

# Invoice Capture Center

## Line Item recognition and PO Line Item Mapping

Gisela Hammann, SAP Solutions Group  
March 2012

### Abstract

This document gives insight in how line item recognition works in Invoice Capture Center (ICC), and how mapping with Purchase Order/Goods Receipt Data is working in detail. This document applies for versions > ICC 5.2 SP5, and > ICC 6.0 SP1.

Highlights:

- Line item recognition
- Mapping with downloaded PO/GR data
- Customizing settings



## Contents

1.	<b>Introduction</b> .....	3
2.	<b>OCR invoice item recognition</b> .....	4
3.	<b>DOKuStar Invoice Items</b> .....	6
4.	<b>PO line item Mapping</b> .....	7
	Evaluation of customer business scenario .....	7
5.	<b>ICC LineItemMapping Algorithm</b> .....	8
	Data set used for line item mapping .....	8
	OCR data .....	8
	SAP download data .....	8
	Line item mapping algorithm .....	9
	Determine relevant POItems for Mapping .....	9
	Assignment mapping .....	10
	SnapMatch mapping and Assignment mapping .....	10
	Mapping result .....	10
6.	<b>Tips for successful use of invoice line item mapping</b> .....	13
	<b>About Open Text</b> .....	14

[www.opentext.com](http://www.opentext.com)

## 1. Introduction

Line item recognition is a feature which is considered as crucial for Invoice Capture Center (ICC), as it has an impact on automation rate of the end to end system.

In baseline, Vendor Invoice Management (VIM) expects OCR line item data delivered with PO number and PO line item number.

ICC line item mapping has been conceived to provide automation support for standard invoices with few line items and with good relation to PO. Other options should be considered, if relationship between PO data and invoice data is not obvious, due to specific business scenarios running in the customer environment.

Alternative VIM customizing options exist, which will not at all require invoice line item data delivered by recognition, or which will do PO line item mapping using VIM methods. For more details please refer to VIM 6.0 Configuration Guide chapter 5.1.7.

[www.opentext.com](http://www.opentext.com)

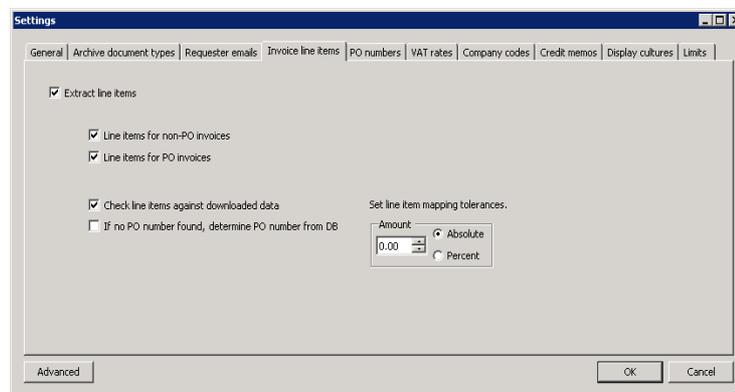
## 2. OCR invoice item recognition

In general we discern invoice header and invoice item data.

Invoice header data in general are fields like invoice date, invoice number, which occur once, mainly on the first page of the invoice, and which are determined by searching for structured data related to keywords and phrases.

Invoice item data recognition is more complex, as multiple line items may exist on an invoice, and line items very often have not a regular layout, but are printed in different variations.

The following parameters can be set for defining how line item recognition should work in a specific customer environment:



- **Extract line items:** Invoice item recognition can be switched on and off.
- **Line items for PO invoices / line items for non-PO invoices:** ICC will recognize line items, but only transmit line item data to VIM for the option which is checked. E.g. if line items for nonPO data would be unchecked, ICC would **not** transmit data for invoices where no PO number has been found.
- **Check line items against downloaded data:** This parameter activates line item mapping with Purchase Order/Goods Receipt detail data downloaded from SAP system. For mapping of amounts, an absolute or relative tolerance can be defined.
- **If no PO found determine PO number from DB (available since ICC 5.2 SP7 and since ICC 6.0 SP2):** This parameter is not activated by default in a newly created application. Because of compatibility reasons it is activated when importing older applications, but it is highly recommended to deactivate this parameter. If the parameter is not activated, it stops LineltemMapping algorithm, if no PO number has been

[www.opentext.com](http://www.opentext.com)

# OPEN TEXT

The Content Experts™

found on invoice. If the parameter is activated, ICC LineItemMapping tries to map with all PO items available for the current vendor. This may take a long time, if many PO items exist in PO download data for the current vendor, and it may cause erroneous mapping results, especially same vendor sends PO and NonPO invoices.

[www.opentext.com](http://www.opentext.com)

For more information about Open Text products and services, visit [www.opentext.com](http://www.opentext.com). Open Text is a publicly traded company on both NASDAQ (OTEX) and the TSX (OTC).  
Copyright © 2009 by Open Text Corporation. Open Text and The Content Experts are trademarks or registered trademarks of Open Text Corporation. This list is not exhaustive. All other trademarks or registered trademarks are the property of their respective owners. All rights reserved. SKU#\_EN

Team  
Suite

### 3. DOKuStar Invoice Items Recognition

The underlying DOKuStar KnowledgeBase method for table recognition is generic and works out of the box. For specific customer business scenarios it may be appropriate to extend line item recognition by customizing, if e.g. line item mapping should rely on additional line item information like material numbers etc for providing better mapping results. For more detailed information about customizing for line items please refer to ICC Customizing Guide.

Invoice line item recognition has been elaborated by analyzing a big volume of various invoices from different vendors and different countries. Invoice line item recognition works independently from PO download data, PO download data are not considered for optimizing recognition result. Invoice line item recognition considers local writing styles for all supported countries. The DOKuStar method for line item recognition takes text lines found by pure OCR as a starting point

These text lines are analyzed for occurrences of syntactical structures like table header keywords, amounts, quantities, units of measure, phrases typical for summarizing lines at the end of the table a.s.o. Results of this first step are used to find table header and end of table as well as horizontal position of lines and vertical position of columns. Logical checks like calculation of unit price x quantity = item price are also used in this second step. All this results in a list of line items, each line item consisting of one or more text lines.

For invoices, some line items currently are filtered out and not considered as valid line items:

- Line items with negative value (except for credit memos). For credit memos, negative line items are not filtered out.
- Line items with keywords like “expenses, discount”, showing incidental costs, discounts or additional costs

For ICC 7.0 it is planned to develop a business scenario framework providing customizing options for handling line items currently filtered out.

## 4. PO line item Mapping

Line item recognition using the table recognition algorithm, described in the precedent chapter, will have very good recognition results. Nevertheless, by default, VIM requires for PO invoices, for every line item PO number and PO line item number. The fact is, that PO line item number is not printed on the invoice, but can only be determined from PO data. Also for some invoice items, especially multiline invoice items on invoices referring to multiple POs, it is difficult to assign PO numbers to line items by means of optical character recognition. This is the reason why data mapping with downloaded PO and GR data is used in order to complete data to be delivered to VIM.

### Evaluation of customer business scenario

Before deciding to activate PO line item mapping, it is crucial to test, if this generic feature will fit to specific customer environment, if customizing is needed for adapting to customer business scenario, or if PO line item mapping cannot be used in ICC, but must be used in VIM. ICC and VIM interaction is very flexible, so in some cases it may be better to do line item mapping in VIM.

Currently, before activating PO line item mapping in ICC, the following business scenarios should be checked:

- Discount handling on invoice line item level vs. availability of discount data in PO download
- Handling of additional costs on invoice line item level vs. availability of additional cost data in PO download
- If multiple invoices refer to one PO, check availability of GR data in download data. Without GR data, line item mapping will not work.

## 5. ICC LineltemMapping Algorithm

ICC Lineltem mapping has been built using innovative generic SnapMatch algorithm. This is an OCR full text based pattern matching algorithm, working language independent. ICC Lineltem mapping works based on PO details data downloaded from SAP and stored in ICC internal data base. ICC Lineltem Mapping tries to map data contained in data base with DOKuStar invoice item data.

### Data set used for line item mapping

#### OCR data

ICC invoice line item mapping uses data extracted by DOKuStar invoice items table recognition. Line item mapping considers structured table data for mapping and evaluates against PO details download data. Depending on parameter settings, in some cases, line item mapping considers additional OCR data, which have not been acknowledged by table recognition, as valid line candidates, with lower priority.

#### SAP download data

For line item mapping, data listed below are downloaded from SAP. For description of the ABAP download report in SAP please refer to VIM Configuration Guide.

#### PO Header data

	Column Name	Data Type	Allow Nulls
?	Id	int	<input type="checkbox"/>
	ORDERID	nvarchar(255)	<input type="checkbox"/>
	VENDORID	nvarchar(255)	<input type="checkbox"/>
	DELIVERY	nvarchar(255)	<input type="checkbox"/>
	DELIVERYDATE	nvarchar(255)	<input type="checkbox"/>
	CURRENCY	nvarchar(255)	<input type="checkbox"/>
	RESERVE1	nvarchar(255)	<input type="checkbox"/>
	RESERVE2	nvarchar(255)	<input type="checkbox"/>
	CUSTOM1	nvarchar(255)	<input type="checkbox"/>
	CUSTOM2	nvarchar(255)	<input type="checkbox"/>
	COMPANYCODE	nvarchar(255)	<input type="checkbox"/>
	SYSTEM	nvarchar(255)	<input type="checkbox"/>

## PO item data (PO details)

Column Name	Data Type	Allow Nulls
Id	int	<input type="checkbox"/>
ORDERID	nvarchar(255)	<input type="checkbox"/>
POSITION	nvarchar(255)	<input type="checkbox"/>
ITEMTYPE	nvarchar(255)	<input type="checkbox"/>
DELIVERY	nvarchar(255)	<input type="checkbox"/>
AMOUNT	nvarchar(255)	<input type="checkbox"/>
CONDITION	nvarchar(255)	<input type="checkbox"/>
DESCRIPTION	nvarchar(255)	<input type="checkbox"/>
QUANTITY	nvarchar(255)	<input type="checkbox"/>
UNITPRICE	nvarchar(255)	<input type="checkbox"/>
UNIT	nvarchar(255)	<input type="checkbox"/>
RESERVE1	nvarchar(255)	<input type="checkbox"/>
RESERVE2	nvarchar(255)	<input type="checkbox"/>
CUSTOM1	nvarchar(255)	<input type="checkbox"/>
CUSTOM2	nvarchar(255)	<input type="checkbox"/>
SYSTEM	nvarchar(255)	<input type="checkbox"/>

There are 2 different item types:

- „1“ PO item
- „2“ GR item (posted)

The column “reserve1” is intended to be used for implementing custom specific extensions, by adding an additional line item field to be considered for mapping, e.g. material number. If line item table recognition will be extended to recognize e.g. material number in item table (for description how to do this, please refer to ICC Customizing Guide: Adding a Custom Column to the Invoice Items Custom Field), line item mapping will work for this new column, if download report fills material numbers in column “reserve1”.

Line item mapping will not work, if one of the table columns will be missing in download data, but it will work, if additional columns exist. Only reserve1 column will be considered for mapping, additional custom fields will not be considered.

## Line item mapping algorithm

Line item mapping algorithm is working in subsequent steps:

### Determine relevant POItems for Mapping

Based on determined vendor, Company Code and PO numbers (if available in download: Delivery note numbers) from the invoice, download data are filtered for relevant items. Line item mapping is dependent on good PO/DN number recognition.

For getting a good data quality for mapping, the algorithm refines PO download data as follows:

[www.opentext.com](http://www.opentext.com)

- If in PO item  $QUANTITY * UNITPRICE \neq AMOUNT$ , then a clone of the POItem with the correct amount value  $QUANTITY * UNITPRICE$  is inserted, and also considered for mapping with OCR data.
- If PO items and GR items with identical  $QUANTITY, UNITPRICE$  and  $AMOUNT, ORDERID$  and  $POSITION$  are found, PO items are deleted for not generating duplicated mapping.

With this set of relevant PO and GR items, mapping with OCR data is set up.

## Assignment mapping

Assignment mapping is applied, if the sum of the extracted invoice item amounts equals the net amount of the invoice. Then only the corresponding PO Numbers and PO Item Numbers need to be assigned,

This assignment is based on the criteria listed below (and will be applied in the mentioned order)

1. Quantity, UnitPrice, ItemAmount on POItems of type 1 & 2
2. ItemAmount of type 1 & 2

Based on these criteria, groups of POItems and InvoiceItems are created.

If another criteria-identical group is found, the mapping algorithm assigns POItems of the group with best fitting invoice items in regard to Description, ItemUnit and Reserve1. For mapping, the absolute or relative tolerance, if defined in customizing, is considered for the amount fields ItemUnitPrice and ItemAmount.

After an item group was mapped, its POItems and InvoiceItems will be removed from the list of mapping candidates.

## SnapMatch mapping and Assignment mapping

SnapMatch mapping and Assignment mapping is applied in combination, if the sum of the extracted invoice item amounts does not equal the net amount of the invoice.

## Mapping result

The result of the mapping is, that the columns  $ORDERID, POSITION, DELIVERY,$  are written to the invoice item result data.

### Example

The following example shows how mapping algorithm is working:

[www.opentext.com](http://www.opentext.com)

Invoice data:



**C.E.B. Berlin**

C.F.R. Berlin - Kolping Str. 15 - 10115 Berlin

Innovate Corporation  
Mainzer Landstraße 189  
D-60441 Frankfurt am Main

**Kundenservice**  
Telefon: 030/670091-0  
Fax: 030/670091-99  
Email: ceb@berlin.de

Datum: 20.07.2009

## Rechnung

<b>Besteller</b> requesterA@innovate.com	<b>Bestellnummer</b> 4500017277	<b>Kunden Nr.</b> 463432	<b>Rechnungs Nr.</b> MN-10077	<b>Lieferdatum</b> 03.03.2009
---	------------------------------------	-----------------------------	----------------------------------	----------------------------------

Wir bedanken uns für Ihren Auftrag und stellen Ihnen folgende Positionen in Rechnung:

Pos.	Menge	Artikelbezeichnung	Einheit	Einzelpreis	Gesamtpreis
10	2	Antrieb	Stk.	102,30	204,60 €
20	1	Rohling für Welle	Stk.	40,00	40,00 €
30	5	Kugellager	Stk.	9,20	46,00 €
40	2	Batterie	Stk.	102,30	204,60 €
50	1	Schrauben	Stk.	40,00	40,00 €
60	2	Bremse	Stk.	102,30	204,60 €
				<b>Summe</b>	739,80 €
				<b>zzgl 19% MWSt</b>	140,56 €
				<b>Rechnungsbetrag</b>	<b>880,36 €</b>

Der Rechnungsbetrag ist in 30 Tagen fällig. Wir weisen daraufhin, dass die oben aufgeführten Artikel bis zur vollständigen Bezahlung in unserem Eigentum verbleiben

**Postanschrift**  
C.E.B. Berlin  
Kolping Str. 15  
10115 Berlin  
Deutschland

**Sitz der Gesellschaft**  
C.E.B. New York  
7890 Broad Street  
New York, NY 10011, USA

**Vorstand:**  
Ralf Bauer  
Lydia Schmidt  
Thomas Schiffer  
Chris Walker

**Bankverbindung**  
Konto-Nr.:  
90010000  
BLZ: 100 500 33  
Sparkasse Berlin

**Ust-IdNr.**  
DE322156687  
Handelsregister  
Amtsgericht Berlin

[www.opentext.com](http://www.opentext.com)

Download data:

ORDERID	POSITION	ITEMTYPE	AMOUNT	DESCRIPTION	QUANTITY	UNITPRICE	UNIT
4500017277	00001	1	204.60	Bremse	2.000	102.30	ST
4500017277	00002	1	40.00	Schrauben	1.000	40.00	ST
4500017277	00003	1	46.00	Lager	5.000	9.20	ST
4500017277	00004	1	204.60	22 Antrieb xyz234	2.000	102.30	ST
4500017277	00005	1	40.00	Rohling Welle	1.000	40.00	ST
4500017277	00006	1	24.00	Muttern	20.000	1.20	ST
4500017277	00007	1	204.60	Batterie	2.000	102.30	ST

Results:

- Invoice item 30 maps for QUANTITY, UNITPRICE, AMOUNT completely with PO item 00003 and is assigned without considering DESCRIPTION.
- Download items 00001, 00004, 00007 and Download items 00002 and 00005 are grouped as they show identical triples QUANTITY, UNITPRICE, AMOUNT.
- The triple 2.000, 102.30, 204.60 is found three times on the invoice (invoice items 10, 40, 60). Assignment is done using item description text: 10 ->00004, 40->00007, 60->00001. For assignment it is sufficient to find the same words in PO descriptions and invoice descriptions, independent of the word sequence.

In the same way positions 20 and 50 are assigned to PO items 00005 and 00002.

#### Specific exceptions:

No Company Code determined

If no CC could be determined, PO line item mapping will not work for the invoice. Depending on settings for line item recognition, line items will or will not be exported to VIM. If exported to VIM, fields PO item number and PO number will be empty.

#### No vendor determined

If no vendor could be determined, PO line item mapping will not work for the current invoice. Depending on settings for line item recognition, line items will or will not be exported to VIM. If exported to VIM, fields PO item number and PO number will be empty.

#### No PO number found

If no PO numbers have been determined for the invoice, all PO/DN items for all PO/DN numbers of the relevant vendor may be considered for mapping. This is the default setting, which is recommended to deactivate, especially if in customer environment same vendor sends PO and nonPO invoices.

[www.opentext.com](http://www.opentext.com)

## 6. Tips for successful use of invoice line item mapping

Understanding how line item mapping is working, for every customer environment it should be checked, if ICC line item mapping can be used successfully. This is important, as line item mapping may abuse validation resources in case it doesn't work properly.

- Is PO number recognition working with high recognition rate? Good PO/DN recognition will enable filtering of the right items from data base for mapping. Same for vendor recognition.
- Does a good relation between PO/GR data and invoice data exist to enable line item mapping? If PO/GR data do not map at all with invoice data, line item mapping will not work. This may be the case for invoices with discounts or additional costs, as, in standard, discount and expense items are currently not considered for mapping.
- For framework contracts, data mapping will only succeed, if goods receipt data exist.
- It should also be considered how many data make sense to be downloaded using the PO download report: Define carefully selection parameters for download report: for most business scenarios selection parameters should be applied as follows:
  - Open POs: download only PO data from PO which are not completed
  - Delta report: The first report should run for complete PO data of the selected period, later on delta download should be applied for several weeks. A full download should be applied regularly e.g. once a month for consolidating data in VIM staging table and ICC data base.
  - Time period: the time period should be selected as short as possible – ideally several months, not several years. This reduces running time for collecting PO data in VIM staging table, and also reduces time to import PO data from staging table into ICC data base.

## About Open Text

Open Text is a leader in Enterprise Content Management (ECM). With two decades of experience helping organizations overcome the challenges associated with managing and gaining the true value of their business content, Open Text stands unmatched in the market.

Together with our customers and partners, we are truly The Content Experts,™ supporting 46,000 organizations and millions of users in 114 countries around the globe. We know how organizations work. We have a keen understanding of how content flows throughout an enterprise, and of the business challenges that organizations face today.

It is this knowledge that gives us our unique ability to develop the richest array of tailored content management applications and solutions in the industry. Our unique and collaborative approach helps us provide guidance so that our customers can effectively address business challenges and leverage content to drive growth, mitigate risk, increase brand equity, automate processes, manage compliance, and generate competitive advantage. Organizations can trust the management of their vital business content to Open Text, The Content Experts.

[www.opentext.com/\[solution group URL\]](http://www.opentext.com/[solution group URL])

**Sales:** [email address]  
[phone number]

**Support:** [email address]  
[phone number]

[www.opentext.com](http://www.opentext.com)

