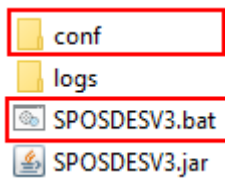


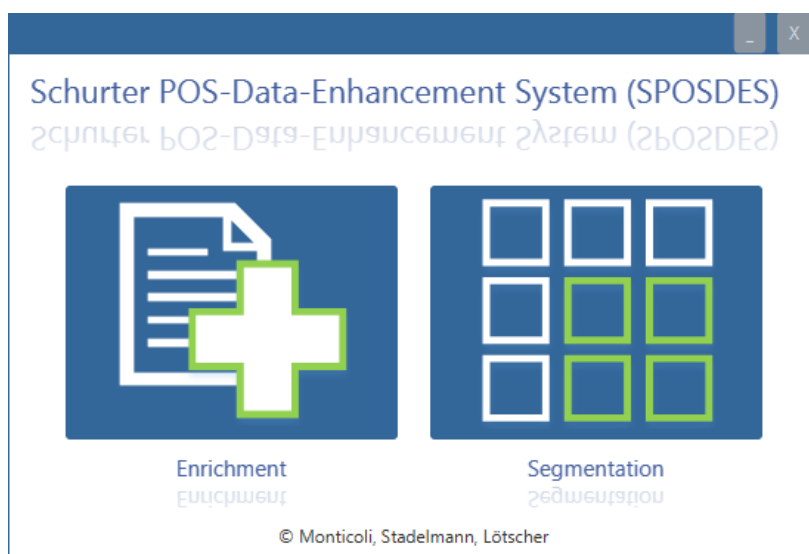
# User Manual SPOSDDES V3.0

## Run SPOSDDES



To run SPOSDDES simply double click on **SPOSDDESv3.bat**  
Make sure the .bat file and the .jar file have the name “SPOSDDESv3” and are located in the same folder.  
Furthermore, the folder named “conf” must be in the same folder as the .bat-file.

## Startmenu



After starting the .bat file the first window opens with the options of either enrich or segment data.

<b>Enrichment</b>	Internet connection is <b>required</b> .
<b>Segmentation</b>	<b>No requirements</b> in this phase.

## Segmentation

### 1. POS-file

Point-of-Sales Excel-file that contains at least the customer ID and the customer name

Format requirements:

	A	B	C	D	E	G	H
1	POS Customer ID						
2	Jan 2014 - Jul 2015						
3	POS Customer ID	Country	ZIP Code	Customer Name short	Customer Name	Customer Ident. Lvl	
4	AE10966ABUDHAB	AE Utd.Arab Emir.	109667	ABUDHAB	ABU DHABI AUTONOMOUS SYSTEMS INVEST	Country+PC+Name	

There are two major format requirements for the Excel:

1. The **POS Customer ID** list must start in cell **A4**
2. The **Customer Name** list must start in cell **G4**

Furthermore, the POS-file has to be an ".xlsx" Excel-file. The program is unable to work with ".xls" files. A possible error can appear in the later deduplication process if the file was converted from a ".xls" file to a ".xlsx".

To make it 100% certain that no errors happen:

**Take all the POS Data and copy it into a new ".xlsx" Excel-file.**

### 2. Industry segmentation file

Excel-file that contains the company names with their industry assignment.

Format requirements:

	A	B
1	Medical	12635 E. Montview Blvd.
2	Food Service	Zummo Inc.
3	Audio	Morgan Hill Music
4	Medical	180 Medical

There are two major format requirements for the Excel-sheet:

1. The **Industry Name** list must to start in cell **A1**
2. The **Company Name** list must start in cell **B1**

## Properties

**Data cleansing**  
**Remove special characters**  
☒ Removing Special Characters is activated  
**Deduplicate list**  
☒ Deduplicating is activated

---

**Segmentation**  
**Segmentation margin**

We highly recommend to keep the standard configuration as seen above.

### Remove special characters

If activated removes all special characters. Only letters from A to Z remain. This configuration helps to get a higher match rate for the segmentation.

### Deduplicate list

Deduplication will merge customers with the same customer ID but different customer name into one row.

**Requirement for the deduplication:** The POS-file must be sorted by the customer ID, else the deduplication process will no work properly.

Example of deduplication:

Situation: There are two customers with the same customer ID **CH6436MIKEMON** but different Customer names.

POS Customer ID	Country		ZIP Code	Customer Name short	Customer Name	Distance
CH6436MIKEMON	CH	Switzerland	6436	MIKEMON	MIKE MONTICOLI	0
CH6436MIKEMON	CH	Switzerland	6436	MIKEMON	MIKE MONTICOLI, LTD	0.1

If “Deduplicate list” is activated, SPOSDS merges the two rows into one as follows:

POS Customer ID	Country		ZIP Code	Customer Name short	Customer Name	Distance
CH6436MIKEMON	CH	Switzerland	6436	MIKEMON	MIKE MONTICOLI, MIKE MONTICOLI, LTD	0

Note:

- The row with the **higher distance** will be **removed** and the customer name of this row assigned to the row with the lower distance.

### Segmentation margin

We use the Jaro-Winkler distance for measuring the similarity of the POS-file customer names and the company names in the segmentation file.

**The lower the distance, the more similar are the compared strings.**

The Industry names will be assigned to the customers if the distance is equals or lower the segmentation margin.