extraction Benchmark - Extraction Benchmark (All Classes)

Legend:

- A Correct valid fields
- B Correct invalid fields
- C Incorrect invalid fields
- D Incorrect valid fields

Summary:

	All Fields	InvoiceNumber	InvoiceDate	OrderNumber	Totalamount
A	88.72%	93.90%	100.00%	71.95%	89.02%
B	7.01%	3.66%	0.00%	21.95%	2.44%
C	3.96%	1.22%	0.00%	6.10%	8.54%
D	0.30%	1.22%	0.00%	0.00%	0.00%

Details - Sorted by Column Content

Filename	Class	A	B	C	D	InvoiceNumber	InvoiceDate	OrderNumber	Totalamount
FedEx34.xdc	FedEx	100.00%	0.00%	0.00%	0.00%	425772185	11/24/2003	296175	52.52
FedEx35.xdc	FedEx	75.00%	25.00%	0.00%	0.00%	426120293	11/17/2003	-/-	1762.59
FedEx36.xdc	FedEx	75.00%	25.00%	0.00%	0.00%	426120335	11/17/2003	-/-	721.08
FedEx37.xdc	FedEx	75.00%	25.00%	0.00%	0.00%	426120343	11/17/2003	-/-	1508.54
FedEx38.xdc	FedEx	100.00%	0.00%	0.00%	0.00%	426120384	11/17/2003	299013	436.76
FedEx39.xdc	FedEx	100.00%	0.00%	0.00%	0.00%	426120525	11/19/2003	299207	158.44
FedEx4.xdc	FedEx	100.00%	0.00%	0.00%	0.00%	570874695	07/22/2004	454673	181.66
FedEx40.xdc	FedEx	75.00%	25.00%	0.00%	0.00%	476333010	12/01/2003	-/-	1062.69
FedEx41.xdc	FedEx	75.00%	25.00%	0.00%	0.00%	476334133	11/25/2003	-/-	79.23
FedEx42.xdc	FedEx	75.00%	25.00%	0.00%	0.00%	476334158	11/26/2003	-/-	167.46
FedEx43.xdc	FedEx	75.00%	25.00%	0.00%	0.00%	476334232	11/25/2003	-/-	473.60
FedEx44.xdc	FedEx	100.00%	0.00%	0.00%	0.00%	476335130	12/09/2003	299627	451.74
FedEx45.xdc	FedEx	100.00%	0.00%	0.00%	0.00%	632324093	07/23/2004	033229	81.69

Progress:

Extraction complete

Processed 82 documents

Elapsed Time: 25 seconds

Buttons: Start, Stop, Save..., Export..., Help

Первый проект - Усовершенствования

FirstTry* - Kofax Transformation - Project Builder

Project Tree

- Invoices
 - Fields
 - InvoiceNumber
 - InvoiceDate
 - OrderNumber
 - Totalamount
 - Locators
 - FL_InvoiceNumber
 - FL_InvoiceDate**
 - FL_OrderNumber
 - FL_Totalamount
 - Validation Rules
 - Structure Rules
 - Instructions
 - Classes
 - Dell
 - FedEx
 - Oki

Details

Locator "FL_InvoiceDate" in Class "Invoices"

Settings

Definition Class: Invoices

Locator Method: Format Locator

The locator is assigned to the following fields:

[InvoiceDate](#)

Documents - Benchmark Set: "C:\Documents\unsorted"

Filename	Classification Re:	Confidence	Assigned Class
000001FC.xdc*	Oki	100.0%	
000001FD.xdc*	Oki	100.0%	
000001FE.xdc*	Oki	91.5%	
000001FF.xdc*	Oki	89.2%	
00000200.xdc*	Oki	88.8%	
00000201.xdc*	Oki	100.0%	
00000202.xdc*	Oki	84.1%	
0000024F.xdc*	Dell	86.6%	
00000250.xdc*	Dell	84.7%	
00000251.xdc*	Dell	78.1%	
00000252.xdc*	Dell	89.4%	
00000253.xdc*	Dell	86.5%	
00000254.xdc*	Dell	75.5%	
00000255.xdc*	Dell	100.0%	
00000256.xdc*	Dell	91.3%	

Instructional Text:

Снова обработайте документы (щелкните по стрелке рядом с «Обработка и обработка всей партии»)

Нажмите «Проверить» и попробуйте проверить все документы.

Проверьте время, необходимое для этого.

Первый проект - Усовершенствования

Test Validation

Batch Edit View Folder Document Page Field Options Help

Batch Content

Root Folder

- 000001FC.xdc Page 1
- 000001FD.xdc Page 1
- 000001FE.xdc Page 1
- 000001FF.xdc Page 1
- 00000200.xdc Page 1
- 00000201.xdc Page 1
- 00000202.xdc Page 1
- 00000204F.xdc Page 1
- 00000250.xdc

Classification Result

The document was classified to the following Oki

3 fields valid, 1 field invalid (0 fields invisible, 0 fields read-only)

InvoiceNumber	310264 RI	310264 RI
InvoiceDate	05/21/03	05/21/2003
OrderNumber	20030032	20030032
Totalamount	21,553.42	21553.42

Ok Semiconductor

700 North Alhambra Avenue
Sunnyvale, CA 94085-5000
Tel: 408-720-1000
Fax: 408-720-1001

202180

RECEIVED

ENTERED MAY 29 2003

Subtotal: 21,553.42 USD
Sales Taxes: USD
Amount Due: 21,993.42 USD

Page: -/1 Invalid Docs: 33

Если вы подтвердите значение, система перейдет в следующее недопустимое (красное) поле и пропустит все зеленые поля.

Даже если следующее недопустимое поле содержит более 100 документов, система пропускает все документы, находящиеся между ними

Сначала вам потребовалось более
15 минут
для валидации пакета

Теперь это занимает
2 минуты!

13 минут сохранено!



Первый проект - Усовершенствования

- ◆ Задание
 - ◆ Используйте эталонный тест извлечения для поиска ошибок
 - ◆ Усовершенствуйте локаторы
 - ◆ Перезапустите эталонный тест извлечения
 - ◆ Посмотрите результаты

Первый проект - Зеленые и красные поля (Правила валидации)

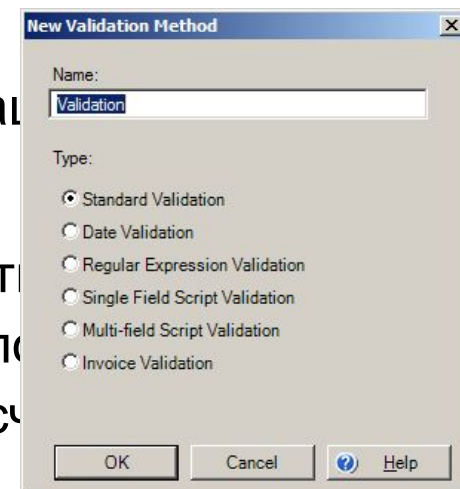
Как сделать значения зелеными или красными автоматически

Первый проект - Зеленые и красные поля (Правила валидации)

- ◆ Поля могут быть валидными и невалидными
Как уже обсуждалось в улучшениях
- ◆ Почему поля зеленые или красные?
 - ◆ Сначала проверьте уверенность в извлечении
 - ◆ Во вторых проверьте расстояние между первым и вторым результатом
 - ◆ В третьих проверьте форматирование
 - ◆ В четвертых проверьте правило валидации
- ◆ Что такое правило валидации?
 - ◆ Давайте обсудим...

Первый проект - Зеленые и красные поля (Правила валидации)

- ◆ Метод валидации это настройка проекта и может быть использована
 - ◆ из локатора для поиска значения (об этом мы поговорим в более позднем модуле)
 - ◆ также он может быть связан с полем, чтобы изменить состояние поля на красный или зеленый.
 - ◆ Это называется ПРАВИЛО ВАЛИДАЦИИ
- ◆ Использовать существующие знания для валидации
- ◆ Доступными способами являются: Стандартное регулярное выражение, Проверка на одно поле, Многопользовательский скрипт и проверка с



Первый проект - Зеленые и красные поля (Правила валидации)

Invoice Number	Keyword	Invoice Date	Keyword	PONumber	Keyword	Total amount	Keyword
\d{9}	Invoice Number	US DATE	Invoice Date	\d{8}	Purchase Order	US Amount	Invoice Total
\d{9}	Freight Bill Number	US DATE	Date	\d{6} \d{8} \d{12} A\d{4}l	P.O. Number	US Amount	Please pay this amount
\d{6}[A-Z]{2}	Invoice NO	US DATE	Invoice Date	\d{8}	Customer PO NO	US Amount	Amount Due



Notes:

Fedex: Poor Keyword for Invoice Date, but it is below Invoice number
PO Number can be empty

Oki: Poor Keyword for Total Amount, because it has white text on black background
Total Amount is always below Sales Taxes

Первый проект - Зеленые и красные поля (Правила валидации)

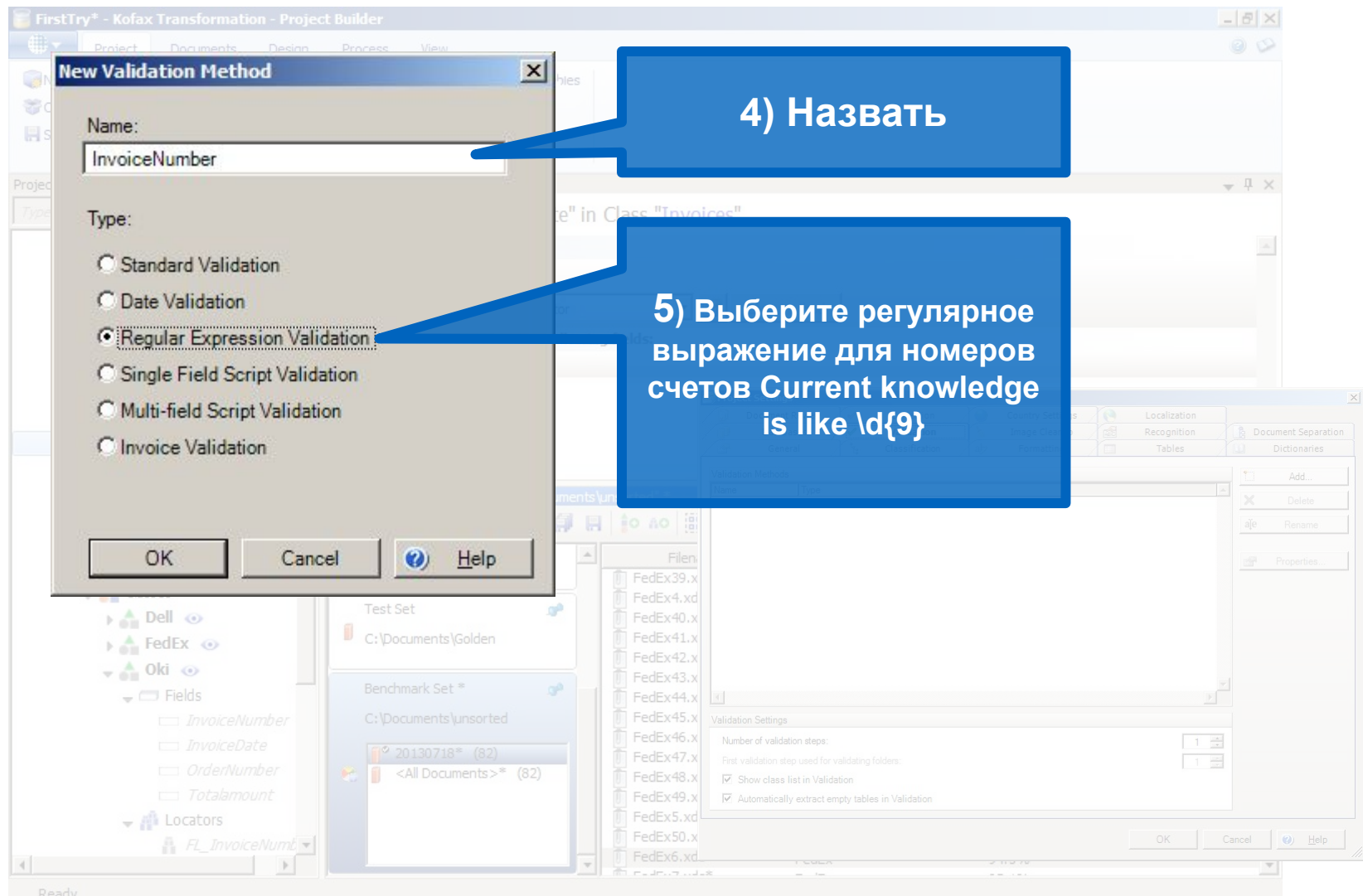
The screenshot shows the 'FirstTry* - Kofax Transformation - Project Builder' application. The 'Project' tab is active, showing a 'Project Tree' on the left and a 'Project Settings' window on the right. The 'Project Tree' lists various components: Fields (InvoiceNumber, OrderNumber, Totalamount), Validation Rules, Structure Rules, Instructions, Classes (Dell, FedEx, Oki), and Locators (FL_InvoiceDate, FL_OrderNumber, FL_Totalamount). The 'Project Settings' window is open, showing the 'Validation' tab. The 'Validation Methods' section is visible, with a list of methods and an 'Add...' button. The 'Validation Settings' section shows options for 'Number of validation steps' and 'Show class list in Validation'.

1) Нажмите на Project setting (часть ленты в Project Tab)

2) Изменить на валидацию

3) Добавить новый метод валидации

Первый проект - Зеленые и красные поля (Правила валидации)



Первый проект - Зеленые и красные поля (Правила валидации)

6) Добавьте свое регулярное выражение (аналогично локатору формата)

Список всех добавленных форматов

Format	Description	Ignore Case	Ignore Blanks	Ignore Chars
*\d{6}[A-Z]{2}		✓	✓	
*\d{9}		✗	✓	

Проверьте регулярные выражения

Testing
Text input string: ✓

OK Cancel Help

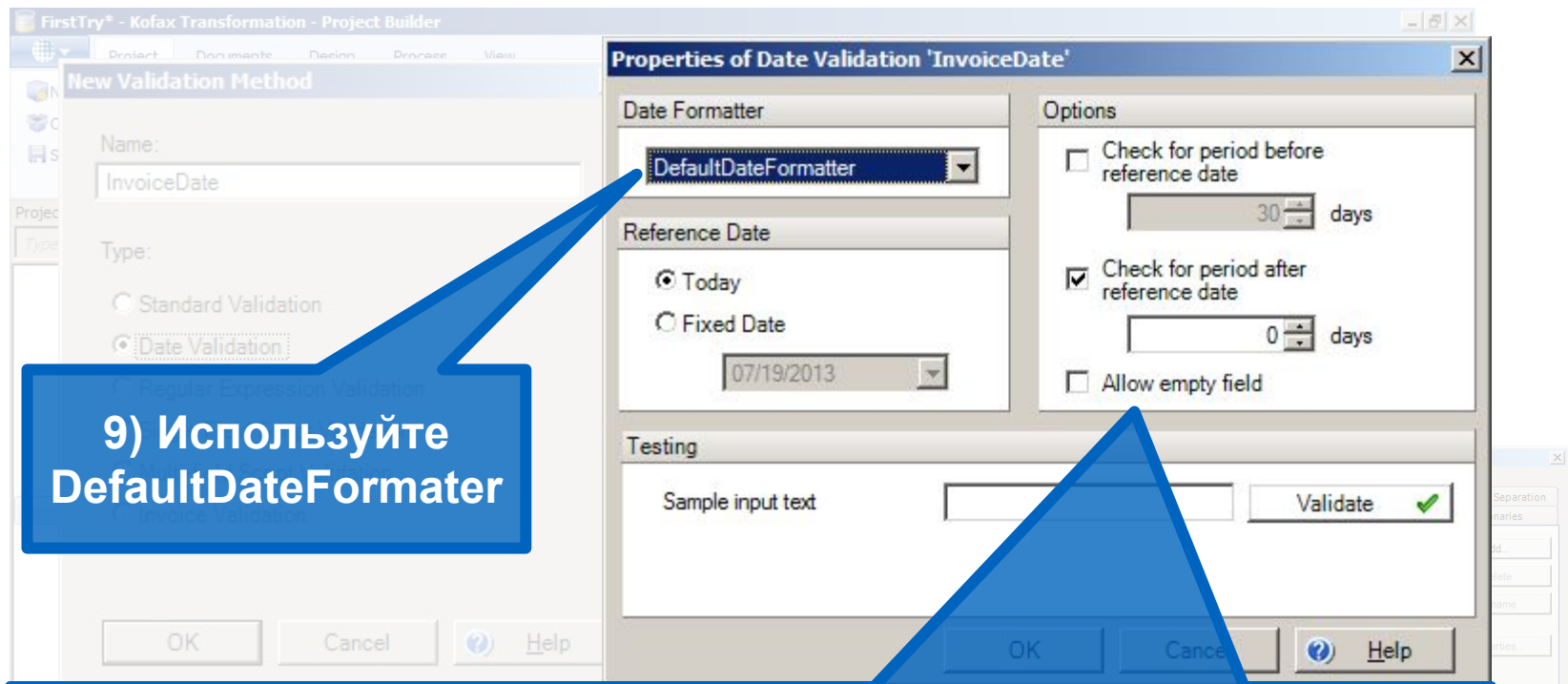
Первый проект - Зеленые и красные поля (Правила валидации)

The screenshot displays the 'FirstTry* - Kofax Transformation - Project Builder' application. A 'New Validation Method' dialog is open, showing the 'Name' field set to 'InvoiceDate' and the 'Type' field with 'Date Validation' selected. A blue callout box points to the 'Date Validation' radio button with the text: '8) Выбрав Date Validation для дат счет-фактуры'. Another blue callout box points to the 'Name' field with the text: '7) Создайте другой Метод Валидации для дат счет фактуры'. In the background, the 'Project Settings' dialog is visible, showing the 'Validation' tab with a table of validation methods. The table has columns 'Name' and 'Type'. Below the table, the 'Validation Settings' section shows 'Number of validation steps' set to 1, 'First validation step used for validating folders' set to 1, and checkboxes for 'Show class list in Validation' and 'Automatically extract empty tables in Validation' both checked. The main interface shows a project tree on the left with 'Dell', 'FedEx', and 'Oki' folders, and a 'Test Set' on the right with 'C:\Documents\Golden' and 'C:\Documents\unsorted' folders. A file list on the right shows various FedEx files.

7) Создайте другой Метод Валидации для дат счет фактуры

8) Выбрав Date Validation для дат счет-фактуры

Первый проект - Зеленые и красные поля (Правила валидации)



10) Значение даты счета может быть старой, но не в будущем

- Снимите first option (Выберите period before ref date)
- Check second (Check for period after ref date)
- Снимите «allow empty fields»

Первый проект - Зеленые и красные поля (Правила валидации)

11) Создать метод валидации для номеров заказов

12) Добавьте регулярные выражения

Отметьте флажки пустыми полями, потому что у нас есть документы без PONumber

Properties of Regular Expression Validation 'Ordernumber'

Formats

Format:

☐ High performance

Description:

Error description:

Ignore characters:

☐ Ignore case

☒ Ignore blanks

Format	Description	Ignore Case	Ignore Blanks	Ignore Chars
*\d{6}		X	X	
*\d{8}		X	X	
*\d{12}		X	✓	
*\A\d{4}I		✓	X	

☒ Allow empty field

Testing

Text input string:

Validate ☒

OK Cancel Help

Первый проект - Зеленые и красные поля (Правила валидации)



Все методы валидации созданы

Давайте привяжем их к полям

Первый проект - Зеленые и красные поля (Правила валидации)

1) Щелкните правой кнопкой мыши на Правилах валидации (ниже счетов-фактур класса)

2) Добавить правило проверки отдельного поля

3) Назовите его InvoiceNumber

Filename	Classification Res	Confidence	Assigned Class
FedEx39.xdc*	FedEx	83.8%	
FedEx4.xdc*	FedEx	96.3%	
FedEx40.xdc*	FedEx	85.2%	
FedEx41.xdc*	FedEx	83.8%	
FedEx42.xdc*	FedEx	83.7%	
FedEx43.xdc*	FedEx	83.0%	
FedEx44.xdc*	FedEx	84.1%	
FedEx45.xdc*	FedEx	98.2%	
FedEx46.xdc*	FedEx	97.2%	
FedEx47.xdc*	FedEx	97.7%	
FedEx48.xdc*	FedEx	96.7%	
FedEx49.xdc*	FedEx	96.5%	
FedEx5.xdc*	FedEx	96.2%	
FedEx50.xdc*	FedEx	95.6%	
FedEx6.xdc*	FedEx	94.5%	

Первый проект - Зеленые и красные поля (Правила валидации)

The screenshot displays the 'FirstTry* - Kofax Transformation - Project Builder' application. The 'Project Tree' on the left shows a hierarchy with 'Invoices' > 'Fields' > 'InvoiceNumber' selected. The 'Details' pane on the right shows the configuration for a 'Single Field Validation Rule "InvoiceNumber" in Class "Invoice"'. The 'Field' dropdown is set to 'InvoiceNumber'. Under 'Validation Methods', a yellow star icon is highlighted. A 'Validation Method' dialog box is open in the foreground, showing 'InvoiceNumber' selected in the 'Please select a validation method:' list. At the bottom right, a table shows document processing results.

4) Выберите поле InvoiceNumber

5) Добавьте метод валидации

6) Добавьте метод InvoiceNumber

Filename	Classification Res	Confidence	Assigned Class
FedEx39.xdc*	FedEx	83.8%	
FedEx	FedEx	95.2%	
FedEx	FedEx	83.8%	
FedEx	FedEx	83.0%	
FedEx	FedEx	84.1%	
FedEx	FedEx	98.2%	
FedEx	FedEx	97.2%	
FedEx	FedEx	97.7%	
FedEx	FedEx	96.7%	
FedEx	FedEx	96.5%	
FedEx	FedEx	96.2%	
FedEx	FedEx	95.6%	
FedEx	FedEx	94.5%	

Первый проект - Зеленые и красные поля (Правила валидации)

The screenshot shows the FirstTry Kofax Transformation Project Builder interface. The 'Project Tree' on the left shows the 'Invoices' class with fields 'InvoiceNumber', 'InvoiceDate', 'OrderNumber', and 'Totalamount'. The 'Details' pane shows the 'Single Field Validation Rule "InvoiceDate" in Class "Invoices"'. The 'General' tab is selected, showing the 'Field' as 'InvoiceDate'. The 'Validation Methods' section shows a list of methods, with 'InvoiceDate' selected. A blue callout box points to the 'Validation Method' dialog box, which is open. The dialog box has a title bar 'Validation Method' and a message 'Please select a validation method:'. The 'InvoiceDate' method is selected in the list. The 'OK' button is highlighted.

Создайте новое правило для invoice date

Validation Method

Please select a validation method:

InvoiceDate

OK Cancel ? Help

Documents - Benchmark Set: "C:\Documents\unsorted" *

Recent Documents

C:\Documents\sorted

Test Set

C:\Documents\Golden

Benchmark Set *

C:\Documents\unsorted

20130718* (82)

<All Documents>* (82)

Filename	Classification Res	Confidence	Assigned Class
FedEx39.xdc*	FedEx	83.8%	
FedEx4.xdc*	FedEx	96.3%	
FedEx40.xdc*	FedEx	85.2%	
FedEx41.xdc*	FedEx	83.8%	
FedEx42.xdc*	FedEx	83.7%	
FedEx43.xdc*	FedEx	83.0%	
FedEx44.xdc*	FedEx	84.1%	
FedEx45.xdc*	FedEx	98.2%	
FedEx46.xdc*	FedEx	97.2%	
FedEx47.xdc*	FedEx	97.7%	
FedEx48.xdc*	FedEx	96.7%	
FedEx49.xdc*	FedEx	96.5%	
FedEx5.xdc*	FedEx	96.2%	
FedEx50.xdc*	FedEx	95.6%	
FedEx6.xdc*	FedEx	94.5%	

Первый проект - Зеленые и красные поля (Правила валидации)

FirstTry* - Kofax Transformation - Project Builder

Project | Documents | Design | Process | View

New Project | Save Project As | Validate Project | Script Variables
Open Project | Close Project | License Utility
Save Project | Project Settings | Script Resources

File | Configuration

Project Tree

Type to search

- Invoices
 - Fields
 - InvoiceNumber
 - InvoiceDate
 - OrderNumber
 - Totalamount
 - Locators
 - FL_InvoiceNumber
 - FL_InvoiceDate
 - FL_OrderNumber
 - FL_Totalamount
 - Validation Rules
 - InvoiceNumber
 - InvoiceDate
 - PONumber
 - Structure Rules
 - Instructions
 - Classes
 - Dell
 - FedEx
 - Oki

Details

Single Field Validation Rule "PONumber" in Class "Invoices"

General

☒ Validation rule makes uncertain extraction results valid

Field: OrderNumber

Validation Methods

Selected validation methods:

OrderNumber

Documents - Benchmark Set: "C:\Documents\unsorted" *

Recent Documents

C:\Documents\sorted

Test Set

C:\Documents\Golden

Benchmark Set *

C:\Documents\unsorted

20130718* (82)

<All Documents>* (82)

File Name	Classification Result	Confidence	Assigned Class
FedEx39.xdc*	FedEx	96.3%	FedEx
FedEx40.xdc*	FedEx	96.3%	FedEx
FedEx41.xdc*	FedEx	96.3%	FedEx
FedEx42.xdc*	FedEx	96.3%	FedEx
FedEx43.xdc*	FedEx	96.3%	FedEx
FedEx44.xdc*	FedEx	96.3%	FedEx
FedEx45.xdc*	FedEx	96.3%	FedEx
FedEx46.xdc*	FedEx	96.3%	FedEx
FedEx47.xdc*	FedEx	96.3%	FedEx
FedEx48.xdc*	FedEx	96.3%	FedEx
FedEx49.xdc*	FedEx	96.3%	FedEx
FedEx50.xdc*	FedEx	96.3%	FedEx
FedEx51.xdc*	FedEx	96.3%	FedEx
FedEx52.xdc*	FedEx	96.3%	FedEx
FedEx53.xdc*	FedEx	96.3%	FedEx
FedEx54.xdc*	FedEx	96.3%	FedEx
FedEx55.xdc*	FedEx	96.3%	FedEx
FedEx56.xdc*	FedEx	96.3%	FedEx
FedEx57.xdc*	FedEx	96.3%	FedEx
FedEx58.xdc*	FedEx	96.3%	FedEx
FedEx59.xdc*	FedEx	96.3%	FedEx
FedEx60.xdc*	FedEx	96.3%	FedEx
FedEx61.xdc*	FedEx	96.3%	FedEx
FedEx62.xdc*	FedEx	96.3%	FedEx
FedEx63.xdc*	FedEx	96.3%	FedEx
FedEx64.xdc*	FedEx	96.3%	FedEx
FedEx65.xdc*	FedEx	96.3%	FedEx
FedEx66.xdc*	FedEx	96.3%	FedEx
FedEx67.xdc*	FedEx	96.3%	FedEx
FedEx68.xdc*	FedEx	96.3%	FedEx
FedEx69.xdc*	FedEx	96.3%	FedEx
FedEx70.xdc*	FedEx	96.3%	FedEx
FedEx71.xdc*	FedEx	96.3%	FedEx
FedEx72.xdc*	FedEx	96.3%	FedEx
FedEx73.xdc*	FedEx	96.3%	FedEx
FedEx74.xdc*	FedEx	96.3%	FedEx
FedEx75.xdc*	FedEx	96.3%	FedEx
FedEx76.xdc*	FedEx	96.3%	FedEx
FedEx77.xdc*	FedEx	96.3%	FedEx
FedEx78.xdc*	FedEx	96.3%	FedEx
FedEx79.xdc*	FedEx	96.3%	FedEx
FedEx80.xdc*	FedEx	96.3%	FedEx
FedEx81.xdc*	FedEx	96.3%	FedEx
FedEx82.xdc*	FedEx	96.3%	FedEx
FedEx83.xdc*	FedEx	96.3%	FedEx
FedEx84.xdc*	FedEx	96.3%	FedEx
FedEx85.xdc*	FedEx	96.3%	FedEx
FedEx86.xdc*	FedEx	96.3%	FedEx
FedEx87.xdc*	FedEx	96.3%	FedEx
FedEx88.xdc*	FedEx	96.3%	FedEx
FedEx89.xdc*	FedEx	96.3%	FedEx
FedEx90.xdc*	FedEx	96.3%	FedEx
FedEx91.xdc*	FedEx	96.3%	FedEx
FedEx92.xdc*	FedEx	96.3%	FedEx
FedEx93.xdc*	FedEx	96.3%	FedEx
FedEx94.xdc*	FedEx	96.3%	FedEx
FedEx95.xdc*	FedEx	96.3%	FedEx
FedEx96.xdc*	FedEx	96.3%	FedEx
FedEx97.xdc*	FedEx	96.3%	FedEx
FedEx98.xdc*	FedEx	96.3%	FedEx
FedEx99.xdc*	FedEx	96.3%	FedEx
FedEx100.xdc*	FedEx	96.3%	FedEx

Ready

И последняя для OrderNumber
Выберите “Validation rule makes
uncertain results valid”

Это полезно для пустых
номеров: даже если
поле пустое, наше правило будет
отображать значения зеленым
цветом

Первый проект - Зеленые и красные поля (Правила валидации)

Запустите эталонный тест извлечения

Правило валидации Order Number делает пустые поля действительными для номеров заказов. Никто не должен проверять пустые поля

Сохраните и сравните

Extraction Benchmark - Extraction Benchmark (All Classes)

Legend

- A Correct valid fields
- B Correct invalid fields
- C Incorrect invalid fields
- D Incorrect valid fields

Summary

	OrderNumber	Totalamount
A	94.21%	93.90%
B	0.00%	0.00%
C	0.00%	0.00%
D	1.83%	1.22%

Details - Sorted by Column Content

	OrderNumber	InvoiceDate	OrderNumber	Totalamount
FedEx40.xdc	425772185	11/24/2003	296175	52.52
FedEx35.xdc	426120293	11/17/2003	-/-	1762.59
FedEx32.xdc	426120293	11/17/2003	-/-	721.08
FedEx37.xdc	426120343	11/17/2003	-/-	1508.54
FedEx38.xdc	426120384	11/17/2003	299013	436.76
FedEx39.xdc	426120525	11/19/2003	299207	158.44
FedEx4.xdc	570874695	07/22/2004	454673	181.66
FedEx40.xdc	476333010	12/01/2003	-/-	1062.69
FedEx41.xdc	476334133	11/25/2003	-/-	79.23
FedEx42.xdc	476334158	11/26/2003	-/-	167.46
FedEx43.xdc	476334232	11/25/2003	-/-	473.60
FedEx44.xdc	476335130	12/09/2003	299627	451.74
FedEx45.xdc	632324093	07/23/2004	033229	81.69

Progress

Extraction complete

100 %

Start Stop

Save... Export... Help

Elapsed time: 23 seconds

Первый проект - Зеленые и красные поля (Правила валидации)

FirstTry* - Kofax Transformation - Project Builder

Extraction Benchmark - Extraction Benchmark (All Classes)

Legend

A Correct valid fields B Correct invalid fields C Incorrect invalid fields D Incorrect valid fields

Benchmark 1

Select Benchmark 1: BM_2013_7_19_14_10_2

Benchmark 1 details

Date: 2013-07-19 14:10:02

Comment:

Benchmark 2

Select Benchmark 2: BM_2013_7_19_19_24_54

Benchmark 2 details

Date: 2013-07-19 19:24:54

Comment: With validation rules

☒ Baseline

Summary Statistics Comparison

	All Fields	InvoiceNumber	InvoiceDate	OrderNumber	Totalamount
A	94.21% [51.83%]	93.90% [93.90%]	100.00% [60.98%]	93.90% [52.44%]	89.02%
B	1.52% [-27.44%]	3.66% [3.66%]	0.00% [-60.98%]	0.00% [-52.44%]	2.44%
C	2.44% [-26.22%]	1.22% [-98.78%]	0.00%	0.00% [-6.10%]	8.54%
D	1.83% [-1.83%]	1.22% [-1.22%]	0.00%	6.10% [-6.10%]	0.00%

Details Statistics Comparison

Filename	Class	A	B	C	D	InvoiceNumber	InvoiceDate	OrderNumber	Total
C:\Documents\unsorted\000001FE.xdc	Ok	100.00% [25.00%]	0.00%	0.00% [-25.00%]	0.00%	307243RI	01/29/2003	20020280	2154
C:\Documents\unsorted\000001FF.xdc	Ok	100.00% [25.00%]	0.00%	0.00% [-25.00%]	0.00%	306091RI	12/09/2002	15800051	2154
C:\Documents\unsorted\00000200.xdc	Ok	100.00% [25.00%]	0.00%	0.00% [-25.00%]	0.00%	306091RI	09/27/2002	15800051	2154
C:\Documents\unsorted\00000201.xdc	Ok	75.00% [50.00%]	0.00%	25.00%	0.00%	3018	07/25/2002	20020020	-/-
C:\Documents\unsorted\00000202.xdc	Ok	100.00% [25.00%]	0.00%	0.00% [-25.00%]	0.00%	30060	06/19/2002	15800051	2155
C:\Documents\unsorted\0000024F.xdc	Ok	100.00% [50.00%]	0.00%	25.00%	0.00%	1967782	12/26/2002	20026306	2038
C:\Documents\unsorted\00000250.xdc	Ok	100.00% [50.00%]	0.00%	25.00%	0.00%	347784804	05/18/2003	20036121	1702
C:\Documents\unsorted\00000251.xdc	Ok	100.00% [50.00%]	0.00%	25.00%	0.00%	345004370	05/08/2003	20036111	1404

Elapsed Time: 25 seconds

Save... Export... Help Stop

Демонстрация и задание

First project – Green & red fields (Validation Rules)

- ◆ Задание
 - ◆ Создайте метод валидации
 - ◆ Создайте правило валидации (Свяжите поле с методом)
 - ◆ Запустите Эталонный тест извлечения
 - ◆ Просмотрите результаты