



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

Общая сумма
всегда в
формате US
Ключевое слово:
Invoice Total

Invoice Total | \$ 2,038.66

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

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

FID Number: 74-2516905 Sales Rep: JENNER ROCHNER Customer Number: 000318193 Purchase Order: 20026306 Order Number: 196328182 Invoice Number: 196328182 Order Date: 12/23/02 Invoice Date: 12/26/02 Payment Terms: NET 30 DAYS Shipped Via: STAIR HAWK GROUND Waybill Number: V5410005034218

For Customer Service: (877) 471-3635 Fax Technical Support: (877) 471-3635 Dell Online: www.dell.com

SOLD TO: 1919 306 00224020 1 FP 0.352 01 ACCOUNTS PAYABLE KOFAX IMAGE PRODUCTS ACCOUNTS PAYABLE PO 20026306 16245 LAGUNA CANYON RD IRVINE, CA 92618

SHIP TO: JM McDONALD KOFAX IMAGE PRODUCTS PO 20026306 16245 LAGUNA CANYON RD IRVINE, CA 92618-3603

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 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/ C840 Notebooks, Factory Install	EA	0.00	0.00
1		340-3017	30GB Hard Drive, 9.5MM for Dell Latitude C800/C500/C610/C510/ C840 Notebooks, Factory Install	EA	0.00	0.00
1		420-0656	Windows XP Professional, English for Dell Latitude C840 Notebooks, Factory Install	EA	0.00	0.00
1		313-5205	Internal 56K Modem for Dell Latitude C810/C510/C840 Notebooks, Factory Install	EA	0.00	0.00
1		313-1174	8-24-10-24X DVD w/Software CD Read-Write Combo for Dell Latitude C810/C810/C400/C510/C840 Notebooks, Factory Inst	EA	0.00	0.00
1		312-0042	9-Cell Primary Battery for Dell Latitude C840 Notebooks, Factory Install	EA	0.00	0.00
1		420-0153	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 On-Site Response, Infil Year	EA	0.00	0.00
1		900-6232	*Type 3 Contract - Next Business Day Parts and Labor On-Site Response, 2YR Extended	EA	0.00	0.00
		System Service Tags	BD1M721			

RECEIVED
DEC 30 2002

36738.00 \$ 1857.00
 01-1570-500 \$ 181.66
 62038.66

Ship. &/or Handling	\$ 35.00
Subtotal	\$ 1,892.00
Taxable	\$ 1,892.00
Tax	146.66
Invoice Total	2,038.66

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 91106

Ship. &/or Handling	\$ 35.00
Subtotal	\$ 1,892.00
Taxable	\$ 1,892.00
Tax	146.66
Invoice Total	2,038.66
Balance Due	2,038.66
Am't. Enclosed	\$

00019632818200000002038660500003161935

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

PO Number
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

Tracking Number: 846167

Date: 07/22/04

Origin: AVL Dest: TW

Shipper: 21344610 COATS NORTH AMERICA / AMERICAN THREAD RD / RION NC 28752

Bill To: COATS AMERICAN / TRENDSET INC / P O BOX 1208 / MAULDIN SC 29662

Consignee: 58413913

PO Number: 58413913

BL Number: 213

Pay: 62.76

Routing: TX 7614

PCS	HM	DESCRIPTION	WT(LBS)	N M F C	CLASS	RATE	TOTAL CHARGES
18		CARTON, THREAD, SEWING, CON ON CONES, CORES OR SPOOLS SKIDS SWS STC 18 CTNS 458018 BL 00676 FUEL SURCHG LTL SHPT8 207 DISCOUNT ARFM 6000 0129433 0014 REFXF 1000 06/14/04 LS 31277 ** FAK RATES APPLIED **	213	182220-01	060	FLAT	\$ 58 00
18		TERMS PPD	213				\$ 62.76

Reliable, Responsive Regional and Interregional LTL

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

Freight Bill Number: 632 325 439

Date: 07/22/04

Please Pay This Amount: 62 76

Tracking Number: 84616794

Shipping Party: COATS AMERICAN / TRENDSET INC / P O BOX 1208 / MAULDIN

Send Payment to: FedEx Freight East, P O Box 406708, Atlanta GA 30384-8708

Direct Billing Inquiries to: P O Box 840, Harrison AR 72602-0840

Phone 1 870 741 9000 Fax 1 870 365 4354 Toll-Free 1 866 393 4585

FedEx Freight (FXFE)

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}	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

3585
PA

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. 1031K
REP. NAME KOFAX IMAGE PRODUCTS

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* 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



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 Re:	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%	
FedEx2.xdc	FedEx	84.0%	

Первый проект– Извлечение по форматам и ключевым словам

3) Клик на свойства локатора

2) Используйте Format Locator как Locator Method

Project Tree: Invoices > Locators > FL_InvoiceNumber

Details: Locator "FL_InvoiceNumber" in Class "Invoices"

Settings: Definition Class: Invoices, Locator Method: <none>

The locator is as follows:

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

Первый проект– Извлечение по форматам и ключевым словам

The screenshot displays the 'Properties of Format Locator' dialog box for 'FL_InvoiceNumber'. The 'Formats' tab is active, showing a list of regular expressions. A blue callout box with the text '4) Добавьте к нему новое регулярное выражение' points to the 'Add a new regular expression to the list' button. The background shows a project tree with 'Invoices' and 'FL_InvoiceNumber' selected, and a table of document processing results.

Document	Format	Percentage
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%
FedEx2.xdc	FedEx	84.0%

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:

File Name	Format	Percentage
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

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

Use	D	Keyword	Relation	Distance	Weight
<input checked="" type="checkbox"/>		Invoice Number	W, N, NW	Near	100

Document	Keyword	Percentage
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%
FedEx2.xdc	FedEx	84.0%

Первый проект – Извлечение по форматам и ключевым словам

Properties of Format Locator 'FL_InvoiceNumber'

Keywords

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

Keyword: INVOICE NO

Distance: Near

Use	D	Keyword	Relation	Distance	Weight	Search Back	Match Phrase
<input checked="" type="checkbox"/>		FREIGHT BILL NUM...	W, N, NW	Near		<input type="checkbox"/>	<input 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, N, NW	Near	100	<input type="checkbox"/>	<input checked="" type="checkbox"/>

Test

Теперь все
ключевые в
списке

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

Первый проект– Извлечение по форматам и ключевым словам

Результаты теста в просмотрщике

Результаты теста в свойствах локатора

12) Выберите другой документ из тестового набора и повторите попытку

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

Document Viewer: Classification result: FedEx

INVOICE

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

FEDERATED MAIL SERVICE INC. 300001294

300dpi / 300dpi 1/82 1/1 TIFF

Первый проект– Извлечение по форматам и ключевым словам

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

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

Первый проект– Извлечение по форматам и ключевым словам

The screenshot displays the Kofax Transformation Project Builder interface. A document viewer shows an invoice with the following extracted data:

- Invoice Number: 333323807
- Invoice Date: 04/27/03
- Payment Terms: NET DUE 30 DAYS
- Shipped Via: 2ND DAY
- Waybill Number: 24282586860

The 'Properties of Format Locator' dialog for 'FL_InvoiceNumber' is open, showing 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 contains the text: **Первый результат исходит со страницы 2, и это номер накладной** (The first result comes from page 2, and this is the invoice number).

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 result. A second window, 'Properties of Format Locator 'FL_InvoiceNumber'', shows a table of detected alternatives. A red callout box points to the second row of the table, indicating it is the correct result from page 1.

	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 shows the Kofax Transformation Project Builder interface. The main window displays an invoice document with extracted fields. A blue box highlights the first result, and a red box highlights the third result. A red callout box contains text explaining the third result.

Page: 1 of 2

1496 *
Invoice Number: 333323807
Invoice Number (98.95%): 333323807 (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
100.00 %	242825868
96.95 %	333323807
96.95 %	242825868
0.00 %	000318193
0.00 %	333323807
0.00 %	333323807
0.00 %	000318193
0.00 %	333323807
0.00 %	000318193
0.00 %	333323807

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 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 'Invoice Number' field. Below the document, the 'Properties of Format Locator 'FL_InvoiceNumber'' dialog is open, showing the 'Evaluation Settings' tab. The 'Keyword' is 'Invoice Number', and the 'Relation' is 'W, N, NW'. The 'Distance' is 'Near' and the 'Weight' is '100'. A blue callout box points to the 'Relation' field, explaining the importance of the 'W' (West) relation.

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

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 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, N, NW	Near	100	<input type="checkbox"/>	<input checked="" type="checkbox"/>

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 'Invoice Number' field and its extraction result. Below the main window, the 'Properties of Format Locator' dialog box is open, showing the 'Evaluation Settings' tab. The 'Keywords' section includes a list of keywords and their relations. A blue callout box points to the 'Invoice Number' keyword, which has a relation of 'W' (West).

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

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"/>

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 an invoice from OICE. A format locator, 'FL_InvoiceNumber', has been applied to the 'Invoice Number' field, which contains the value '333323807'. The 'Properties of Format Locator' dialog box is open, showing the 'Test Results' tab. A table of detected alternatives is visible, with the first row highlighted, indicating a 96.95% confidence level for the text '333323807'.

	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

У вас есть только один результат с высокой степенью уверенности/ Отлично!

Первый проект– Извлечение по форматам и ключевым словам

The screenshot shows the 'Kofax Transformation - Project Builder' interface. The 'Project Tree' on the left shows a project structure with 'Project Class' containing 'Invoices' and 'Fields'. The 'Details' pane on the right is titled 'Multiple Selection: Fields' and contains a table with the following columns: 'Field', 'Language', 'Display name', 'Field populated by', and 'Fixed field value'. Four fields are selected: 'InvoiceNumber (Inv...', 'InvoiceDate (Invoice', 'OrderNumber (Invoik', and 'Totalamount'. The 'Field populated by' dropdown for 'OrderNumber' is open, showing options: '<none>', '<fixed value>', and 'FL_InvoiceNumber'. A file explorer at the bottom shows a list of files with columns for 'Filename' and 'Classification'.

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

Filename	Classification	Percentage
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	88.4%

Назначьте свое поле локатору
1) Дважды щелкните по полям ниже Счета-фактуры

2) Выберите графу "Field populated by" to FL_InvoiceNumber

Первый проект– Извлечение по форматам и ключевым словам

3) Контролируйте новые настройки извлечения

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

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	88.4%	

Первый проект– Извлечение по форматам и ключевым словам

Значительное улучшение

Summary	InvoiceNumber	InvoiceDate	OrderNumber	Totalamount
A 23.44%	62.50%	18.75%	6.25%	6.25%
B 17.19%	31.25%	6.25%	6.25%	25.00%
C 58%	6.25%	75.00%	87.50%	68.75%
D 0.00%	0.00%	0.00%	0.00%	0.00%

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
000001FF.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%	0.00%	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###2T [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.

Первый проект – Извлечение по форматам и ключевым словам

◆ Настройки для FL_InvoiceDate:

Используйте одно или несколько predetermined регулярных выражений

Используйте ключевые слова

Keywords

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

Keyword: [DEST]

Match each word exactly (not fuzzy)

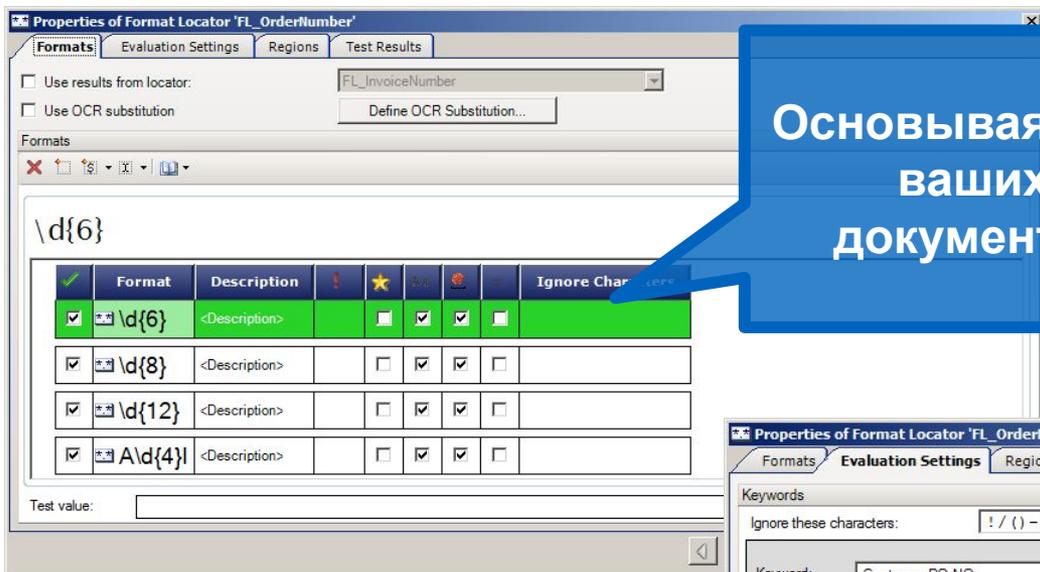
Match all words as a phrase

Weight: [31] Distance: [Near]

Use	D	Keyword	Relation	Distance	Weight	Search Exact	Whole Phrase
<input checked="" type="checkbox"/>		Date	W, N, NW	Near	31	<input type="checkbox"/>	<input type="checkbox"/>
<input checked="" type="checkbox"/>		DEST	SE	Near	31	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
<input checked="" type="checkbox"/>		Invoice Date	W, N, NW	Near	100	<input type="checkbox"/>	<input checked="" type="checkbox"/>
<input checked="" type="checkbox"/>		ORIGIN	SW	Near	31	<input type="checkbox"/>	<input type="checkbox"/>

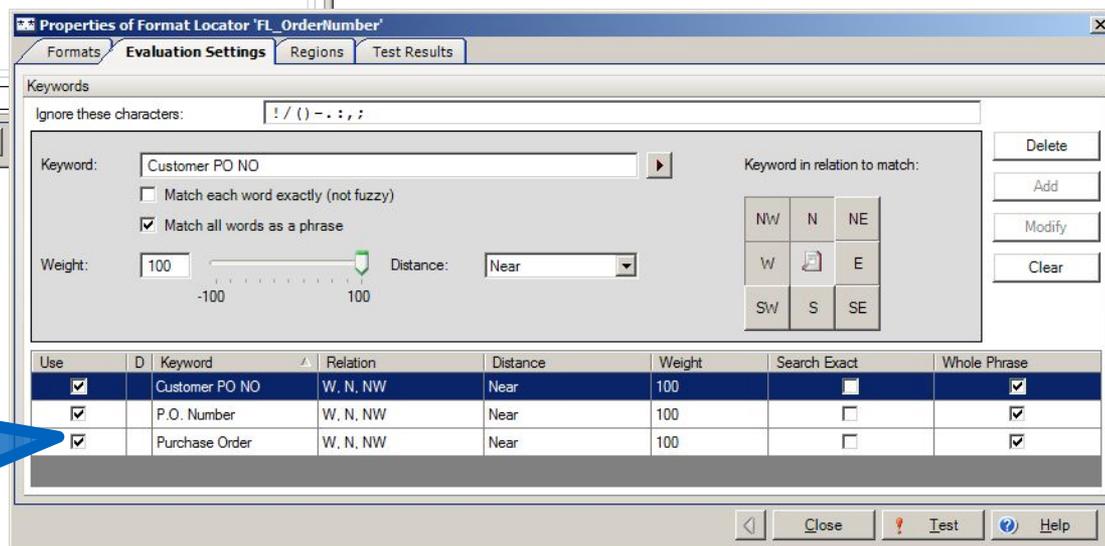
Первый проект– Извлечение по форматам и ключевым словам

◆ Настройки для FL_OrderNumber:



Основываясь на
ваших
документах

Используйте
ключевые слова



Первый проект– Извлечение по форматам и ключевым словам

◆ Настройки для FL_OrderNumber:

The screenshot shows the 'Properties of Format Locator' dialog box for 'FL_Totalamount'. The 'Formats' tab is active, showing a list of formats under the 'Amounts' category. A blue callout box points to this list with the text: **Предопределенные регулярные выражения для доступных сумм** (Predefined regular expressions for available amounts).

The 'Evaluation Settings' tab is also shown, featuring a 'Keywords' section. A blue callout box points to this section with the text: **Используйте ключевые слова** (Use keywords). The 'Keywords' section includes a text box for 'Ignore these characters' containing '!/()-.:;'. Below it, a 'Keyword' field contains 'Please pay this amount'. There are checkboxes for 'Match each word exactly (not fuzzy)' and 'Match all words as a phrase'. A 'Weight' slider is set to 100, and a 'Distance' dropdown is set to 'Near'. To the right, a grid of directional buttons (NW, N, NE, W, E, SW, S, SE) is visible. At the bottom, a table lists the configured keywords.

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

The screenshot shows the Kofax Transformation Project Builder interface. The 'Project Tree' on the left shows a project with a class named 'Invoices' containing several fields. The 'Details' pane on the right shows the configuration for these fields. A blue callout box with white text is overlaid on the interface, pointing to the 'Field populated by' column in the table. The text in the callout reads: 'Дважды щелкните по полям ниже Счета-фактуры класса Свяжите все свои поля с локаторами'. The table below shows the configuration for the 'Invoices' class fields.

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

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

Filename	Classification Re:	Confidence	Assigned Class
FedEx		84.2%	
FedEx		83.3%	
FedEx		83.6%	
FedEx		83.7%	
FedEx		85.5%	
FedEx		96.5%	
FedEx		96.2%	
FedEx		96.7%	
FedEx		97.2%	
FedEx		97.7%	
FedEx		94.8%	
FedEx		96.5%	
FedEx		95.1%	

Первый проект – Извлечение по форматам и ключевым словам

Запустите extraction benchmark еще раз

**Прекрасный результат
Но это недостаточно хорошо 😊**

	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%

Filename	Class	A	B	C	D	InvoiceNumber	InvoiceDate	OrderNumber	Totalamount
000001FF.xd	Oki	100.00%	0.00%	0.00%	0.00%	307707RI	02/17/2003	15800051	21553.42
00000201.xd	Oki	50.00%	0.00%	50.00%	0.00%	307243RI	01/29/2003	20020280	-/- [21590.66]
00000202.xd	Oki	100.00%	0.00%	0.00%	0.00%	307243RI	01/29/2003	20020280	21540.00
000002024F.xd	Dell	75.00%	25.00%	0.00%	0.00%	306091RI	12/09/2002	15800051	21540.00
00000250.xd	Dell	75.00%	25.00%	0.00%	0.00%	303947RI	09/27/2002	15800051	21540.00
00000251.xd	Dell	75.00%	25.00%	0.00%	0.00%	301814RI	07/25/2002	20020020	-/- [0.00]
00000252.xd	Dell	75.00%	25.00%	0.00%	0.00%	300603RI	06/19/2002	15800051	21555.74
00000253.xd	Dell	75.00%	25.00%	0.00%	0.00%	196328182	12/26/2002	20026306	2038.66
00000254.xd	Dell	75.00%	25.00%	0.00%	0.00%	347784804	05/18/2003	20036121	1702.46
00000255.xd	Dell	75.00%	25.00%	0.00%	0.00%	345004270	05/08/2003	20036114	1404.28
00000256.xd	Dell	75.00%	25.00%	0.00%	0.00%	344964623	05/09/2003	20036113	1318.86
00000257.xd	Dell	75.00%	25.00%	0.00%	0.00%	341350479	05/08/2003	20036106	1053.80

Progress: Extraction complete 100%
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

Green fields are correct (will be skipped on validation)

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%		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		

Test Validation

The document was classified to the following:

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

InvoiceNumber: 306091 RI

InvoiceDate: 12/09/2002

OrderNumber: 15800051

Totalamount: 21,540.00

Progress

Extraction complete

Processed 82 documents

Elapsed Time: 24 seconds

KOFAX

Первый проект - Усовершенствования

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

Желтые поля - это неправильные поля (они не соответствуют значению в наборе ссылок). Но они отображаются в валидации, и кто-то должен подтвердить свои значения.

InvoiceDate	OrderNumber	Totalamount
41.46%	89.02%	
52.44%	2.44%	
6.10%	8.54%	
0.00%	0.00%	

InvoiceNumber	InvoiceDate	OrderNumber	Totalamount
307707RI	05/21/2003	20030032	21553.42
307707RI	02/17/2003		
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		

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

Test Validation
Classification Result
2 fields valid, 2 fields invalid (0 fields invisible, 0 fields read-only)
InvoiceNumber: [307707RI] (Invalid)
InvoiceDate: 05/21/2003 (Valid)
OrderNumber: [20030032] (Invalid)
Totalamount: 21553.42 (Valid)

Ok! Semiconductor
RECEIVED
MAY 21 2003

Первый проект - Усовершенствования

Extraction Benchmark - Extraction Benchmark (All Classes)

Legend

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

Class	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	0.30%	1.22%	0.00%	6.10%	8.54%
D	0.00%	0.00%	0.00%	0.00%	0.00%

Filename	Class	A	B	C	D	InvoiceNumber	InvoiceDate	OrderNumber	Totalamount
0000015D.xd	Ok	50.00%	0.00%	0.00%	0.00%	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%	0.00%	0.00%	303947RI	09/27/2002		
00000202.xd	Ok	100.00%	0.00%	0.00%	0.00%	301814RI	07/25/2002		
00000202E.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

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: [Red Box]

The amount field does not contain a valid amount.

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)

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 30.30%	60.98%	52.44%	2.44%	0.00%
C 3.96%	1.22%	0.00%	6.10%	8.54%
D 0.00%	0.00%	0.00%	0.00%	0.00%

Test Validation

The document was classified to the following [Ok]

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

InvoiceNumber: 301814 RI (Valid) / 301814RI (Invalid)

InvoiceDate: 07/25/02 (Valid) / 07/25/2002 (Invalid)

OrderNumber: 20020020 (Valid) / 20020020 (Invalid)

Totalamount: [Empty]

Progress

Extraction complete

Processed 82 documents

Elapsed Time: 24 seconds

Annotations:

- Синие поля - это поля, которые являются правильными, НО кто-то должен их подтвердить.** (Blue fields are correct, but they need confirmation.)
- Красный восклицательный знак - значение неверное** (Red exclamation mark - incorrect value)

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	Okki	50.00%	0.00%	50.00%	0.00%	-/- [310264 RI]	05/21/2003	-/- [20030032]	21553.42
000001FD.xd	Okki	50.00%	0.00%	25.00%	25.00%	307707RI [307707 RI]	02/17/2003	15800051	-/- [21590.66]
000001FE.xd	Okki	100.00%	0.00%	0.00%	0.00%	307243RI	01/29/2003	20020280	21540.00
000001FF.xd	Okki	100.00%	0.00%	0.00%	0.00%	306091RI	12/09/2002	15800051	21540.00
00000200.xd	Okki	75.00%	25.00%	0.00%	0.00%	303947RI	09/27/2002	15800051	21540.00
00000201.xd	Okki	50.00%	25.00%	25.00%	0.00%	301814RI	07/25/2002	20020020	-/- [0.00]
00000202.xd	Okki	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	Oki	50.00%	0.00%
000001FD.xd	Oki	50.00%	0.00%
000001FE.xd	Oki	100.00%	0.00%
000001FF.xd	Oki	100.00%	0.00%
00000200.xd	Oki	75.00%	25.00%
00000201.xd	Oki	50.00%	25.00%
00000202.xd	Oki	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%	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	Oki	50.00%	0.00%	50.00%	0.00%	-/-	05/21/2003	-/-
C:\Documents\unsorted\000001FD.xdc	Oki	50.00%	0.00%	25.00% [-25.00%]	25.00% [25.00%]	307707RI [-/-]	02/17/2003	15800051
C:\Documents\unsorted\000001FE.xdc	Oki	100.00% [25.00%]	0.00%	0.00% [-25.00%]	0.00%	307243RI [-/-]	01/29/2003	20020280
C:\Documents\unsorted\000001FF.xdc	Oki	100.00% [25.00%]	0.00%	0.00% [-25.00%]	0.00%	306091RI [-/-]	12/09/2002	15800051
C:\Documents\unsorted\00000200.xdc	Oki	75.00% [25.00%]	25.00%	0.00% [-25.00%]	0.00%	303947RI [-/-]	09/27/2002	15800051
C:\Documents\unsorted\00000201.xdc	Oki	50.00% [25.00%]	25.00%	25.00% [-25.00%]	0.00%	301814RI [-/-]	07/25/2002	20020020
C:\Documents\unsorted\00000202.xdc	Oki	100.00% [25.00%]	0.00%	0.00% [-25.00%]	0.00%	300603RI [-/-]	06/19/2002	15800051

Elapsed Time: 25 seconds

Save... Export... Help

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

Первая проверка Порог валидации

Третья проверка Форматтер

Вторая проверка Расстояние между двумя результатами

Первый проект - Усовершенствования

Четвертая проверка Правил валидации, СВЯЗАННЫЕ С ПОЛЯМИ

Either there is nothing selected in the Project Tree or incompatible items are selected. As a result, nothing can be displayed in the Details window.

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	Oki	50.00%	0.00%	50.00%	0.00%	-/ [310264 RI]	05/21/2003	-/ [20030032]	21553.42
000001FD.xd	Oki	50.00%	0.00%	25.00%	25.00%	307707RI [307707 RI]	02/17/2003	15800051	-/ [21590.66]
000001FE.xd	Oki	100.00%	0.00%	0.00%	0.00%	307243RI	01/29/2003	20020280	21540.00
000001FF.xd	Oki	100.00%	0.00%	0.00%	0.00%	306091RI	12/09/2002	15800051	21540.00
00000200.xd	Oki	75.00%	25.00%	0.00%	0.00%	303947RI	09/27/2002	15800051	21540.00
00000201.xd	Oki	50.00%	25.00%	25.00%	0.00%	301814RI	07/25/2002	20020020	-/ [0.00]
00000202.xd	Oki	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%

	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

Classification result: Oki

ACCOUNTS PAYABLE DEPT. 102
16245 LAGUNA CANYON ROAD
IRVINE CA 92618-3603 IRV

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

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

Протестируйте его в средстве просмотра
документов
Вы увидите два ключевых слова,
выделенных оранжевым цветом, и значения
рядом с ним

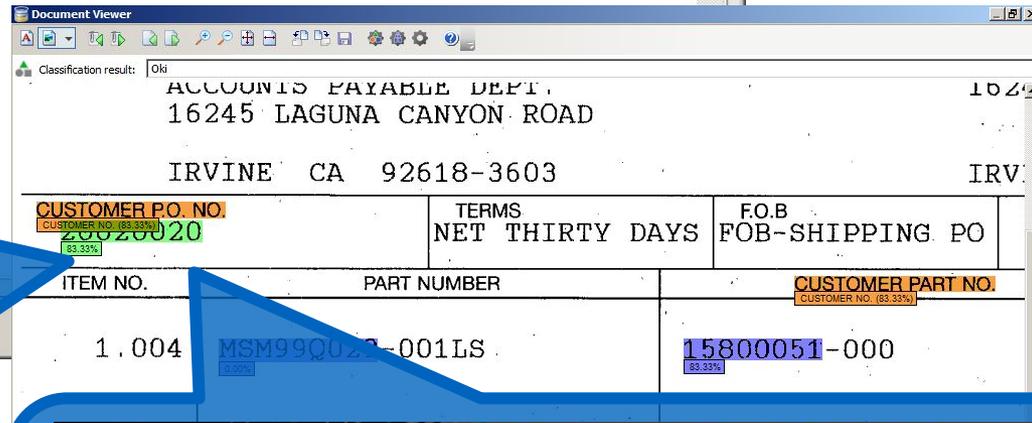
Первый проект - Усовершенствования

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



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

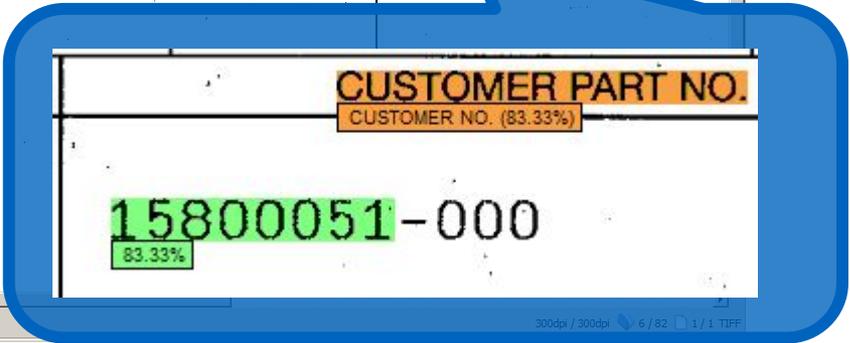
IRVINE CA 92618-3603 IR

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

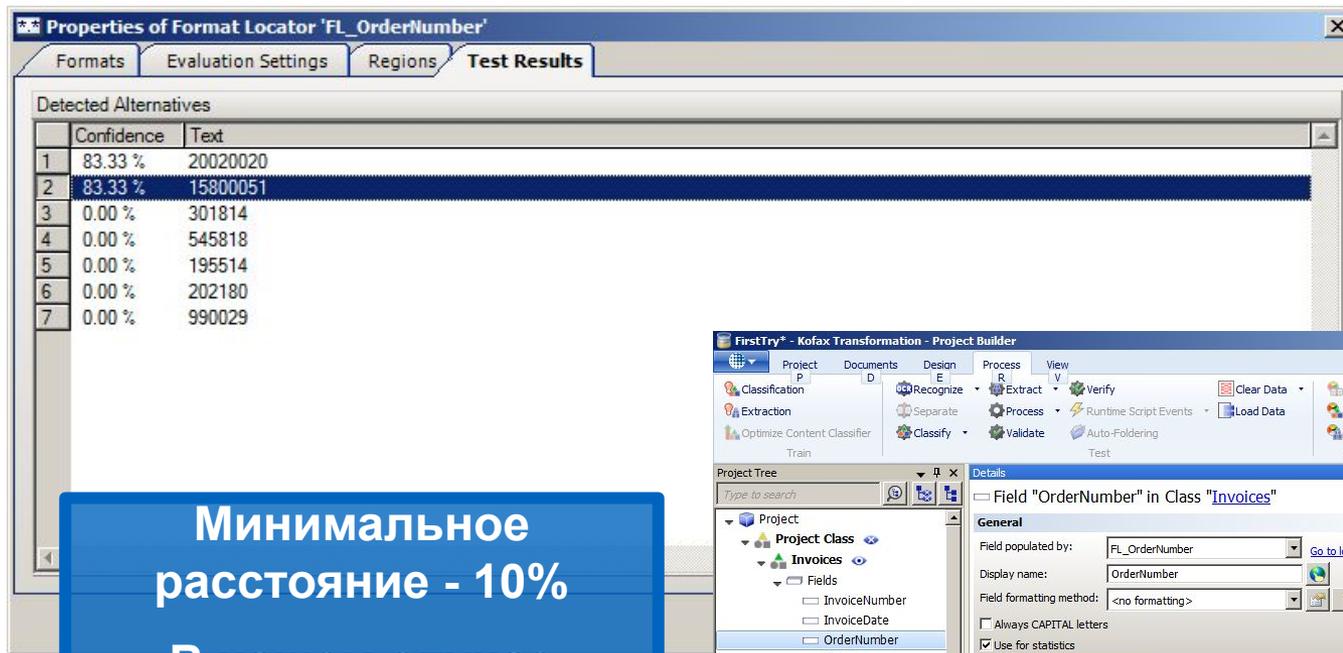
CUSTOMER PART NO.
CUSTOMER NO. (83.33%)

15800051-000
83.33%

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

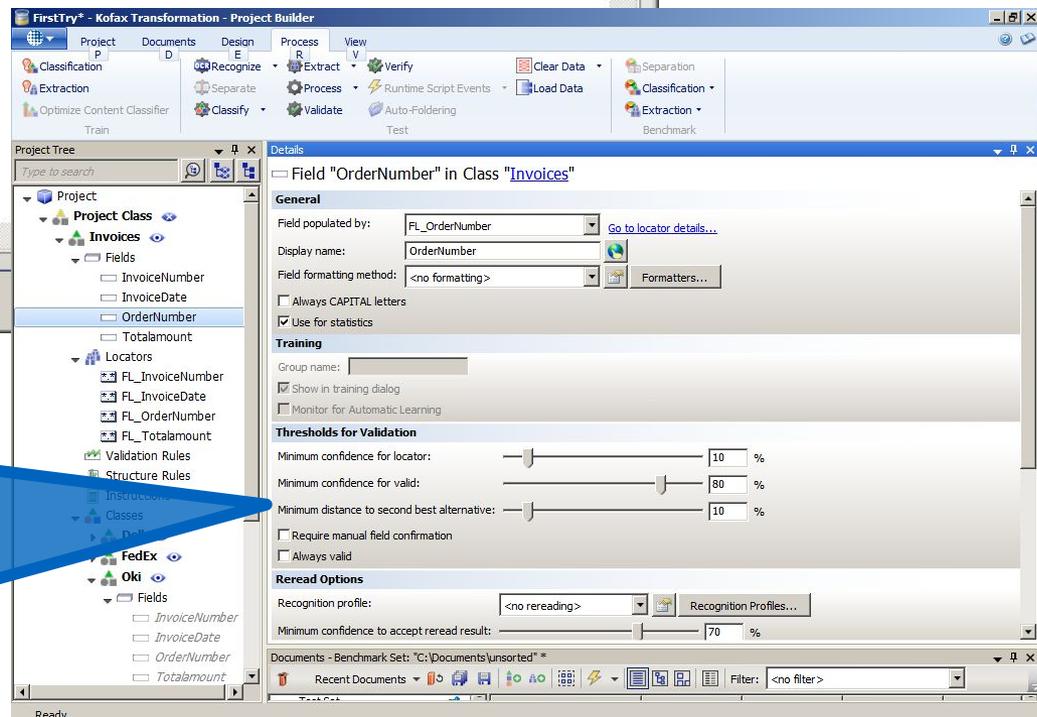


Первый проект - Усовершенствования



	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



Project Tree: Invoices > Fields > OrderNumber

Details: Field "OrderNumber" in Class "Invoices"

General

Field populated by: FL_OrderNumber
Display name: OrderNumber
Field formatting method: <no formatting>

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 %

Reread Options

Recognition profile: <no rereading>
Minimum confidence to accept reread result: 70 %

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	Okki	50.00%	0.00%	50.00%	0.00%	-/- [310264 RI]	05/21/2003	-/- [20030032]	21553.42
000001FD.xd	Okki	50.00%	0.00%	25.00%	25.00%	307707RI [307707 RI]	02/17/2003	1580005	-/- [21590.60]
000001FE.xd	Okki	100.00%	0.00%	0.00%	0.00%	307243RI	01/29/2003	20020280	21540.00
000001FF.xd	Okki	100.00%	0.00%	0.00%	0.00%	306091RI	12/09/2002	15800051	21540.00
00000200.xd	Okki	75.00%	25.00%	0.00%	0.00%	303947RI	09/27/2002	15800051	21540.00
00000201.xd	Okki	50.00%	25.00%	25.00%	0.00%	301814RI	07/25/2002	20020020	-/- [0.00]
00000202.xd	Okki	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

The screenshot displays the Kofax Transformation Project Builder interface. The main window shows a document viewer with an invoice from Dell. A blue callout box points to the 'FL_OrderNumber' locator in the Project Tree, with the text: "Двойной клик на локатор FL_OrderNumber".

The 'Properties of Format Locator 'FL_OrderNumber'' dialog box is open, showing the 'Formats' tab. It includes a dropdown menu for 'FL_InvoiceNumber' and a 'Define OCR Substitution...' button. Below this, there is a table of formats:

Format	Description	Star	Flag	Ignore Characters
<input checked="" type="checkbox"/> \d{6}	<Description>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
<input checked="" type="checkbox"/> \d{8}	<Description>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>

A second blue callout box points to the table with the text: "Тестируем".

At the bottom of the dialog, there is a 'Test value:' field and buttons for 'Close', 'Test', and 'Help'.

Первый проект - Усовершенствования

The screenshot displays the Kofax Transformation Project Builder interface. The main window shows a document with the text "This is your INVOICE" and "Customer Number: 000318193". A blue callout box points to the "Purchase Order" field, stating: "Вторая альтернатива имеет ключевые слова к северо-западу от ее местоположения". Below the document, the "Properties of Format Locator 'FL_OrderNumber'" dialog is open, showing a table of detected alternatives.

	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 row of the table, stating: "Альтернатива 1 - правильный номер счета-фактуры".

Первый проект - Усовершенствования

The screenshot displays the Kofax Transformation Project Builder interface. The main window shows a document with the text "This is your INVOICE" and various fields like "Customer Number", "Purchase Order", "Order Number", and "Order Date". A blue callout box points to the "Evaluation Settings" tab of the "Properties of Format Locator 'FL_OrderNumber'" dialog. This dialog shows the "Keywords" section with a table of settings for different keywords.

Просмотрите настройки ключевых слов. Только отношение W, не используйте W, N, NW (настройка по умолчанию)

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 classification results for 'Dell'. The document text includes fields like 'Customer Number', 'Purchase Order', 'Order Number', 'Order Date', and 'SHIP TO:'. A 'Properties of Format Locator' dialog is open, showing a table of detected alternatives for the 'FL_OrderNumber' locator.

Confidence	Text
100.00 %	20026306
100.00 %	20026306
0.00 %	261680
0.00 %	00031819
0.00 %	19632818
0.00 %	19632818
0.00 %	42050342
0.00 %	00031819
0.00 %	00024020
0.00 %	20026306
0.00 %	20026283
0.00 %	19632818
0.00 %	00031819
0.00 %	19632818

The dialog also shows 'Detected Alternatives' and 'Evaluation Settings' tabs. The 'Evaluation Settings' tab is active, showing 'The locator is a OrderNumber'.

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 a 'Benchmark' test results table for 'Extraction Benchmark - Extraction Benchmark (All Classes)'. The table includes columns for 'All Fields', 'InvoiceNumber', 'InvoiceDate', 'OrderNumber', and 'Totalamount'. The results are categorized into four groups: A (Correct valid fields), B (Correct invalid fields), C, and D. Below the summary table is a 'Details - Sorted by Column Content' table with columns for 'Filename', 'Class', 'A', 'B', 'C', 'D', 'InvoiceNumber', 'InvoiceDate', 'OrderNumber', and 'Totalamount'. The progress bar at the bottom indicates 'Extraction complete' and 'Processed 82 documents'.

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

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	[00000000]	1553.42
000001FD.xd	Oki	50.00%	0.00%	25.00%	25.00%	307707RI [307707 RI]	02/17/2003	1500061	-/- [21590.66]
000001FE.xd	Oki	100.00%	0.00%	0.00%	0.00%	307243RI	01/29/2003	200280	21540.00
000001FF.xd	Oki	100.00%	0.00%	0.00%	0.00%	306091RI	12/09/2002	1500051	21540.00
00000200.xd	Oki	75.00%	25.00%	0.00%	0.00%	303947RI	09/27/2002	1580061	21540.00
00000201.xd	Oki	50.00%	25.00%	25.00%	0.00%	301814RI	07/25/2002	200280	-/- [0.00]
00000202.xd	Oki	100.00%	0.00%	0.00%	0.00%	300603RI	06/19/2002	1580000	21555.74
0000024F.xd	Dell	100.00%	0.00%	0.00%	0.00%	196328182	12/26/2002	2002630	2038.66
00000250.xd	Dell	100.00%	0.00%	0.00%	0.00%	347784804	05/18/2003	2003612	1702.46
00000251.xd	Dell	100.00%	0.00%	0.00%	0.00%	345004270	05/08/2003	20036114	1404.28
00000252.xd	Dell	100.00%	0.00%	0.00%	0.00%	344964623	05/09/2003	20036113	1318.86
00000253.xd	Dell	100.00%	0.00%	0.00%	0.00%	341350479	05/08/2003	20036106	1053.80
00000254.xd	Dell	100.00%	0.00%	0.00%	0.00%	333323807	04/27/2003	20036099	2523.52

Первый проект - Усовершенствования

The screenshot displays the 'Extraction Benchmark' window in the Kofax Transformation Project Builder. It shows a summary table and a details table for various fields across different document classes.

Legend:

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

Summary Table:

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	15722
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	13288 [313288212400] -/ [115 40]
FedEx2.xdc	FedEx	75.00%	25.00%	0.00%	0.00%	610129688	07/22/2004	4582
FedEx20.xdc	FedEx	50.00%	25.00%	25.00%	0.00%	352783429	04/06/2004	313525 [31328212167] 319

Progress: Extraction complete, 100% processed, 82 documents.

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 displays the Kofax Transformation Project Builder interface. The main window shows a document viewer with a FedEx invoice. The classification result is 'FedEx'. The invoice details include:

- SHIPPER: COATS NORTH AMERICA, 650 AMERICAN THREAD RD, MARION, NC 28752
- CONSIGNEE: AVCOR, 1520 EVERMAN PARKWAY, FORT WORTH, TX 76140
- BILL TO: COATS AMERICAN, 3 TRENDSET INC, P O BOX 1208, MAULDIN, SC 29662
- TRACKING NUMBER: 652 325 439
- SHIP DATE: 07/22/04 (31.00% confidence)
- ORIGIN: WEST (100%)
- REV: 62.76
- DESCRIPTION: 18 CARTONS, THREAD, SEWING, COTTON ON PONES, COBES OR SPINLE
- WT(LBS): 213, N M F C: 182220-01, CLASS: 060, RATE: 88.72%, TOTAL CHARGES: 58.00

The 'Properties of Format Locator 'FL_InvoiceDate'' window is open, showing the 'Test Results' tab. The 'Detected Alternatives' table is as follows:

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

A blue callout box points to the first alternative (07/22/04) with the text: "Уверенность в альтернативе слишком низкая. Давайте снова проверять ключевые слова" (Confidence in the alternative is too low. Let's check the key words again).

Первый проект - Усовершенствования

The screenshot shows the Kofax Transformation Project Builder interface. The main window displays a FedEx shipping label with the following details:

Classification result: FedEx

SHIPPER: COATS NORTH AMERICA, 630 AMERICAN THREAD RD, MARION, NC 28752

SHIPPER ID: 21344610

BILL TO: COATS AMERICAN, 3 TRENDSET INC, P. O. BOX 1208, MAULDIN, SC 29662

TRACKING NUMBER: 652 325 439

DATE: 07/22/04

ORIGIN: AVL, **DEST:** FTW

SHIPPER ID: 84616794

The **Properties of Format Locator 'FL_InvoiceDate'** window is open, showing the **Evaluation Settings** tab. The **Keywords** list includes:

Use	D	Keyword	Relation	Distance	Weight	Search Exact	Whole Phrase
<input checked="" type="checkbox"/>		Date	W, N, NW	Near	100	<input type="checkbox"/>	<input type="checkbox"/>
<input checked="" type="checkbox"/>		DEST	S, SE	Near	31	<input type="checkbox"/>	<input type="checkbox"/>
<input checked="" type="checkbox"/>		Invoice Date	W, N, NW	Near	100	<input type="checkbox"/>	<input type="checkbox"/>
<input checked="" type="checkbox"/>		ORIGIN	S, SW	Near	31	<input type="checkbox"/>	<input type="checkbox"/>

The blue callout box contains the following text:

Ключевое слово DEST и ORIGIN был установлен на юго-восток и юго-запад. Отношение OCR (хdoc) 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 document viewer with a FedEx Freight invoice. The classification result is 'FedEx'. A 'Properties of Format Locator' window is open, showing the 'Test Results' tab. The 'Detected Alternatives' table is as follows:

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

A blue callout box with the text 'Идеально! 😊' points to the first row of the table, indicating a high confidence in the date extraction.

Первый проект - Усовершенствования

Запускаем эталонный тест извлечения

Посмотрите на сводку результатов тестирования:

88.72 %

являются правильными и пропущены!

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

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.**

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

100 %

Processed 82 documents

Elapsed Time: 25 seconds

Первый проект - Усовершенствования

The screenshot displays the Kofax Transformation Project Builder interface. The 'Project Tree' on the left shows a project named 'Invoices' with various fields and locators. The 'Details' pane shows the configuration for the 'FL_InvoiceDate' locator, which is assigned to the 'Invoices' class and uses the 'Format Locator' method. A blue callout box with a white arrow points to the 'Verify' button in the 'Process' menu, containing the following 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.

The 'Documents' pane at the bottom shows a benchmark set for 'C:\Documents\unsorted'. The table below lists the classification results for various documents:

Filename	Classification Re:	Confidence	Assigned Class
000001FC.xdc*	Ok	100.0%	
000001FD.xdc*	Ok	100.0%	
000001FE.xdc*	Ok	91.5%	
000001FF.xdc*	Ok	89.2%	
00000200.xdc*	Ok	88.8%	
00000201.xdc*	Ok	100.0%	
00000202.xdc*	Ok	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%	

Первый проект - Усовершенствования

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

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)

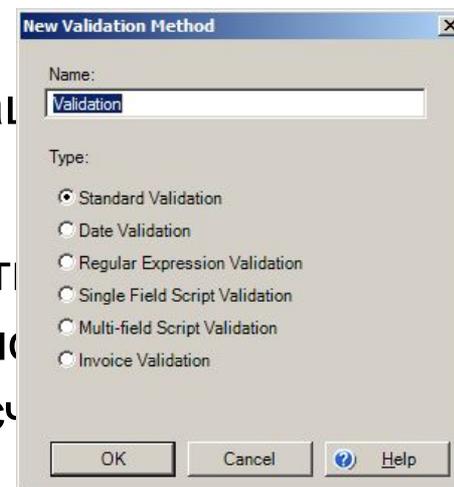
Как сделать значения зелеными или красными автоматически

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)

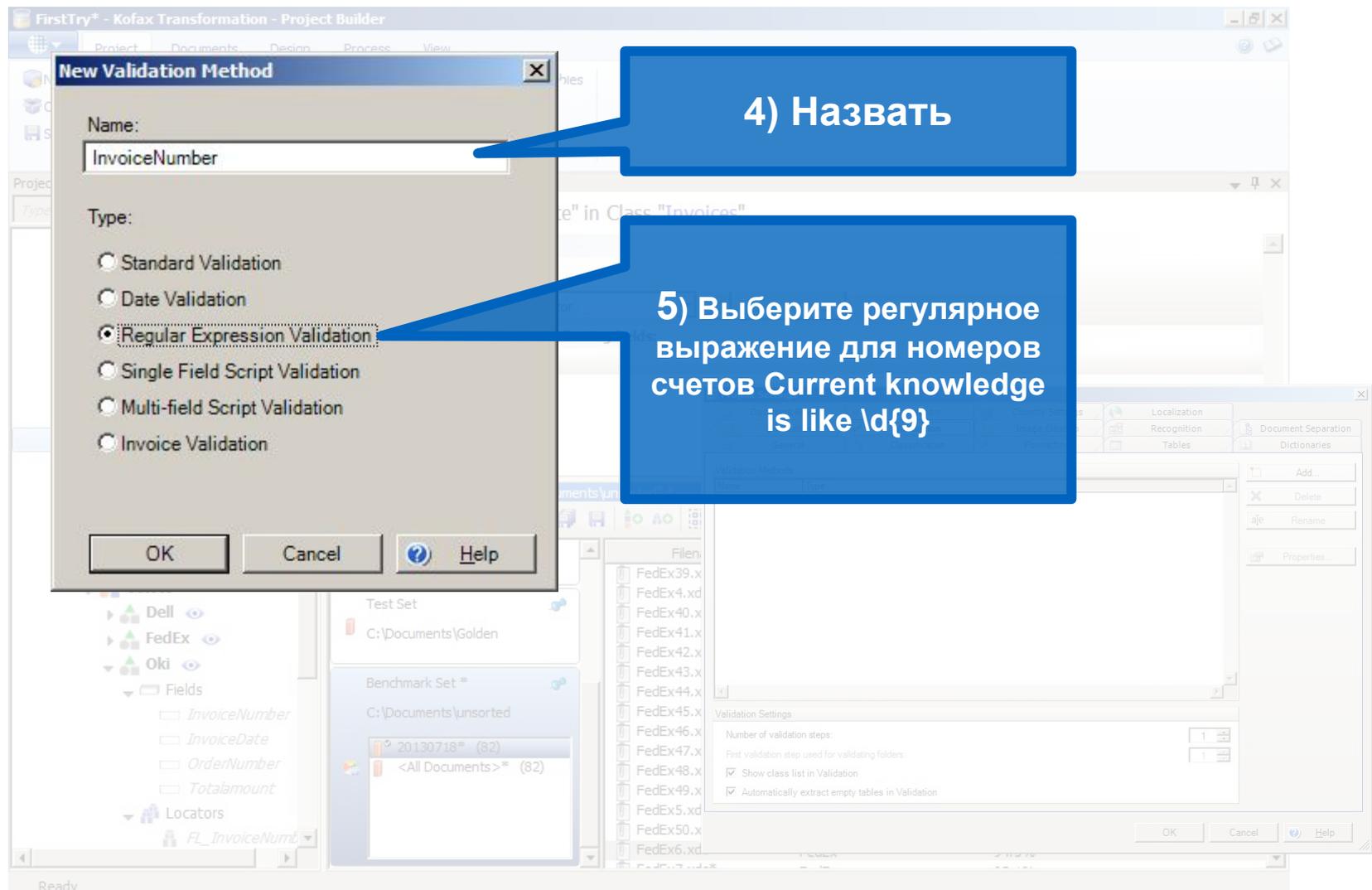
The screenshot shows the Kofax Transformation Project Builder interface. The top menu bar includes Project, Documents, Design, Process, and View. The Project Tree on the left shows a hierarchy of Fields (InvoiceNumber, OrderNumber, Totalamount) and Locators (FL_InvoiceDate, FL_OrderNumber, FL_Totalamount). The Project Settings window is open, showing the 'Validation' tab. The 'Validation Methods' table is empty, and the 'Add...' button is highlighted. The 'Validation Settings' section shows 'Number of validation steps' set to 1 and 'Show class list in Validation' checked.

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

Проверьте регулярные выражения

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

Первый проект - Зеленые и красные поля (Правила валидации)

7) Создайте другой Метод Валидации для дат счет фактуры

8) Выбрав Date Validation для дат счет-фактуры

The screenshot shows the 'New Validation Method' dialog box with the following details:

- Name: InvoiceDate
- Type: Date Validation
- Other options: Standard Validation, Regular Expression Validation, Single Field Script Validation, Multi-field Script Validation, Invoice Validation

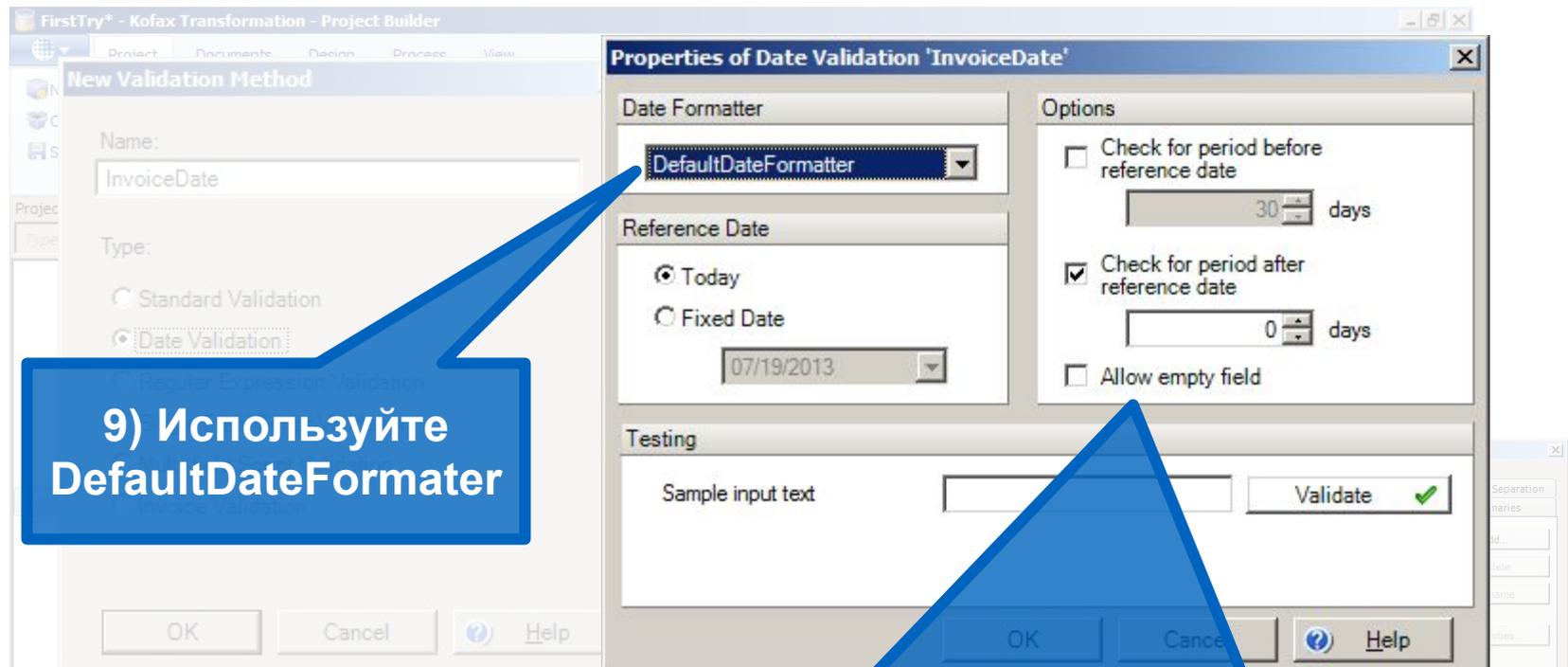
The 'Project Settings' dialog box shows the 'Validation' tab with the following details:

- Validation Methods table:

Name	Type

- Validation Settings:
 - Number of validation steps: 1
 - First validation step used for validating folders: 1
 - Show class list in Validation
 - Automatically extract empty tables in Validation

Первый проект - Зеленые и красные поля (Правила валидации)



9) Используйте DefaultDateFormatter

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

Name: Ordernumber

Type:

- Standard Validation
- Date Validation
- Regular Expression Validation
- Single Field Script Validation
- Multi-field Script Validation
- Invoice Validation

OK Cancel Help

Formats

Format: \d{12}

High performance

Description:

Error description: Field item does not match any of the given expressions.

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

12) Добавьте регулярные выражения

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

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 shows the Kofax Transformation Project Builder interface. The 'Project Tree' on the left shows the 'Invoices' folder expanded to 'Fields', with 'InvoiceNumber' selected. The 'Details' pane shows the configuration for a 'Single Field Validation Rule "InvoiceNumber"'. The 'Field' dropdown is set to 'InvoiceNumber'. The 'Validation Methods' section shows a yellow star icon, indicating that no validation methods are currently selected. A 'Validation Method' dialog box is open in the foreground, showing a list of validation methods with 'InvoiceNumber' selected. The background shows a 'Documents' window with a table of data.

4) Выберите поле InvoiceNumber

5) Добавьте метод валидации

6) Добавьте метод InvoiceNumber

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

Первый проект - Зеленые и красные поля (Правила валидации)

The screenshot displays the Kofax Transformation Project Builder interface. The 'Project Tree' on the left shows a project named 'Invoices' with several fields: InvoiceNumber, InvoiceDate, OrderNumber, and Totalamount. The 'Details' pane shows a 'Single Field Validation Rule "InvoiceDate" in Class "Invoices"'. The 'General' section has a checkbox for 'Validation rule makes uncertain extraction results valid' and a 'Field' dropdown set to 'InvoiceDate'. The 'Validation Methods' section shows a star icon, a red X icon, and an orange arrow icon. Below these are 'Selected validation methods:' and 'InvoiceDate'. A 'Validation Method' dialog box is open, prompting the user to 'Please select a validation method:' with a dropdown menu showing 'InvoiceDate' and 'Configure...' button. The 'Documents' pane at the bottom shows a benchmark set for 'C:\Documents\unsorted' with a table of files and their classification results.

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%	

Создайте новое правило для invoice date

Первый проект - Зеленые и красные поля (Правила валидации)

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

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

File Name	Classification Res	Confidence	Assigned Class
FedEx39.xdc*			
FedEx4.xdc*	FedEx	96.3%	
FedEx40.xdc*	FedEx	83.0%	
FedEx41.xdc*	FedEx	83.7%	
FedEx42.xdc*	FedEx	84.1%	
FedEx43.xdc*	FedEx	84.1%	
FedEx44.xdc*	FedEx	97.7%	
FedEx45.xdc*	FedEx	96.7%	
FedEx46.xdc*	FedEx	96.5%	
FedEx47.xdc*	FedEx	96.2%	
FedEx48.xdc*	FedEx	95.6%	
FedEx49.xdc*	FedEx	94.5%	
FedEx50.xdc*	FedEx		
FedEx51.xdc*	FedEx		
FedEx52.xdc*	FedEx		
FedEx53.xdc*	FedEx		
FedEx54.xdc*	FedEx		
FedEx55.xdc*	FedEx		
FedEx56.xdc*	FedEx		
FedEx57.xdc*	FedEx		
FedEx58.xdc*	FedEx		
FedEx59.xdc*	FedEx		
FedEx60.xdc*	FedEx		
FedEx61.xdc*	FedEx		
FedEx62.xdc*	FedEx		
FedEx63.xdc*	FedEx		
FedEx64.xdc*	FedEx		
FedEx65.xdc*	FedEx		
FedEx66.xdc*	FedEx		
FedEx67.xdc*	FedEx		
FedEx68.xdc*	FedEx		
FedEx69.xdc*	FedEx		
FedEx70.xdc*	FedEx		
FedEx71.xdc*	FedEx		
FedEx72.xdc*	FedEx		
FedEx73.xdc*	FedEx		
FedEx74.xdc*	FedEx		
FedEx75.xdc*	FedEx		
FedEx76.xdc*	FedEx		
FedEx77.xdc*	FedEx		
FedEx78.xdc*	FedEx		
FedEx79.xdc*	FedEx		
FedEx80.xdc*	FedEx		
FedEx81.xdc*	FedEx		
FedEx82.xdc*	FedEx		
FedEx83.xdc*	FedEx		
FedEx84.xdc*	FedEx		
FedEx85.xdc*	FedEx		
FedEx86.xdc*	FedEx		
FedEx87.xdc*	FedEx		
FedEx88.xdc*	FedEx		
FedEx89.xdc*	FedEx		
FedEx90.xdc*	FedEx		
FedEx91.xdc*	FedEx		
FedEx92.xdc*	FedEx		
FedEx93.xdc*	FedEx		
FedEx94.xdc*	FedEx		
FedEx95.xdc*	FedEx		
FedEx96.xdc*	FedEx		
FedEx97.xdc*	FedEx		
FedEx98.xdc*	FedEx		
FedEx99.xdc*	FedEx		
FedEx100.xdc*	FedEx		

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

OrderNumber	InvoiceDate	OrderNumber	Totalamount
425772185	11/24/2003	296175	52.52
426120293	11/17/2003	-/-	1762.59
426120343	11/17/2003	-/-	721.08
426120384	11/17/2003	299013	436.76
426120525	11/19/2003	299207	158.44
570874695	07/22/2004	454673	181.66
476333010	12/01/2003	-/-	1062.69
476334133	11/25/2003	-/-	79.23
476334158	11/26/2003	-/-	167.46
476334232	11/25/2003	-/-	473.60
476335130	12/09/2003	299627	451.74
632324093	07/23/2004	033229	81.69

Первый проект - Зеленые и красные поля (Правила валидации)

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%	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%	25.00%	0.00%	1967782	12/26/2002	20026306	2038
C:\Documents\unsorted\00000250.xdc	Ok	100.00% [50.00%]	0.00%	25.00%	25.00%	0.00%	347784804	05/18/2003	20036121	1702
C:\Documents\unsorted\00000251.xdc	Ok	100.00% [50.00%]	0.00%	25.00%	25.00%	0.00%	345004370	05/09/2003	20036111	1702

Elapsed Time: 25 seconds

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