



Informatica® SSA-NAME3
10.0.0

Getting Started Guide

© Copyright Informatica LLC 1993, 2018

This software and documentation contain proprietary information of Informatica LLC and are provided under a license agreement containing restrictions on use and disclosure and are also protected by copyright law. Reverse engineering of the software is prohibited. 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. This Software may be protected by U.S. and/or international Patents and other Patents Pending.

Use, duplication, or disclosure of the Software by the U.S. Government is subject to the restrictions set forth in the applicable software license agreement and as provided in DFARS 227.7202-1(a) and 227.7702-3(a) (1995), DFARS 252.227-7013(1)(ii) (OCT 1988), FAR 12.212(a) (1995), FAR 52.227-19, or FAR 52.227-14 (ALT III), as applicable.

The information in this product or documentation is subject to change without notice. If you find any problems in this product or documentation, please report them to us in writing.

Informatica, Informatica Platform, Informatica Data Services, PowerCenter, PowerCenterRT, PowerCenter Connect, PowerCenter Data Analyzer, PowerExchange, PowerMart, Metadata Manager, Informatica Data Quality, Informatica Data Explorer, Informatica B2B Data Transformation, Informatica B2B Data Exchange Informatica On Demand, Informatica Identity Resolution, Informatica Application Information Lifecycle Management, Informatica Complex Event Processing, Ultra Messaging and Informatica Master Data Management are trademarks or registered trademarks of Informatica LLC in the United States and in jurisdictions throughout the world. All other company and product names may be trade names or trademarks of their respective owners.

Portions of this software and/or documentation are subject to copyright held by third parties, including without limitation: Copyright DataDirect Technologies. All rights reserved. Copyright © Sun Microsystems. All rights reserved. Copyright © RSA Security Inc. All Rights Reserved. Copyright © Ordinal Technology Corp. All rights reserved. Copyright © Aandacht c.v. All rights reserved. Copyright Genivia, Inc. All rights reserved. Copyright Isomorphic Software. All rights reserved. Copyright © Meta Integration Technology, Inc. All rights reserved. Copyright © Intalio. All rights reserved. Copyright © Oracle. All rights reserved. Copyright © Adobe Systems Incorporated. All rights reserved. Copyright © DataArt, Inc. All rights reserved. Copyright © ComponentSource. All rights reserved. Copyright © Microsoft Corporation. All rights reserved. Copyright © Rogue Wave Software, Inc. All rights reserved. Copyright © Teradata Corporation. All rights reserved. Copyright © Yahoo! Inc. All rights reserved. Copyright © Glyph & Cog, LLC. All rights reserved. Copyright © Thinkmap, Inc. All rights reserved. Copyright © Clearpace Software Limited. All rights reserved. Copyright © Information Builders, Inc. All rights reserved. Copyright © OSS Nokalva, Inc. All rights reserved. Copyright Edifecs, Inc. All rights reserved. Copyright Cleo Communications, Inc. All rights reserved. Copyright © International Organization for Standardization 1986. All rights reserved. Copyright © ej-technologies GmbH. All rights reserved. Copyright © Jaspersoft Corporation. All rights reserved. Copyright © International Business Machines Corporation. All rights reserved. Copyright © yWorks GmbH. All rights reserved. Copyright © Lucent Technologies. All rights reserved. Copyright (c) University of Toronto. All rights reserved. Copyright © Daniel Veillard. All rights reserved. Copyright © Unicode, Inc. Copyright IBM Corp. All rights reserved. Copyright © MicroQuill Software Publishing, Inc. All rights reserved. Copyright © PassMark Software Pty Ltd. All rights reserved. Copyright © LogiXML, Inc. All rights reserved. Copyright © 2003-2010 Lorenzi Davide, All rights reserved. Copyright © Red Hat, Inc. All rights reserved. Copyright © The Board of Trustees of the Leland Stanford Junior University. All rights reserved. Copyright © EMC Corporation. All rights reserved. Copyright © Flexera Software. All rights reserved. Copyright © Jinfonet Software. All rights reserved. Copyright © Apple Inc. All rights reserved. Copyright © Telerik Inc. All rights reserved. Copyright © BEA Systems. All rights reserved. Copyright © PDFlib GmbH. All rights reserved. Copyright © Orientation in Objects GmbH. All rights reserved. Copyright © Tanuki Software, Ltd. All rights reserved. Copyright © Ricebridge. All rights reserved. Copyright © Sencha, Inc. All rights reserved. Copyright © Scalable Systems, Inc. All rights reserved. Copyright © jqWidgets. All rights reserved. Copyright © Tableau Software, Inc. All rights reserved. Copyright © MaxMind, Inc. All Rights Reserved. Copyright © TMate Software s.r.o. All rights reserved. Copyright © MapR Technologies Inc. All rights reserved. Copyright © Amazon Corporate LLC. All rights reserved. Copyright © Highsoft. All rights reserved. Copyright © Python Software Foundation. All rights reserved. Copyright © BeOpen.com. All rights reserved. Copyright © CNRI. All rights reserved.

This product includes software developed by the Apache Software Foundation (<http://www.apache.org/>), and/or other software which is licensed under various versions of the Apache License (the "License"). You may obtain a copy of these Licenses at <http://www.apache.org/licenses/>. Unless required by applicable law or agreed to in writing, software distributed under these Licenses is distributed on an "AS IS" BASIS, WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied. See the Licenses for the specific language governing permissions and limitations under the Licenses.

This product includes software which was developed by Mozilla (<http://www.mozilla.org/>), software copyright The JBoss Group, LLC, all rights reserved; software copyright © 1999-2006 by Bruno Lowagie and Paulo Soares and other software which is licensed under various versions of the GNU Lesser General Public License Agreement, which may be found at <http://www.gnu.org/licenses/lgpl.html>. The materials are provided free of charge by Informatica, "as-is", without warranty of any kind, either express or implied, including but not limited to the implied warranties of merchantability and fitness for a particular purpose.

The product includes ACE(TM) and TAO(TM) software copyrighted by Douglas C. Schmidt and his research group at Washington University, University of California, Irvine, and Vanderbilt University, Copyright (©) 1993-2006, all rights reserved.

This product includes software developed by the OpenSSL Project for use in the OpenSSL Toolkit (copyright The OpenSSL Project. All Rights Reserved) and redistribution of this software is subject to terms available at <http://www.openssl.org> and <http://www.openssl.org/source/license.html>.

This product includes Curl software which is Copyright 1996-2013, Daniel Stenberg, <daniel@haxx.se>. All Rights Reserved. Permissions and limitations regarding this software are subject to terms available at <http://curl.haxx.se/docs/copyright.html>. Permission to use, copy, modify, and distribute this software for any purpose with or without fee is hereby granted, provided that the above copyright notice and this permission notice appear in all copies.

The product includes software copyright 2001-2005 (©) MetaStuff, Ltd. All Rights Reserved. Permissions and limitations regarding this software are subject to terms available at <http://www.dom4j.org/license.html>.

The product includes software copyright © 2004-2007, The Dojo Foundation. All Rights Reserved. Permissions and limitations regarding this software are subject to terms available at <http://dojotoolkit.org/license>.

This product includes ICU software which is copyright International Business Machines Corporation and others. All rights reserved. Permissions and limitations regarding this software are subject to terms available at <http://source.icu-project.org/repos/icu/icu/trunk/license.html>.

This product includes software copyright © 1996-2006 Per Bothner. All rights reserved. Your right to use such materials is set forth in the license which may be found at <http://www.gnu.org/software/kawa/Software-License.html>.

This product includes OSSP UUID software which is Copyright © 2002 Ralf S. Engelschall, Copyright © 2002 The OSSP Project Copyright © 2002 Cable & Wireless Deutschland. Permissions and limitations regarding this software are subject to terms available at <http://www.opensource.org/licenses/mit-license.php>.

This product includes software developed by Boost (<http://www.boost.org/>) or under the Boost software license. Permissions and limitations regarding this software are subject to terms available at http://www.boost.org/LICENSE_1_0.txt.

This product includes software copyright © 1997-2007 University of Cambridge. Permissions and limitations regarding this software are subject to terms available at <http://www.pcre.org/license.txt>.

This product includes software copyright © 2007 The Eclipse Foundation. All Rights Reserved. Permissions and limitations regarding this software are subject to terms available at <http://www.eclipse.org/org/documents/epl-v10.php> and at <http://www.eclipse.org/org/documents/edl-v10.php>.

This product includes software licensed under the terms at <http://www.tcl.tk/software/tcltk/license.html>, <http://www.bosrup.com/web/overlib/?License>, <http://www.stlport.org/doc/license.html>, <http://asm.ow2.org/license.html>, <http://www.cryptix.org/LICENSE.TXT>, <http://hsqldb.org/web/hsqLicense.html>, <http://httpunit.sourceforge.net/doc/license.html>, <http://jung.sourceforge.net/license.txt>, http://www.gzip.org/zlib/zlib_license.html, <http://www.openldap.org/software/release/license.html>, <http://www.libssh2.org>, <http://slf4j.org/license.html>, <http://www.sente.ch/software/OpenSourceLicense.html>, <http://fusesource.com/downloads/license-agreements/fuse-message-broker-v-5-3-license-agreement>, <http://antlr.org/license.html>, <http://aopalliance.sourceforge.net/>, <http://www.bouncycastle.org/licence.html>, <http://www.jgraph.com/jgraphdownload.html>, <http://www.jcraft.com/jsch/LICENSE.txt>, http://jotm.objectweb.org/bsd_license.html, <http://www.w3.org/Consortium/Legal/2002/copyright-software-20021231>, <http://www.slf4j.org/license.html>, <http://nanoxml.sourceforge.net/orig/copyright.html>, <http://www.json.org/license.html>, <http://forge.ow2.org/projects/javaservice/>, <http://www.postgresql.org/about/licence.html>, <http://www.sqlite.org/copyright.html>, <http://www.tcl.tk/software/tcltk/license.html>, <http://www.jaxen.org/faq.html>, <http://www.jdom.org/docs/faq.html>, <http://www.slf4j.org/license.html>, <http://www.iodbc.org/dataspace/iodbc/wiki/IODBC/License>, <http://www.keplerproject.org/md5/license.html>, <http://www.toedter.com/en/jcalendar/license.html>, <http://www.edankert.com/bounce/index.html>, <http://www.net-snmp.org/about/license.html>, <http://www.openmdx.org/#FAQ>, http://www.php.net/license/3_01.txt, <http://srp.stanford.edu/license.txt>, <http://www.schneider.com/blowfish.html>, <http://www.jmock.org/license.html>, <http://xsom.java.net>, <http://benalman.com/about/license/>, <https://github.com/CreateJS/EaselJS/blob/master/src/easeljs/display/Bitmap.js>, <http://www.h2database.com/html/license.html#summary>, <http://jsoncpp.sourceforge.net/LICENSE>, <http://jdbc.postgresql.org/license.html>, <http://protobuf.googlecode.com/svn/trunk/src/google/protobuf/descriptor.proto>, <https://github.com/rantav/hector/blob/master/LICENSE>, <http://web.mit.edu/Kerberos/krb5-current/doc/mitK5license.html>, <http://jibx.sourceforge.net/jibx-license.html>, <https://github.com/lyokato/libgeohash/blob/master/LICENSE>, <https://github.com/hjiang/jsonxx/blob/master/LICENSE>, <https://code.google.com/p/lz4/>, <https://github.com/jedisct1/libsodium/blob/master/LICENSE>, <http://one-jar.sourceforge.net/index.php?page=documents&file=license>, <https://github.com/EsotericSoftware/kryo/blob/master/license.txt>, <http://www.scala-lang.org/license.html>, <https://github.com/tinkerpop/blueprints/blob/master/LICENSE.txt>, <http://gee.cs.oswego.edu/dl/classes/EDU/oswego/cs/dl/util/concurrent/intro.html>, <https://aws.amazon.com/ssl/>, <https://github.com/twbs/bootstrap/blob/master/LICENSE>, <https://sourceforge.net/p/xmlunit/code/HEAD/tree/trunk/LICENSE.txt>, <https://github.com/documentcloud/underscore-contrib/blob/master/LICENSE>, and <https://github.com/apache/hbase/blob/master/LICENSE.txt>.

This product includes software licensed under the Academic Free License (<http://www.opensource.org/licenses/afl-3.0.php>), the Common Development and Distribution License (<http://www.opensource.org/licenses/cddl1.php>), the Common Public License (<http://www.opensource.org/licenses/cpl1.0.php>), the Sun Binary Code License Agreement Supplemental License Terms, the BSD License (<http://www.opensource.org/licenses/bsd-license.php>), the new BSD License (<http://opensource.org/licenses/BSD-3-Clause>), the MIT License (<http://www.opensource.org/licenses/mit-license.php>), the Artistic License (<http://www.opensource.org/licenses/artistic-license-1.0>) and the Initial Developer's Public License Version 1.0 (<http://www.firebirdsql.org/en/initial-developer-s-public-license-version-1-0/>).

This product includes software copyright © 2003-2006 Joe Walnes, 2006-2007 XStream Committers. All rights reserved. Permissions and limitations regarding this software are subject to terms available at <http://xstream.codehaus.org/license.html>. This product includes software developed by the Indiana University Extreme! Lab. For further information please visit <http://www.extreme.indiana.edu/>.

This product includes software Copyright (c) 2013 Frank Balluffi and Markus Moeller. All rights reserved. Permissions and limitations regarding this software are subject to terms of the MIT license.

See patents at <https://www.informatica.com/legal/patents.html>.

DISCLAIMER: Informatica LLC provides this documentation "as is" without warranty of any kind, either express or implied, including, but not limited to, the implied warranties of noninfringement, merchantability, or use for a particular purpose. Informatica LLC does not warrant that this software or documentation is error free. The information provided in this software or documentation may include technical inaccuracies or typographical errors. The information in this software and documentation is subject to change at any time without notice.

NOTICES

This Informatica product (the "Software") includes certain drivers (the "DataDirect Drivers") from DataDirect Technologies, an operating company of Progress Software Corporation ("DataDirect") which are subject to the following terms and conditions:

1. THE DATADIRECT DRIVERS ARE PROVIDED "AS IS" WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESSED OR IMPLIED, INCLUDING BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND NON-INFRINGEMENT.
2. IN NO EVENT WILL DATADIRECT OR ITS THIRD PARTY SUPPLIERS BE LIABLE TO THE END-USER CUSTOMER FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, CONSEQUENTIAL OR OTHER DAMAGES ARISING OUT OF THE USE OF THE ODBC DRIVERS, WHETHER OR NOT INFORMED OF THE POSSIBILITIES OF DAMAGES IN ADVANCE. THESE LIMITATIONS APPLY TO ALL CAUSES OF ACTION, INCLUDING, WITHOUT LIMITATION, BREACH OF CONTRACT, BREACH OF WARRANTY, NEGLIGENCE, STRICT LIABILITY, MISREPRESENTATION AND OTHER TORTS.

Publication Date: 2018-06-09

Table of Contents

Preface	5
Learning About Informatica SSA-NAME3.	5
What Do I Read If.	7
Informatica Resources.	8
Informatica My Support Portal.	8
Informatica Documentation.	8
Informatica Product Availability Matrixes.	8
Informatica Web Site.	8
Informatica How-To Library.	9
Informatica Knowledge Base.	9
Informatica Support YouTube Channel.	9
Informatica Marketplace.	9
Informatica Velocity.	9
Informatica Global Customer Support.	9
 Chapter 1: SSA-NAME3 Implementation Overview.....	 10
SSA-NAME3 Software Components.	10
Implementation Architecture.	13
Resources Required to Implement SSA-NAME3.	14
 Chapter 2: SSA-NAME3 Installation.....	 16
Installation Overview.	16
The Prototype Application.	17
Choosing the Data to Search.	18
Choosing the Mode of Search.	18
Testing the Standard Population Rules.	18
Defining a New Database Table, Column or File.	19
Developing the Application Programs.	20
After the Prototype Is Working.	20
 Index.....	 22

Preface

Welcome to the Informatica SSA-NAME3 Getting Started Guide. This guide is intended to be the first technical material a new user reads before installing or using the SSANAME3 software, regardless of the platform or environment.

Learning About Informatica SSA-NAME3

This section provides details of documentation available with the SSA-NAME3 product.

Introduction to SSA-NAME3

Provides an overview of SSA-NAME3. It is written in a way that can be read by someone who has no prior experience of the product and wants a general overview of SSA-NAME3. It explains the problems SSA-NAME3 overcomes and provides an overview of how this is done. One chapter is dedicated to providing an overview for Application Programmers.

Getting Started

This manual is intended to be the first technical material a new developer or designer reads before installing or using the SSA-NAME3 software, regardless of the platform or environment. Its goal is to help a new user get the software installed and produce a working prototype application that calls SSA-NAME3 and executes searches against their own data.

To achieve this it provides a "script" to follow which includes pointers to pertinent sections of the other manuals.

Application & Database Design

This manual contains tips and techniques useful for setting up and optimizing a name search and matching application, including database issues, and illustrates best-practice techniques, common pitfalls, and strategies regarding the subject of name and address matching.

Installation Guide

This manual provides information on how to install the SSA-NAME3 product.

SSA-NAME3 Workbench User Guide

This is a guide to using the SSA-NAME3 Workbench - a Java GUI tool that helps a programmer understand and prototype SSA-NAME3 calls. The Workbench is also used for:

- Generating Sample Program Code;
- Executing SSA-NAME3 Calls;
- Testing different SSA-NAME3 run-time options;
- Producing debugging and support information for Informatica Corporation

Note: The Workbench in itself is not a search and match application. It assists the developer build a search and match application.

API Reference

The ultimate goal of an SSA-NAME3 implementation is for application programs to be able to call SSA-NAME3's API Functions to build keys and search strategies and to compute match scores and decisions.

This manual describes a typical program process flow for building an identity search application, and also lists in detail each of the API Functions. It describes the parameters required by these functions and the information returned.

Population Override Manager User's Guide

This is a guide to using the SSA-NAME3 Population Override Manager - a Java GUI tool that allows a trained data analyst to override some of the Standard Population rules that are supplied with the product, or provided in the form of a Custom Population. The types of rules that can be overridden using this tool are:

- Edit-list rules
- Frequency tables
- Scalar Frequency Tables
- Matching Purposes

Note: Use of this tool without proper training from Informatica should not be attempted, as improper use can adversely affect the reliability and performance of the search application(s).

Edit Rule Wizard User's Guide

This is a guide to using the SSA-NAME3 Edit Rule Wizard - a Java GUI tool that helps a business user safely add certain types of Edit Rules to the Standard or Custom Population without requiring specific knowledge of SSA-NAME3 or support from a programmer or data analyst. The types of rules that can be added using this tool are:

- Discard a word or phrase when searching and matching (e.g. a new "noise" word)
- Add a new replacement word or phrase when searching and matching (e.g. a new "abbreviation", "nickname" or "acronym")
- Add a new compound name marker word

Release Notes

The Release Notes contain information about what's new in this version of SSA-NAME3. It is also used to summarize any documentation updates as they are published.

What Do I Read If. . .

I am. . .

. . . a business manager

The INTRODUCTION TO SSA-NAME3 will address questions such as "Why have we got SSA-NAME3?", "What does SSA-NAME3 do"?

I am. . .

. . . a system designer or DBA

The INTRODUCTION TO SSA-NAME3 will address questions such as "What resources are needed to implement SSA-NAME3?". The APPLICATION & DATABASE DESIGN manual will lead you through many of the design considerations of name search and matching applications.

I am. . .

. . . installing SSA-NAME3

Before attempting to install SSA-NAME3 you should read the Getting Started document. This will describe the pre-requisites and help you plan the installation and implementation of SSA-NAME3. The actual installation steps for your platform are documented in the Installation Guide.

I am. . .

. . . an Analyst or Application Programmer

A high-level overview is provided specifically for Application Programmers in the INTRODUCTION TO SSA-NAME3 manual. Before attempting to develop programs that interface with SSA-NAME3, you should also read the GETTING STARTED and APPLICATION & DATABASE DESIGN manuals, as well as experimenting with calls in the WORKBENCH USER GUIDE.

When developing the application program(s), use the API REFERENCE manual which describes a typical application and the Function parameters.

Working example programs that illustrate the calls to SSA-NAME3 in various languages are available by using the Sample Program button on the Workbench.

I want to know. . .

. . . what SSA-NAME3 does

The INTRODUCTION TO SSA-NAME3 manual gives an overview of what SSA-NAME3 does and how it does it.

I want to know. . .

. . . how to setup the database

Refer to the APPLICATION & DATABASE DESIGN manual for tips and techniques on configuring the database to store SSA-NAME3 Keys and optimizing it for searching and matching.

I want to know. . .

. . . how to code a search application

The INTRODUCTION TO SSA-NAME3 manual contains a specific section designed to get application programmers familiar with the concepts of developing an SSA-NAME3 search and match application.

The API REFERENCE GUIDE details the Function calls required and their parameters. The SSA-NAME3 WORKBENCH USER GUIDE shows how to generate a sample program in a variety of programming languages.

Informatica Resources

Informatica My Support Portal

As an Informatica customer, the first step in reaching out to Informatica is through the Informatica My Support Portal at <https://mysupport.informatica.com>. The My Support Portal is the largest online data integration collaboration platform with over 100,000 Informatica customers and partners worldwide.

As a member, you can:

- Access all of your Informatica resources in one place.
- Review your support cases.
- Search the Knowledge Base, find product documentation, access how-to documents, and watch support videos.
- Find your local Informatica User Group Network and collaborate with your peers.

Informatica Documentation

The Informatica Documentation team makes every effort to create accurate, usable documentation. If you have questions, comments, or ideas about this documentation, contact the Informatica Documentation team through email at infa_documentation@informatica.com. We will use your feedback to improve our documentation. Let us know if we can contact you regarding your comments.

The Documentation team updates documentation as needed. To get the latest documentation for your product, navigate to Product Documentation from <https://mysupport.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. You can access the PAMs on the Informatica My Support Portal at <https://mysupport.informatica.com>.

Informatica Web Site

You can access the Informatica corporate web site at <https://www.informatica.com>. The site contains information about Informatica, its background, upcoming events, and sales offices. You will also find product and partner information. The services area of the site includes important information about technical support, training and education, and implementation services.

Informatica How-To Library

As an Informatica customer, you can access the Informatica How-To Library at <https://mysupport.informatica.com>. The How-To Library is a collection of resources to help you learn more about Informatica products and features. It includes articles and interactive demonstrations that provide solutions to common problems, compare features and behaviors, and guide you through performing specific real-world tasks.

Informatica Knowledge Base

As an Informatica customer, you can access the Informatica Knowledge Base at <https://mysupport.informatica.com>. Use the Knowledge Base to search for documented solutions to known technical issues about Informatica products. You can also find answers to frequently asked questions, technical white papers, and technical tips. If you have questions, comments, or ideas about the Knowledge Base, contact the Informatica Knowledge Base team through email at KB_Feedback@informatica.com.

Informatica Support YouTube Channel

You can access the Informatica Support YouTube channel at <http://www.youtube.com/user/INFASupport>. The Informatica Support YouTube channel includes videos about solutions that guide you through performing specific tasks. If you have questions, comments, or ideas about the Informatica Support YouTube channel, contact the Support YouTube team through email at supportvideos@informatica.com or send a tweet to @INFASupport.

Informatica Marketplace

The Informatica Marketplace is a forum where developers and partners can share solutions that augment, extend, or enhance data integration implementations. By leveraging any of the hundreds of solutions available on the Marketplace, you can improve your productivity and speed up time to implementation on your projects. You can access Informatica Marketplace at <http://www.informaticamarketplace.com>.

Informatica Velocity

You can access Informatica Velocity at <https://mysupport.informatica.com>. 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 have questions, comments, or ideas about Informatica Velocity, contact Informatica Professional Services at ips@informatica.com.

Informatica Global Customer Support

You can contact a Customer Support Center by telephone or through the Online Support.

Online Support requires a user name and password. You can request a user name and password at <http://mysupport.informatica.com>.

The telephone numbers for Informatica Global Customer Support are available from the Informatica web site at <http://www.informatica.com/us/services-and-training/support-services/global-support-centers/>.

CHAPTER 1

SSA-NAME3 Implementation Overview

This document is intended to be the first technical material a new user reads before installing or using the SSA-NAME3 software, regardless of the platform or environment.

The goal of this guide is to help a new user implement SSA-NAME3 and produce a working prototype SSA-NAME3 application against their own data

For a general understanding of SSA-NAME3 and its uses, see the *Introduction to SSA-NAME3 Guide*.

SSA-NAME3 Software Components

The SSA-NAME3 software consists of a Callable Routine, Standard Populations, a Developer's Workbench, the Population Override Manager, the Edit Rule Wizard, and the documentation needed to support these facilities, provided in PDF format.

The Callable Routine contains the proprietary algorithms and processes used for building keys and search strategies and for computing match decisions for a calling application. The Callable Routine is provided as a shareable, reentrant object, such as a DLL or Shared Library, and also as a server process. It is invoked via the documented API.

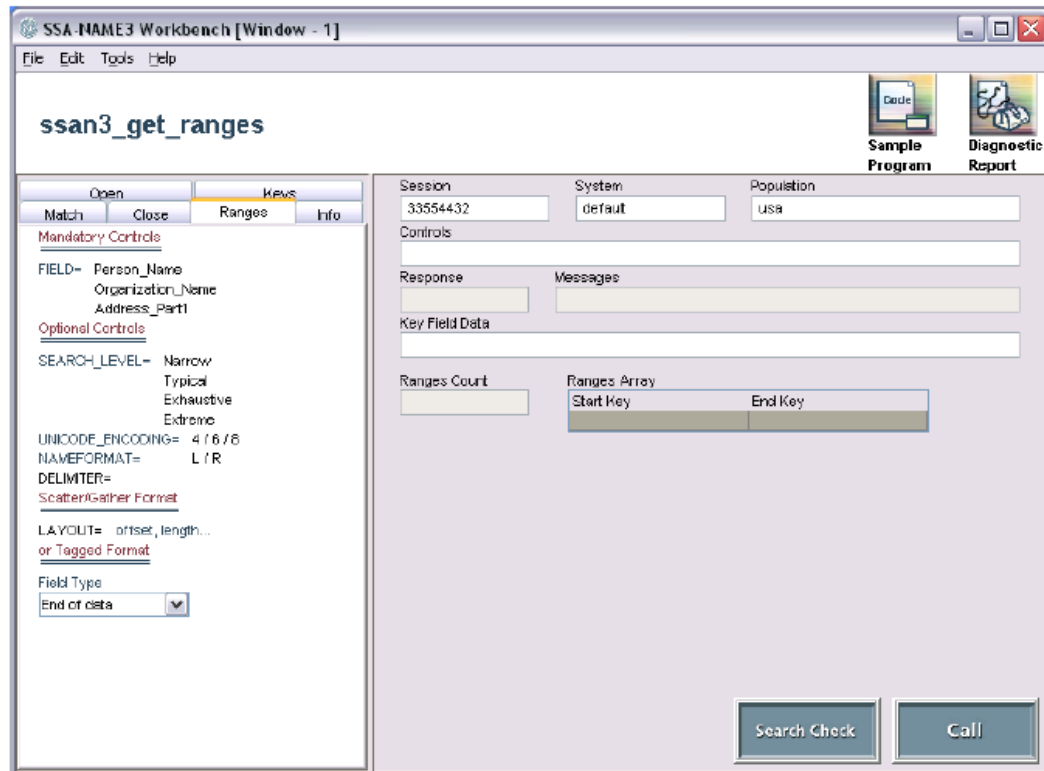
The Standard Populations contain the out-of-the-box Country and Language rules that are applied to the internal algorithms to support different countries, character sets, data types and application needs. The Standard Populations are provided as ASCII text files with the extension `.ysp`, and are accessible only via the Callable Routine, or one of the SSA-NAME3 GUI Clients. They are used by the Callable Routine to influence key-building, search range building and match computation. The Standard Population rules may be overridden by a qualified data analyst through the Population Override Manager. A subset of rules may be safely added to the standard set by a business user using the Edit RuleWizard.

The Standard Populations are normally installed by the person installing SSA-NAME3 into a subdirectory called **default**. This sub-directory name is later specified during runtime use of SSANAME3 as the System parameter. For example, a call to the "USA" Population specifying the "default" System will look for the `usa.ysp` Standard Population in the sub-directory called **default**.

It is highly recommend that users create new sub-directories (and therefore Systems) that correspond to the use and status of the standard populations. For example, a typical installation will use the **default** directory to hold the Informatica Corporation supplied Standard Populations, and user-created directories such as **dev**, **test**, **qa** and **prod** for migration purposes.

The Developer's Workbench

The Developer's Workbench is a Java GUI application intended to help programmers understand and test the API functions. It provides a Function Testing platform for each function call, and allows the user to input parameter values, invoke **SSA-NAME3**, and view the output results (keys, search ranges and match decisions). It also allows the analyst compare search and match levels, thus helping choose the most appropriate purposes and levels. The Workbench also provides access to the SSA-NAME3 documentation and generates sample program code in a variety of programming languages.



The Population Override Manager

The Population Override Manager is a Java-based, GUI application available for use by a data analyst trained in the syntax of the SSA-NAME3 Edit rules, and versed in the consequences of making changes that may affect search and match quality and performance or require indexes to be re-built. It allows the addition or removal of Edit rules from the Standard or Custom Population set. Using this utility, the analyst can also replace the packaged Frequency table with one built from the organization's own data.

SSA-NAME3 Edit List Override Editor - USA Population

Save Overrides
Discard Changes
Test
Commit Changes
Close

Search Rules for Tom
Search

Existing Edit List Rules
Disable Rule

Rule Type	Value 1	Value 2	Value 3
Word Rule	NC - Common Nickname Rep...	TOM	THOMAS
Word Rule	SN - Secondary Lookup Word	THOMAS	TAM
Word Rule	SN - Secondary Lookup Word	THOMAS	TAMI
Word Rule	NC - Common Nickname Rep...	THOS	THOMAS
Word Rule	NN - Nickname Replace Word	TOMASI	THOMAS
Word Rule	NN - Nickname Replace Word	TOMISLAV	THOMAS
Word Rule	NN - Nickname Replace Word	TAMAS	THOMAS
Word Rule	SN - Secondary Lookup Word	TAM	THOMAS
Word Rule	NK - Nickname Replace Di...	TOM	THOMAS
Word Rule	SN - Secondary Lookup Word	TAMI	THOMAS
Word Rule	NN - Nickname Replace Word	THONMO	THOMAS

Table displaying the Edit Rules found from a search

Override Edit List Rules
Add Override
Delete Override

Rule Type	Value 1	Value 2	Value 3	Action	Comment
-----------	---------	---------	---------	--------	---------

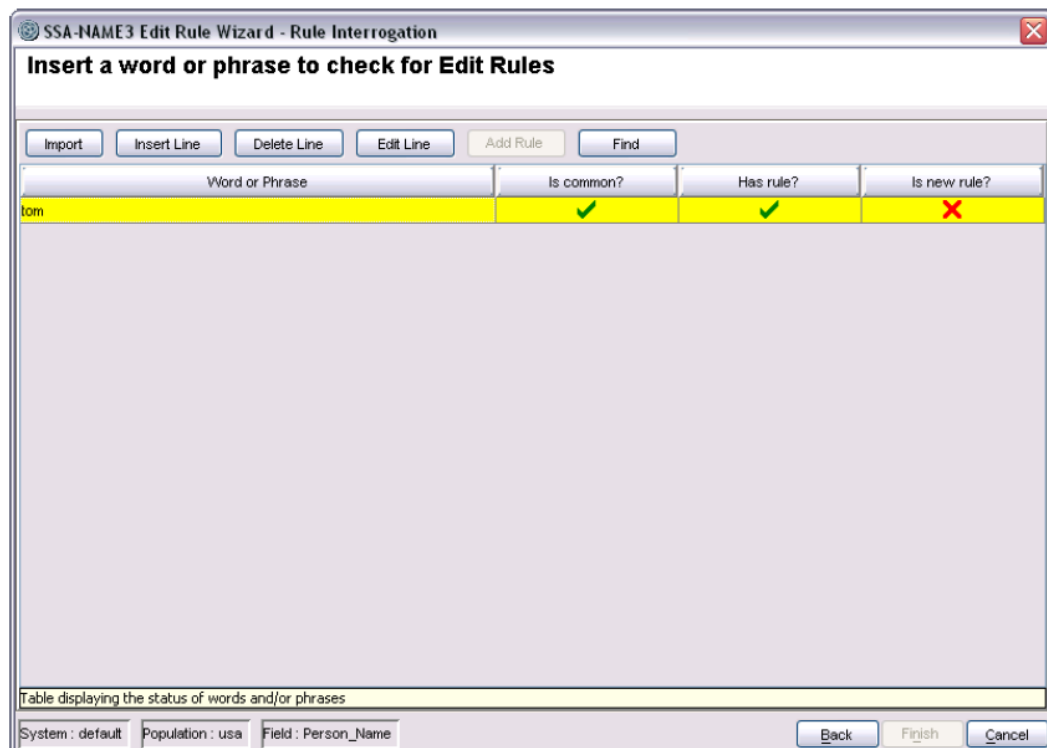
Table showing the uncommitted rules

System : default
Population : usa
Key Field : Person_Name

The action of committing rules via the Population Override Manager results in the generation of a **Local Population** rule-set. This is an ASCII text file with an extension of .y1p. Like .y1p files, .y1p files are accessible only via the Callable Routine, or one of the SSA-NAME3 GUI Clients.

The Edit Rule Wizard

The Edit Rule Wizard is a Java-based, GUI application available for use by a business or non-IT user to safely add certain types of Edit rules, typically new nicknames or synonyms, new noise words, or new phrase replacement rules, without requiring involvement from IT or the need to re-build indexes.



The action of committing rules via the Edit Rule Wizard results in the generation of a **Local Population** rule-set. This is an ASCII text file with an extension of .y1p. Like .y1p files, .y1p files are accessible only via the Callable Routine, or one of the SSA-NAME3 GUI Clients.

Implementation Architecture

The SSA-NAME3 software can be invoked either through a socket interface using TCP/IP or through a local call. A local call will interface directly to the DLL or Shared Library. A call through TCP/IP will interface with the SSA-NAME3 Server.

During development, the use of the SSA-NAME3 server allows the Workbench, Population Override Manager and Edit Rule Wizard to be situated on one or more clients, accessing an SSA-NAME3 runtime installation on a server. It also means that application programs can be developed on different clients and access a single copy of the SSA-NAME3 run-time installation on a server.

This client/server architecture is not intended to be the basis for a production implementation of an SSA-NAME3 application. It is recommended that the application that calls SSA-NAME3, and the SSA-NAME3 Callable Routine, both reside on the same server in a production environment even if the Callable Routine is started as a Server process. Performance degradation will occur if this design is not followed.

A user's production application only requires access to the Callable Routine and the particular Standard Population(s) used. The Workbench is a development only component. The Population Override Manager is also intended to be used in a "development" environment. It is possible to use the Edit Rule Wizard in a production environment.

If the system owner has decided that the Edit Rule Wizard can be used in a production environment, then the SSA-NAME3 server will need to be started in the production environment to service the Edit Rule Wizard requests only. However, for changes made by the Edit Rule Wizard to take effect, the SSA-NAME3 DLL or Shared Library will need to be unloaded and reloaded, or the server restarted.

Resources Required to Implement SSA-NAME3

The following typical resources are required to implement SSA-NAME3 and develop a prototype application.

Hardware, Software and Network Resources

Windows 32-bit Computer

If the SSA-NAME3 GUI clients are to run on Windows, an IBM Compatible Computer with Pentium III or Higher with Minimum 512MB RAM is required.

The Windows platform is typically where the SSA-NAME3 Workbench, Population Override Manager and Edit Rule Wizard would run. It may also be where the user's application is developed. It must be running a Microsoft Windows 32-bit operating system, for example, Windows NT/2000/XP. Users can also deploy complete production applications and systems on the Win32 platform.

Connection from Client to Server Platform(s)

A TCP/IP connection is required if the Population Override Manager or Edit Rule Wizard are to be used (as these clients communicate only with the SSA-NAME3 server), or if the Developer's Workbench is to run on a client and connect to the SSA-NAME3 server on a remote computer.

An application can also access the SSA-NAME3 Server using this network connection, however it is not recommended that remote calls are used in a production system (due to performance degradation).

Server

The server is the computer where the SSA-NAME3 applications will typically be executed, that is where the SSA-NAME3 Callable Routine resides. This is frequently a database server, as performance improves the closer the search and match process is to the data.

Application Development Environment

Using SSA-NAME3 requires an application development language which supports a Call to an external routine. The application needs to be able to call a Dynamic Link Library in the case of MS Windows or a Shared Object/Library on Unix.

In a z/OS environment, the application needs to be able to call a C Load Module. See the *Z/OS INSTALLATION GUIDE* for more details.

Data Resources

Test Data

This is the data that will be used by the user's application to build SSA-NAME3 Keys and perform searches. Preferably this should be a copy of the Production database. It could also be any other data, although if the data is not "real" the results themselves will not be realistic.

Sample of Real Production Data

A data analyst trained in the customization of SSA-NAME3 may decide to override the Frequency Table or Scalar Frequency Tables for one or more of the search fields. If so, a sample of real production names or addresses (depending on the field) will be required. The sample can be as little as 250,000, or the entire file, whichever is the smaller. If it is a sample, it must be a random sample.

A New Database Table

SSA-NAME3 will generate multiple keys for every name and address. Therefore a new table or indexed file needs to be set-up to contain the SSA-NAME3 Keys and other data. The new table needs to have an index defined on the SSA-NAME3 Keys. It is highly recommended that this new table be initially physically sequenced in SSA-NAME3 key order.

It is also important to note that the SSA-NAME3 Keys will not be unique values, so the column created to store the keys needs to take this fact into consideration. For more on Database design, see the *APPLICATION AND DATABASE DESIGN* guide.

Job Functions

The following job functions are required to install and implement SSA-NAME3 in a prototype application.

Installer

The SSA-NAME3 software may need to be installed on both a Client and a Server. See the *Implementation Architecture* section earlier in this guide. While the installation is not difficult, some organizations may have internal procedures regarding who performs such tasks.

Analyst/Programmer

In order to understand the SSA-NAME3 API's and build the prototype application, the Programmer should be fluent in the development language chosen for the prototype and be comfortable with the Operating System command language of both the target and the Client. This role is generally performed by the person who also carries out the installation. Typically this person will also perform the data and business purpose analysis to select Standard Populations, Key/Search/Match Levels and Match Purposes.

Business Analyst

A business analyst or project manager is normally involved to determine the data characteristics and source, the language, along with the business purpose for performing the identity searches. Once this analysis is conducted, the analyst may wish to discuss the results with the local Informatica Corporation support office to select the most appropriate Standard Population, Key Level, Search Level and Match Purposes.

Database Administrator

To define and optimize the new database table(s) to contain SSA-NAME3 Keys and matching information.

CHAPTER 2

SSA-NAME3 Installation

This chapter includes the following topics:

- [Installation Overview, 16](#)
- [The Prototype Application, 17](#)
- [Choosing the Data to Search, 18](#)
- [Choosing the Mode of Search, 18](#)
- [Testing the Standard Population Rules, 18](#)
- [Defining a New Database Table, Column or File, 19](#)
- [Developing the Application Programs, 20](#)
- [After the Prototype Is Working, 20](#)

Installation Overview

This topic provides information about the Installation of SSA-NAME3.

How SSA-NAME3 is Delivered

The SSANAME3 Product consists of the SSA-NAME3 Callable Routine, the Standard Populations, the Developer's Workbench, Population Override Manager and Edit Rule Wizard.

Customers who are evaluating and licensing SSA-NAME3 will receive a download link for the installation media from Informatica Shipping. A 30-day temporary license key will also be provided to customers evaluating the product. The 30-day license will allow for a full installation of the SSA-NAME3 with no restrictions.

Complete documentation for SSA-NAME3 is electronically with the installation media and, by default, will be expanded into a sub-directory called `docs` where SSA-NAME3 is installed.

The Informatica Installer is provided installation on Windows and Unix platforms which may be run as a Java GUI or scripted at the command prompt. For z/OS, a `.zip` file containing a number of load libraries will be provided. Refer to the *SSA-NAME3 Installation Guide* for detailed information on installing the product.

Choices to Be Made Before Installation

If this Installation is an upgrade, note that the Informatica Product and Product Update Installers will overwrite the standard populations in the 'default' sub-directory'. Therefore it is important to store your active populations in other sub-directories/systems folders such as `dev`, `test`, `prod`, etc.

Installation Sequence

The Informatica License Server must be installed and running on the target computer in order to install the SSA-Name3 libraries, server and client components. The 'target' computer is the computer where the SSA-NAME3 Callable Routine and Standard Populations are installed. Client components, consisting of the Developer's Workbench, Population Override Manager and Edit Rule Wizard are also installed on the target computer, but may be installed on other client computers as well.

Both the Informatica License Server and the SSA-NAME3 product are installed using the Informatica Product Installer provided with the installation media.

If you plan to install client components on other computers, you must first complete installation on the target computer and insure that the License Server is running and then install on the client computers providing the hostname for the target computer and port number for the License Server. Before the clients can be used, the SSA-NAME3 server must be started on the target computer.

The Prototype Application

Even though it would be your intention to eventually integrate the SSA-NAME3 search, match and key building facilities into a system, you need to first get a stand-alone prototype application working satisfactorily against your data. This allows familiarization with the software and how it works without confusion from other system requirements.

Once the prototype is working satisfactorily, you will have the necessary knowledge and be in a much better position to integrate the software into the system.

To prototype an SSA-NAME3 application requires:

- Choosing the data to search or match on and choosing whether the prototype application will be online or batch.
- Deciding on a suitable Standard Population and assessing the purpose of the search application. At this point it is highly recommended that you create another subdirectory at the same level as "default". that is if you have installed SSA-NAME3 in the folder `c:\InformaticaIR`, you will have `c:\InformaticaIR\pr\default` available. Create another directory, for example, `c:\InformaticaIR\pr\proto` and copy the appropriate Standard Population library to this directory.

From now on you can access this Standard Population in your prototype application and the SSA-NAME3 Workbench by specifying the system parameter equal to `proto`.

Defining a new Database table to hold the SSA-NAME3 Keys and other data. See the *APPLICATION AND DATABASE DESIGN* guide for more information.

Developing two prototype applications:

- Key-Build/Load program
- Search and Match program

The analyst/programmer can refer to the sample programs available through the Developer's Workbench. Additional example programs for a variety of platforms may be available on the Informatica Corporation Online Customer Support website. For more details and the choice of development examples, contact your Informatica Corporation support office.

Choosing the Data to Search

SSA-NAME3 can be used to search on the following types of names:

- Person names
- Organization (Company/Business) names (can also be used to support mixed Person/Organization names)
- Addresses (typically that part of the Address up to but not including the locality details, otherwise known as the Street address or Address Part 1)

It is important to choose data that is available in both reasonable quality and quantity. It is not recommended to build the prototype to search on the made-up names in the development database. We recommend to use an extract or copy of some real data.

Choosing the Mode of Search

The following factors help you choose between an online search prototype and a batch prototype:

- What is the main purpose for evaluating or purchasing SSA-NAME3?
- How quickly does the prototype need to be delivered?
- What are the analyst/programmers skills?

For example, if the main purpose for using SSA-NAME3 is for batch file matching, or the prototype needs to be delivered quickly and either the analyst/programmer does not have online skills or the online environment is difficult to build in, then a batch program could be the right choice, otherwise an online program is usually more interesting and the results are easier to present to others.

Testing the Standard Population Rules

After the installation of SSA-NAME3 and the Standard Populations are made available, you can test the Standard Population rules.

If connecting to a server then launch the SSA-NAME3 server process on that computer. See the *SSA-NAME3 INSTALLATION GUIDE*.

Launch the Developer's Workbench on the client computer. There will be an icon in the Informatica program group for the Workbench. The Workbench entry screen will be displayed. For more details on Workbench, see the *WORKBENCH USER GUIDE*.

1. Select the appropriate `system` name i.e. default, proto, prod etc. This specifies the location of the Standard Population `.ysp` file, or Custom Population `.ycp` file, you will use.
2. Select the Standard Population that suits the data.
3. From this screen, the Online Documentation is accessible. The *API REFERENCE GUIDE* explains the individual function details in more detail.
4. Click the **Workbench** button and enter the main Function Testing area.
5. Choose a function to test (the logical order is Keys, then Ranges, then Match).

A brief of each of the functions is described below. For more understanding, see the *API REFERENCE GUIDE*.

ssan3_open

Used to initiate a session to the SSA-NAME3 server process. Only used if the SSA-NAME3 server is being used.

ssan3_close

Used to close a session with the SSA-NAME3 server process. Only used if the SSA-NAME3 server is being used.

ssan3_get_keys

This is used by programs that need to build or maintain the SSA-NAME3 keys. The primary function controls are the Field type and Key Level. The user's program must store these keys in a database table specifically designed to hold the SSA-NAME3 Keys.

ssan3_get_ranges

This is used by the search program(s) to request the necessary SSA-NAME3 key ranges that will allow it to retrieve candidate records from the SSA-NAME3 Keys table. These Key Range values are typically inserted into an SQL statement to SELECT rows from the Key table which are between the Start and End Keys. All key ranges from the returned array must be processed for the search to be logically completed. The primary function controls are the Field Type and Search Level.

ssan3_match

As the user's search program retrieves the candidate records defined by key ranges, each candidate record needs to be compared with the original search record in order to compute a Score and Match decision. The Match Purpose chosen depends on the fields available in the data and the business purpose of the search. The primary function controls are the Purpose and Match Level.

ssan3_info

Returns information on the System and Population Rules.

Defining a New Database Table, Column or File

A new database table or indexed file needs to be defined to hold the SSA-NAME3 keys and other data. The SSA-NAME3 keys are 8 bytes in length (by default) and consists of ASCII printable character values.

The minimum number of columns or fields in this new table or file is two, the SSA-NAME3 key, and a foreign key to point back to the master record where the name or address was sourced from. For performance optimization, it is recommended that this table should be de-normalized to include in the search fields based on:

- the data to be used for matching
- any data to be used for filtering
- any data to appear on the user's search results screen

For more information, see the *APPLICATION AND DATABASE DESIGN Guide*.

Developing the Application Programs

There is a **Sample Code** button in the Workbench program. It is available from within any of the functions, except `ssan3_open` and `ssan3_close`. This **Sample Code** button will generate a code sample using the current parameters being used in the Workbench. This sample code can be generated in a number of popular development languages. For more details, see the *WORKBENCH USER GUIDE*.

Before you work on the prototype application, the analyst/programmer should refer to the *API REFERENCE GUIDE*.

How to Use the Sample Programs

The sample programs are intended to be a starting point for the prototype development. These programs can be copied and changed as appropriate, or, by referring to these programs, a new program can be written from scratch.

After the Prototype Is Working

Once there is a working prototype of an SSA-NAME3 search and matching application, it should be demonstrated to users from a functional point of view and any changes made to bring it broadly into line with those expectations.

The next step may be to generate the SSA-NAME3 Keys and use the search application on production data, or a copy of production data, if this had not already been done for the prototype. This will allow users to get a feel for the real performance and reliability of the search results. The feedback from this step will help adjust the Controls on the API calls to the needs of the users and constraints of the system. It may also feed useful information to a data analyst who can override or add Edit rules in the Standard Population (for example, new noise words specific to this population of data) using the Population Override Manager.

Often coinciding with this last phase is planning SSA-NAME3's implementation into the production environment. Besides enabling the search and key-building applications and key loading application in the production environment, one very important step still needs to be completed. In a real system, maintenance of the user's source data is an on-going real-time function. Therefore, the insert/update/ delete transactions which maintain the source data will also need to maintain the SSA-NAME3 Keys associated with that data and keep them synchronized.

These next stages typically require the following extra job functions:

Business Analyst:

Optionally to facilitate finding out what the users really need to achieve with the search and matching software.

Data Analyst:

Using the Population Override Manager, a data analyst trained in the syntax of the SSA-NAME3 Edit rules, and versed in the consequences of making changes that may effect search and match quality and performance, and possibly require indexes to be rebuilt, can add or remove Edit rules from the Population set. With this utility, the analyst can also replace the packaged Frequency table with one built from the organization's own data.

Project Leader / Systems Analyst:

To design and plan the way in which the SSA-NAME3 facilities will be integrated into the system.

User(s):

To use the search application, review its results, and provide feedback that may lead to further customization and tuning of the application or SSA-NAME3.

INDEX

A

analyst [17](#), [18](#)
ASCII [19](#)
ASCII text files [10](#)

B

Business Analyst [20](#)

C

Callable Routine [10](#), [13](#)
client/server [13](#)
Custom Population [18](#)
Customer Support [17](#)

D

data analyst [14](#)
Data Analyst [20](#)
Developer's Workbench [10](#), [16](#), [17](#)
Dynamic Link Library [14](#)

E

Edit Rule Wizard [13](#)

F

Files
 .ylp [10](#)
 .ysp [10](#)
foreign key [19](#)
Frequency Table [14](#)

I

Installation Sequence [16](#)

J

Java GUI [16](#)

K

Key-Build [17](#)

L

license server [16](#)

M

Match Purpose [14](#), [18](#)

N

Network Resource [14](#)

O

online search [18](#)

P

parameter [10](#)
Population Override Manager [13](#)
prototype [17](#), [18](#)

S

Sample Code [20](#)
sample programs [20](#)
search
 Person names
 Organization
 Addresses [18](#)
server software [16](#)
Shared Library [13](#)
Shared Object [14](#)
socket interface [13](#)
SSA-NAME3 [13](#)
SSA-NAME3 Callable Routine [16](#)
SSA-NAME3 Edit rule [10](#)
SSA-NAME3 key [19](#)
SSA-NAME3 keys [18](#)
SSA-NAME3 Keys [20](#)
SSA-NAME3 server [18](#)
SSA-NAME3 software [10](#)
SSA-NAME3 Workbench [14](#)
ssan3_close [18](#), [20](#)
ssan3_get_keys [18](#)
ssan3_info [18](#)
ssan3_open [18](#), [20](#)
SSANAME3 [16](#)
Standard Population [17](#)
Standard Populations [10](#)
Systems Analyst [20](#)

T

TCP/IP [13](#)

W

Win32 [14](#)

Workbench [13](#), [18](#)