



Suche

11.0  
11.0



# Impressum

---

© 1998 – 2019 Comitas AG Schweiz. Alle Rechte vorbehalten.

Suche mit Intrapact 11.0, 1. Auflage, 2019.

Eine Vervielfältigung – auch von Auszügen – in jeglicher Weise bedarf der vorherigen ausdrücklichen Genehmigung von der Comitas AG.

Alle in diesem Dokument verwendeten Markennamen und Bezeichnungen unterliegen im Allgemeinen warenzeichen-, marken- oder patentrechtlichen Schutzbestimmungen. Microsoft und Adobe sind eingetragene Warenzeichen.

# Inhaltsverzeichnis

---

<b>Impressum</b> .....	<b>2</b>
<b>Inhaltsverzeichnis</b> .....	<b>3</b>
<b>1. Grundsätzliches</b> .....	<b>4</b>
<b>2. Funktionsweise des Comitas Indexing Service (CIS)</b> .....	<b>4</b>
<b>3. Suchmöglichkeiten in Intrapact</b> .....	<b>5</b>
3.1 Globale Suche .....	5
3.2 Suche innerhalb der Anwendung .....	5
3.3 Suche im Dokumenten-Center .....	5
3.4 Suche im FileWalker .....	5
<b>4. Suchanfragen</b> .....	<b>6</b>
4.1 Grundsätzliche Suche in der Datenbank und den Dokumenten .....	6
4.2 Zusätzliche Suchmöglichkeiten in den Dokumenten .....	6
4.2.1 <i>Allgemeines</i> .....	6
4.2.2 <i>Fuzzy-Suche</i> .....	7
4.2.3 <i>Phonetische Suche</i> .....	7
4.2.4 <i>Suche mit grammatischen Variationen</i> .....	7
4.2.5 <i>Zahlenbereich-Suche</i> .....	8
4.2.6 <i>Feld-Suche</i> .....	8
4.2.7 <i>Segmentierte Suche</i> .....	8
<b>4. Indizierung: Unterstützte Dateiformate bei den hochgeladenen Dateien</b> .....	<b>9</b>
<b>5. Indizierung: Standardfelder, die indiziert werden</b> .....	<b>10</b>

# Suche mit Intrapact 11.0

## 1. Grundsätzliches

---

Eine der wichtigsten Fähigkeiten eines Portals wie Intrapact ist es, Daten, welche in dieses Portal eingegeben worden sind, wieder finden zu können. In Intrapact werden Daten einerseits in Form von Datensätzen in einer Datenbank und andererseits in – auf den Webserver hochgeladenen – Dokumenten gespeichert. Für die Suche heißt das, dass einerseits die Datenbank über SQL-Abfragen und andererseits die Dokumente über eine Volltext-Suchmöglichkeit durchsucht werden müssen. Bei diesen beiden Fällen ist die Suche in den Dokumenten die komplexere, da die Daten unstrukturiert sind (in reiner Textform mit eventuellen Steuerzeichen dazwischen) und somit über eine intelligente Indizierung aufbereitet werden müssen. Ohne Indizierung müsste jedes Dokument Zeichen für Zeichen, Wort für Wort durchsucht werden, was extrem zeitaufwendig, und bei Dokumentbeständen von mehreren Giga- bzw. Terabyte nicht mehr durchführbar wäre.

Intrapact hat schon immer eine Indizierungstechnologie verwendet: die Microsoft Indexing Services. Diese waren aber ziemlich eingeschränkt, weshalb auch Microsoft mit Windows Server 2012 diese gar nicht mehr anbietet und auch nicht mehr unterstützt. Seit Intrapact 9 wurde deswegen auf eine neue Suchmaschine (**dtsearch**) umgestellt, welche unabhängig vom Betriebssystem ist. Dtsearch ist eines der besten Tools auf dem Markt und sehr schnell aus Terabytes von Daten die gewünschten Dokumente lokalisieren zu können, welche die gesuchten Wörter beinhalten. Um diese Geschwindigkeit zu erreichen, müssen die Daten zuvor intelligent indiziert werden.

Für die Indizierung wurde ein eigener Service programmiert: **Comitas Indexing Service (CIS)**. Dieser Service indiziert unter Verwendung von dtsearch genau die Daten, die im Intrapact benötigt werden. Auf diesen Index greift dann die Suche von Intrapact zu, um die Daten schnell finden zu können.

## 2. Funktionsweise des Comitas Indexing Service (CIS)

---

Damit CIS weiß, was alles zu indizieren ist, gibt es unter `\intrapact\bin` die Datei `intrapact.ilb`, in welche jeder, von Intrapact benötigte, Index eingetragen wird. Diese Datei sollten Sie nur dann manuell anpassen, wenn Sie genau wissen, was zu tun ist. Ansonsten kann es sein, dass die Indexerstellung nicht mehr richtig funktioniert!

CIS arbeitet diese Datei zyklisch ab und erstellt anhand der Einträge die benötigten Suchindizes. Für die Abarbeitung gibt es noch Steuerparameter, die in der Windows Registry hinterlegt sind. Dort gibt es bei 32-Bit-Windows-Systemen unter `HKEY_LOCAL_MACHINE\Software\United Planet\Intrazone\1.00` die Parameter „IndexingSchedule“ und „IndexingTracing“. Bei 64-Bit-Systemen liegen die Parameter unter

`HKEY_LOCAL_MACHINE\Software\Wow6432Node\United Planet\Intrazone\1.00`.

**IndexingSchedule** gibt die Zeit in Minuten an, wann zyklisch eine Indizierung gemacht werden soll. Standardmäßig sind 20 Minuten eingestellt. Wenn bei Ihnen täglich nicht viele Dateien hochgeladen werden, kann man hier diesen Wert auf z.B. 300 stellen (= alle 5 Stunden). Für die Aktivierung muss danach der Dienst neu gestartet werden. Am besten über die Intrapact-Konsole im System-Tray neben der Uhrzeit ganz unten rechts am Desktop. Wichtig ist dabei zu wissen, dass der 1. Lauf nicht sofort ausgeführt wird, sondern erst nach dem einmaligen Verstreichen der angegebenen Zeit (in unserem Beispiel also nach 5 Stunden).

**IndexingTracing** kann folgende Parameterwerte erhalten (nach Änderung wiederum Neustart des Service):

- 0 = Schreibe nichts in die Windows Ereignisanzeige (Standardeinstellung)
- 1 = Schreibe eine ausführliche LOG-Datei (ACHTUNG: die kann tatsächlich mehrere 100 Seiten groß werden!) in das `\intrapact\bin` Verzeichnis mit allen genauen Indizierungs-Informationen
- 2 = Schreibe Start-/Endzeitpunkt und welche Firma indiziert wird für Kontrollzwecke in die Windows-Ereignisanzeige

### 3. Suchmöglichkeiten in Intrapact

---

In Intrapact kann an vier Stellen gesucht werden. Unabhängig davon, welche Suche verwendet wird, werden überall nur die Daten bzw. Dokumente als Suchtreffer angezeigt, für welche der Suchende auch eine Berechtigung hat.

Folgend die vier Such-Möglichkeiten:

#### 3.1 Globale Suche

Diese befindet sich auf der Startseite und es können Daten über die komplette Firma im Intrapact gesucht werden.

#### 3.2 Suche innerhalb der Anwendung

Innerhalb jeder Anwendung kann ein Suchformular implementiert werden, welches nur Daten und Dokumente aus dieser Anwendung als Ergebnis liefert.

#### 3.3 Suche im Dokumenten-Center

Alle Dokumente, die im Dokumenten-Center hochgeladen werden, können auch dort gesucht werden. Das heißt, dass eine strikte Trennung zwischen den anwendungsspezifischen Daten und Dokumenten und den Dokumenten im Dokumenten-Center erfolgt. Für das Dokumenten-Center gibt es auch noch ein Portlet auf der Hauptseite, über welches auch nach Dokumenten gesucht werden kann.

#### 3.4 Suche im FileWalker

Über den FileWalker im Intrapact Manager können freigegebene Windows-Verzeichnisse (sogenannte „Shares“), die vom Intrapact-Server aus erreichbar sein

müssen, auf einfache Art und Weise ins Intrapact eingebunden werden. Damit können diese Dokumente dann über das Web gesucht, heruntergeladen und verwendet werden.

Bei diesen vier Suchmöglichkeiten ist die Suche selber immer identisch implementiert. Man kann einen Suchtext in ein Eingabefeld eingeben und dann auf „Suchen“ klicken. Fertig!

## 4. Suchanfragen

---

Bis auf die FileWalker-Suche wird der eingegebene Suchtext einerseits dazu verwendet, um Daten in der Datenbank zu suchen und andererseits wird derselbe Suchtext für die Volltextsuche in den Dokumenten verwendet. Falls einige „Spezial-Parameter“ der Volltextsuche verwendet werden, kann es deswegen dazu kommen, dass in der Datenbanksuche einiges nicht gefunden wird, da die Datenbanksuche davon ausgeht, dass diese „Spezial-Parameter“ auch Daten in der Datenbank darstellen. Dieses derzeitige Verhalten wird einmal verfeinert werden.

### 4.1 Grundsätzliche Suche in der Datenbank und den Dokumenten

Durch **Eingabe EINES Suchwortes** wird genau dieses Wort gesucht (Bsp: Auto).

Will man mehrere Wörter zusammenhängend suchen, müssen sie unter doppelten Hochkommas eingeben werden (Bsp: "Auto mit Anhänger"). Wörter können auch mit **and, or, not** zu einer logischen Suche verknüpft werden (Bsp: Auto and Pickup not Ford).

Will man **nach Wortfragmenten suchen**, das heißt das Suchwort soll auch als Teil eines zusammengesetzten Wortes gefunden werden, so kann man das Ersetzungszeichen „\*“ (**Stern**) verwenden. „\*“ ersetzt dabei beliebig viele Zeichen (Bsp: rad\* findet Radiergummi, Radieschen, Radmutter usw. aber nicht Fahrrad! \*rad\* findet Fahrrad, Fahrradpumpe usw.). Die Groß-/Kleinschreibung wird in der Suche standardmäßig nicht berücksichtigt.

Neben dem „\*“ gibt es noch folgende 2 spezielle Ersetzungszeichen:

? genau EIN beliebiges Zeichen kann auf dieser Position stehen (Bsp: H?se für Hose, Hase, ...)

= genau EINE beliebige Ziffer kann auf dieser Position stehen (Bsp: 201= für 2010, 2011, 2012, ...)

### 4.2 Zusätzliche Suchmöglichkeiten in den Dokumenten

#### 4.2.1 Allgemeines

Bei der Suche in den Dokumenten kann über spezielle Zeichen auch das Suchverhalten gesteuert werden, und zwar gibt es dabei folgende Möglichkeiten:

- % es handelt sich um eine „Fuzzy-Suche“
- # es handelt sich um eine „phonetische Suche“
- ~ Sucht nach grammatischen Variationen
- ~~ Suche innerhalb eines Zahlenbereiches
- ## Suche über einen regulären Ausdruck

Wie oben bereits angedeutet, gibt es bei der Suche in den Dokumenten erweiterte Parameter. Diese sind speziell bei der Suche im Doku-Center und im FileWalker sehr nützlich.

Apfel w/5 Birne vorkommen	„Apfel“ muss innerhalb von 5 Wörtern mit „Birne“
Apfel not w/10 Birne	„Apfel“ muss vorkommen, aber nicht innerhalb von 10 Wörtern mit „Birne“
filename contains Apfel	Das Feld „Filename“ muss die Zeichenfolge „Apfel“ beinhalten
Apfel w/3 xfirstword	„Apfel“ muss in den ersten 3 Wörtern vorkommen
Apfel w/3 xlastword	„Apfel“ muss in den letzten 3 Wörtern vorkommen

Gibt man mehrere Suchkriterien an, ist es sinnvoll, das ganze so zu klammern, wie man den Ausdruck gerne haben möchte. Beispielsweise liefern die zwei Abfragen unterschiedliche Ergebnisse:

(Apfel and Birne) or Orange      bzw. Apfel and (Orange or Birne)

#### 4.2.2 Fuzzy-Suche

Mit "Fuzzy-Suche" kann man auch Wörter finden, wenn sie falsch geschrieben werden. Beispielsweise wird bei einer Fuzzy-Suche nach „Apfel“ auch „Appppfel“ gefunden. Für die Fuzzy-Suche wird das %-Zeichen verwendet. Die Anzahl der %-Zeichen gibt an, wie viele Unterschiede im gefundenen Wort zum Suchwort sein dürfen. Die Position gibt an, ab wann diese Unterschiede sein dürfen.

Beispiel: Ap%%fel gibt an, das alle Wörter mit "Ap" beginnen müssen und dann dürfen maximal 3 Unterschiede zum Wort Apfel sein. Appppfel, Apfeeel, Apfelll, Apffeell würde alles gefunden werden.

#### 4.2.3 Phonetische Suche

Mit der phonetischen Suche findet man auch Wörter die so ähnlich klingen wie das gesuchte Wort. Für die phonetische Suche stellt man ein #-Zeichen vor das Suchwort.

Beispiel: #Schmid findet auch Schmit, Schmitt, Schmied, Schmidt

#### 4.2.4 Suche mit grammatischen Variationen

Grammatische Variationen sind zum Beispiel für das Wort „Fisch“: „fischen“, „Fischer“, usw. Welche Variationen hier unterstützt werden sollen, muss man sich in einer Datei

stemming.dat definieren, die am Intrapact-Server im Verzeichnis `\intrapact\bin` zu speichern ist. In dieser Datei ist auch ein Kommentar enthalten, wie die Regeln zu definieren sind.

#### 4.2.5 Zahlenbereich-Suche

Eine Zahlenbereich-Suche findet Treffer von Zahlen innerhalb eines definierten Bereiches, wobei die angegebenen Bereichsgrenzen auch als Treffer gelten.

Beispiel: `Apfel w/5 10~15` findet Dokumente, bei denen das Wort „Apfel“ innerhalb von 5 Worten bei einer Zahl zwischen 10 und 15 vorkommt.

#### 4.2.6 Feld-Suche

Alle Dokumente werden mit ihren Standardfeldern (siehe unten) indiziert. Das heißt, dass nicht nur der Text im Dokument indiziert wird, sondern auch Standardfelder wie Autor, Name, Beschreibung, etc. Microsoft, Adobe etc. bezeichnen ihre internen Felder immer mit der englischen Bezeichnung.

Beispiel:

`(filename contains (Protokoll1 or Protokoll2)) and (Author contains („Gottfried Stutz“))`

Dazu gibt es auch eine alternative Schreibform:

`(Filename::Protokoll1 or filename::Protokoll2) and (Author::"Gottfried Stutz")`

#### 4.2.7 Segmentierte Suche

Es lassen sich auch virtuelle Felder in einem Dokument erstellen, indem Beginn- und Endwörter definiert werden, welche den Suchtext begrenzen müssen. Wenn Sie beispielsweise so ein Konstrukt in Ihrer Datei haben:

Name: Gottfried Stutz  
Adresse: Zürcherstrasse 12  
Ort: Zürich

dann können Sie definieren, dass das Suchfeld zwischen den Wörtern „Name“ und „Adresse“ liegt (die „:“ sind dabei egal), und eine Suchanfrage mit

`(name to adresse) contains „Gottfried Stutz“`

würde nur mehr alle Dokumente liefern, die genau dort im Dokument „Gottfried Stutz“ stehen haben. In den „Grenzwörtern“ darf als Operator nur ein „or“ vorkommen oder es muss genau 1 Wort sein. Also folgende Abfrage

`(name to (adresse or xlastword)) contains „Gottfried Stutz“`

Würde alle Dokumente liefern, wo „Gottfried Stutz“ nach Name: und vor Adresse: oder dem letzten Wort des Dokuments vorkommt.



## 4. Indizierung: Unterstützte Dateiformate bei den hochgeladenen Dateien

---

Für folgende Dateiformate kann ein Index generiert werden, damit die Datei durchsucht werden kann:

Adobe Acrobat (\*.pdf)  
Adobe Framemaker MIF (\*.mif)  
Ami Pro (\*.sam)  
Ansi Text (\*.txt)  
ASCII Text  
ASF media files (metadata only) (\*.asf)  
CSV (Comma-separated values) (\*.csv)  
DBF (\*.dbf)  
EBCDIC  
EML files (emails saved by Outlook Express) (\*.eml)  
Enhanced Metafile Format (\*.emf)  
Eudora MBX message files (\*.mbx)  
Flash (\*.swf)  
GZIP (\*.gz)  
HTML (\*.htm, \*.html)  
Ichitaro (versions 5 and later) (\*.jtd, \*.jbw)  
JPEG (\*.jpg)  
Lotus 1-2-3 (\*.123, \*.wk?)  
MBOX email archives such as Thunderbird, including attachments (\*.mbx)  
MHT archives (HTML archives saved by Internet Explorer) (\*.mht)  
MIME messages, including attachments  
MSG files (emails saved by Outlook), including attachments (\*.msg)  
Microsoft Access MDB files (\*.mdb, \*.accdb, including Access 2007 and Access 2010)  
Microsoft Document Imaging (\*.mdi)  
Microsoft Excel (\*.xls)  
Microsoft Excel 2003 XML (\*.xml)  
Microsoft Excel 2007 and 2010 (\*.xlsx)  
Microsoft OneNote 2007 and 2010 (\*.one)  
Microsoft Outlook data files, including attachments (\*.PST)  
Microsoft Outlook/Exchange Messages, Notes, Contacts, Appointments, and Tasks  
Microsoft Outlook Express 5 and 6 (\*.dbx) message stores  
Microsoft PowerPoint (\*.ppt)  
Microsoft PowerPoint 2007 and 2010 (\*.pptx)  
Microsoft Rich Text Format (\*.rtf)  
Microsoft Searchable Tiff (\*.tiff)  
Microsoft Word for DOS (\*.doc)  
Microsoft Word for Windows (\*.doc)  
Microsoft Word 2003 XML (\*.xml)  
Microsoft Word 2007 and 2010 (\*.docx)  
Microsoft Works (\*.wks)  
MP3 (metadata only) (\*.mp3)  
Multimate Advantage II (\*.dox)

Multimate version 4 (\*.doc)  
 OpenOffice versions 1, 2, and 3 documents, spreadsheets, and presentations (\*.sxc, \*.sxd, \*.sxi, \*.sxw, \*.sxcg, \*.stc, \*.sti, \*.stw, \*.stm, \*.odt, \*.ott, \*.odg, \*.otg, \*.odp, \*.otp, \*.ods, \*.ots, \*.odf) (includes OASIS Open Document Format for Office Applications)  
 Quattro Pro (\*.wb1, \*.wb2, \*.wb3, \*.qpw)  
 QuickTime (\*.mov, \*.m4a, \*.m4v)  
 RAR (\*.rar)  
 TAR (\*.tar)  
 TIFF (metadata only) (\*.tif)  
 TNEF (winmail.dat files)  
 Treepad HJT files (\*.hjt)  
 Unicode (UCS16, Mac or Windows byte order, or UTF-8)  
 Visio XML files (\*.vdx)  
 Windows Metafile Format (\*.wmf)  
 WMA media files (metadata only) (\*.wma)  
 WMV video files (metadata only) (\*.wmv)  
 WordPerfect 4.2 (\*.wpd, \*.wpf)  
 WordPerfect (5.0 and later) (\*.wpd, \*.wpf)  
 WordStar version 1, 2, 3 (\*.ws)  
 WordStar versions 4, 5, 6 (\*.ws)  
 WordStar 2000  
 Write (\*.wri)  
 XBase (including FoxPro, dBase, and other XBase-compatible formats) (\*.dbf)  
 XML (\*.xml)  
 XML Paper Specification (\*.xps)  
 XSL  
 XyWrite  
 ZIP (\*.zip)

Verschachtelte Objekte (z.B. ein Excel Datenblatt, eingebettet in eine PowerPoint-Präsentation) werden immer als Teil des Hauptobjektes angegeben, also Treffer, im Beispiel, in der Excel-Datei, werden als Treffer innerhalb der PowerPoint-Datei dargestellt.

## 5. Indizierung: Standardfelder, die indiziert werden

Dateiformat	Felder
Email Dateien (Outlook Express, Eudora, MBOX, EML)	An, CC, BCC, Von, Senden über, Sender, Empfänger, Betreff, Datum, Anhang
Outlook Felder und .MSG Dateien	An, CC, BCC, Von, Senden über, Sender, Empfänger, Betreff, Datum, Übermittlungsdatum, Anhänge, Kontaktfelder (Ortsfelder, Firmenname, etc.)
Microsoft Word, Excel, PowerPoint	Dokumenten Zusammenfassungsfelder
OpenOffice/Open Document Format	Dokumenten Eigenschaftsfelder
HTML	META tags; <TITLE> ist indiziert als HtmlTitle Feld; <H1>, <H2>, <H3> sind indiziert als HtmlH1, HtmlH2, HtmlH3, etc.

XML	Alle Felder
DBF	Alle Felder
CSV	Alle Felder (CSV, Dateien müssen eine .csv Erweiterung, eine Liste von Feldnamen in der ersten Zeile haben und TAB, Semikolon oder Komma als Trenner haben)
PDF Dateien	Dokumenten Eigenschaften
WordPerfect	Dokumenten Zusammenfassungsfelder
MP3	Alle Metadaten Felder
JPG, TIFF	EXIF und IPTC Metadaten Felder; XMP (Vista) Metadaten welche in Version 7.40 unterstützt werden
ASF, WMA, WMV	Alle Metadaten Felder