



Informatica® Identity Resolution 10.1

Introduction Guide

This software and documentation are provided only under a separate license agreement containing restrictions on use and disclosure. No part of this document may be reproduced or transmitted in any form, by any means (electronic, photocopying, recording or otherwise) without prior consent of Informatica LLC.

U.S. GOVERNMENT RIGHTS Programs, software, databases, and related documentation and technical data delivered to U.S. Government customers are "commercial computer software" or "commercial technical data" pursuant to the applicable Federal Acquisition Regulation and agency-specific supplemental regulations. As such, the use, duplication, disclosure, modification, and adaptation is subject to the restrictions and license terms set forth in the applicable Government contract, and, to the extent applicable by the terms of the Government contract, the additional rights set forth in FAR 52.227-19, Commercial Computer Software License.

Informatica and the Informatica logo are trademarks or registered trademarks of Informatica LLC in the United States and many jurisdictions throughout the world. A current list of Informatica trademarks is available on the web at <https://www.informatica.com/trademarks.html>. Other company and product names may be trade names or trademarks of their respective owners.

The information in this documentation is subject to change without notice. If you find any problems in this documentation, report them to us at infa_documentation@informatica.com.

Informatica products are warranted according to the terms and conditions of the agreements under which they are provided. INFORMATICA PROVIDES THE INFORMATION IN THIS DOCUMENT "AS IS" WITHOUT WARRANTY OF ANY KIND, EXPRESS OR IMPLIED, INCLUDING WITHOUT ANY WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND ANY WARRANTY OR CONDITION OF NON-INFRINGEMENT.

Publication Date: 2018-09-26

Table of Contents

- Preface 4**
 - Learning About Informatica Identity Resolution. 4
 - What Do I Read If. 5
 - Informatica Resources. 5
 - Informatica Network. 5
 - Informatica Knowledge Base. 6
 - Informatica Documentation. 6
 - Informatica Product Availability Matrixes. 6
 - Informatica Velocity. 6
 - Informatica Marketplace. 6
 - Informatica Global Customer Support. 6
- Chapter 1: Overview..... 8**
- Chapter 2: Terminology..... 10**

Preface

The *IIR Introduction Guide* provides an overview of the purpose of Informatica Identity Resolution product.

Learning About Informatica Identity Resolution

This section provides details of documentation available with the Informatica Identity Resolution product.

Introduction Guide

Introduces Identity Resolution and its related terminology. It may be read by anyone with no prior knowledge of the product who requires a general overview of Identity Resolution.

Installation Guide

This manual is intended to be the first technical material a new user reads before installing the Identity Resolution software, regardless of the platform or environment.

Design Guide

This is a guide that describes the steps needed to design, define and load an Identity Resolution "System".

Developer Guide

This manual describes how to develop a custom search client application using the Identity Resolution API.

Operations Guide

This manual describes the operation of the run-time components of Identity Resolution, such as servers, search clients and other utilities.

Populations and Controls Guide

This manual describes SSA-Name3 populations and the controls they support. The latter are added to the Controls statement used within an IDX-Definition or Search-Definition section of the SDF.

Release Notes

The Release Notes contain information about what's new in this version of Identity Resolution. It is also summarizes any documentation updates as they are published.

What Do I Read If. . .

I am. . .

. . . a business manager

The INTRODUCTION to Identity Resolution will address questions such as "Why have we got Identity Resolution?", "What does Identity Resolution do"?

I am. . .

. . . installing the product?

Before attempting to install IIR you should read the INSTALLATION GUIDE to learn about the prerequisites and to help you plan the installation and implementation of the Identity Resolution.

I am. . .

...an Analyst or Application Programmer?

A high-level overview is provided specifically for Application Programmers in the INTRODUCTION to Identity Resolution.

When designing and developing the application programs, refer to the DEVELOPER GUIDE which describes a typical application process flow and API parameters. Working example programs that illustrate the calls to IIR in various languages are available under the <IIR_client_installation>/samples directory.

I am. . .

...designing and administering Systems?

The process of designing, defining and creating Systems is described in the DESIGN GUIDE. Administering the servers and utilities is described in the OPERATIONS manual.

Informatica Resources

Informatica Network

Informatica Network hosts Informatica Global Customer Support, the Informatica Knowledge Base, and other product resources. To access Informatica Network, visit <https://network.informatica.com>.

As a member, you can:

- Access all of your Informatica resources in one place.
- Search the Knowledge Base for product resources, including documentation, FAQs, and best practices.

- View product availability information.
- Review your support cases.
- Find your local Informatica User Group Network and collaborate with your peers.

Informatica Knowledge Base

Use the Informatica Knowledge Base to search Informatica Network for product resources such as documentation, how-to articles, best practices, and PAMs.

To access the Knowledge Base, visit <https://kb.informatica.com>. If you have questions, comments, or ideas about the Knowledge Base, contact the Informatica Knowledge Base team at KB_Feedback@informatica.com.

Informatica Documentation

To get the latest documentation for your product, browse the Informatica Knowledge Base at https://kb.informatica.com/_layouts/ProductDocumentation/Page/ProductDocumentSearch.aspx.

If you have questions, comments, or ideas about this documentation, contact the Informatica Documentation team through email at infa_documentation@informatica.com.

Informatica Product Availability Matrixes

Product Availability Matrixes (PAMs) indicate the versions of operating systems, databases, and other types of data sources and targets that a product release supports. If you are an Informatica Network member, you can access PAMs at

<https://network.informatica.com/community/informatica-network/product-availability-matrixes>.

Informatica Velocity

Informatica Velocity is a collection of tips and best practices developed by Informatica Professional Services. Developed from the real-world experience of hundreds of data management projects, Informatica Velocity represents the collective knowledge of our consultants who have worked with organizations from around the world to plan, develop, deploy, and maintain successful data management solutions.

If you are an Informatica Network member, you can access Informatica Velocity resources at <http://velocity.informatica.com>.

If you have questions, comments, or ideas about Informatica Velocity, contact Informatica Professional Services at ips@informatica.com.

Informatica Marketplace

The Informatica Marketplace is a forum where you can find solutions that augment, extend, or enhance your Informatica implementations. By leveraging any of the hundreds of solutions from Informatica developers and partners, you can improve your productivity and speed up time to implementation on your projects. You can access Informatica Marketplace at <https://marketplace.informatica.com>.

Informatica Global Customer Support

You can contact a Global Support Center by telephone or through Online Support on Informatica Network.

To find your local Informatica Global Customer Support telephone number, visit the Informatica website at the following link:

<http://www.informatica.com/us/services-and-training/support-services/global-support-centers>.

If you are an Informatica Network member, you can use Online Support at <http://network.informatica.com>.

CHAPTER 1

Overview

This manual gives an overview of the purpose of Informatica Identity Resolution™ (IIR) and introduces IIR terminology.

Identity Resolution

"Out of the Box" Software

Identity Resolution (IIR) is "out of the box" software to add search and matching applications to identity data stored in databases and DataWarehouses. On-line and Batch Identity Search and Matching Clients are provided for LAN, Intranet and Web Systems.

Without programming or change to existing databases this software provides ready to use on-line inquiry and batch processing of identification data about people, companies, addresses and products. Very high quality results are achieved regardless of the quality, format or variation of the user's data.

Custom Clients can be built in a variety of other development languages.

Overcomes spelling and word variation

When searching and matching names and addresses and other identity data, IIR overcomes spelling, keying, word variation and sequence errors. It addresses the fact that transactions often contain repeated identity fields such as prior names and addresses, aliases, compound names, and multiple identity or phone numbers. The matching of dates and codes compensates for the particular error and variation found in this class of data.

Data does not require cleaning or enhancement

The user's source data can be formatted or unformatted and does not require cleaning or enhancement. IIR supports data from all countries, in local character sets.

Achieves productive results quickly

With initial specialist help, productive results can be achieved in the user's environment in a few days. Because of the rapid setup time, iterative tuning with users can be based upon actual results with production data, rather than the typical lengthy design, build, test and change process.

Large volumes of data can be processed

IIR is a highly scalable system that operates efficiently on servers running Unix and Windows. Large volumes of data can be processed and it is not constrained by operating system file size limits.

Rulebase Driven Implementation

During installation, the definition of the user's database tables is used to choose and customize the required search and matching features. These are then stored as meta-data rules in the user's database. The Rulebase is maintained with a GUI Rulebase editor.

High-Performance Indexes

High performance IIR Indexes are automatically generated and maintained with no change to existing business application programs or databases.

Identity Search Clients for LAN, Intranet and Web Applications

The IIR on-line and batch client applications provide high quality identity search, matching and duplicate finding despite the unavoidable error and variation in name, address and other identity data. The search clients are controlled by customizable rules so that they may be designed to suit the data and needs of the users.

The client applications can be used on local client terminals, LANs and as WEB applets. The user of the on-line client can enter the search data in any convenient format, and view the candidate records in a flexible table or fixed field format.

Additional searches can be conducted using the data from any record found, and a history of current searches is maintained. The batch client supports search and matching from a user's input transaction file; output can be to screen or file in a variety of formats.

Search Server

Search Server provides on-line search and matching services to applications used by multiple concurrent users.

Flexible Search Algorithms & Matching Schemes

The search algorithms use rich, highly compressed keys for names, addresses, titles or descriptions. The search strategies are calculated in real time, based on the data being searched for.

The search clients satisfy business needs that range from the extreme positive, where the search data is reliable and a close match is expected, to the extreme negative where it is essential to prove that there is no match on file.

The matching schemes used to choose candidate records are tuned and tested to mimic an expert human operator making identification decisions that suit the specific business purpose. Matching schemes for persons, households, companies, and all forms of transactions and documents are available.

CHAPTER 2

Terminology

The following terminologies are used in the IIR product:

Batch Search Client - DupFinder

Refers to the rule based search client provided with IIR to run duplicate searches within an IDT. Output is to a file or database table.

Batch Search Client - Relate

Refers to the rule-based batch C search client provided with IIR to run searches against the IIR Identity Indexes via the Search Server. Input is from a file or database tables. Output is written to a file or database table.

IIR Console

The IIR Console is used to administer the operational aspects of IIR, including System Loading, Table Loading, and the startup and shutdown of the servers. The IIR Console is comprised of the Console Server and Console Client.

IIR Identity Table (IDT)

An IIR Identity Table (IDT) is a table stored in the user's database containing denormalized and transformed matching and display data extracted by IIR from the User Source Table(s). The IDT is accessible using native SQL.

IIR Identity Index (IDX)

An IIR Identity Index (IDX) is a table stored in the user's database containing key information plus compressed matching and display data. Each IDX maps to a physical index (IDX_i). This is also known as a heavy index.

IIR Identity Link (IDL)

An IIR Identity Link (IDL) is an IIR table stored in the user's database containing relationships (links) between rows of the IDT.

IIR Nucleus

The IIR Nucleus is a term used to describe the collection of servers that provide the indexing, search and matching facilities of IIR.

IIR Search Server

The IIR Search Server provides multi-user local and remote access to IDT & IDX tables.

IIR Tables

Refers to all of the database tables and indexes created, used and maintained by IIR.

Lite Index

A Lite Index is a non-fuzzy index on an IIR Identity Table column. It has lower storage requirements than a heavy index but requires additional I/O when used in a search.

Online Search Client

Refers to the rule-based Java search client provided with IIR to run searches against the IIR Identity Indexes via the IIR Search Server. Input and output is to a screen.

Population file

Contains the SSA-NAME3 Key-building Algorithms, Search Strategies and Matching Schemes used by IIR.

Rulebase

The IIR Rulebase contains all of the information and rules necessary to run an IIR System. It is stored in tables on the user's database. It is accessed via the IIR Rulebase Server and is maintained by the Rulebase Client.

Source Database

The source database is the database containing the User Source Tables.

System

A system is the highest level logical definition in an IIR Rulebase. It can be thought of as corresponding to the user's view of a system, although this is not enforced. Many systems can be defined to the Rulebase.

System Definition File (SDF)

The SDF is a text file containing the rules for a System. It is loaded to the Rulebase via the System Loader.

System Loader

The System Loader is a set of utilities used to load an IIR System to the Rulebase. It is normally run from the IIR Console.

Table Loader

The Table Loader is a set of utilities used to extract, denormalize, transform and key data into the IIR Identity and Index Tables. It is normally run from the IIR Console.

Target Database

The target database is the database containing the IIR Tables and Indexes

Update Synchronizer

The Update Synchronizer is a program that actively monitors changes to the User Source Tables and synchronizes these changes to the IIR Identity Tables and Indexes.

User Service Group (USG)

Contains the SSA-NAME3 v1 Key-building Algorithms, Search Strategies and Matching Schemes used by IIR.

User Source Table (UST)

A User Source Table (UST) refers to user data stored in a table in the user's database containing the identity data to be searched and matched.