Informatica®
10.1.1 HotFix 1

Referenzdaten-Handbuch

Informatica Referenzdaten-Handbuch
10.1.1 HotFix 1
Juni 2017

© Copyright Informatica LLC 2013, 2018

Diese Software und die Dokumentation werden nur im Rahmen eines eigenen Lizenzvertrags zur Verfügung gestellt, der Beschränkungen für die Verwendung und Weitergabe enthält. Ohne ausdrückliche schriftliche Genehmigung der Informatica LLC darf kein Teil dieses Dokuments zu irgendeinem Zweck vervielfältigt oder übertragen werden, unabhängig davon, auf welche Art und Weise oder mit welchen Mitteln (elektronisch, mechanisch, durch Fotokopieren, Aufzeichnen usw.) dies geschieht.

Den RECHTEN DER REGIERUNG DER VEREINIGTEN STAATEN unterliegende Programme, Software, Datenbanken und zugehörige Dokumentation und technische Daten, die an Kunden der Regierung der Vereinigten Staaten geliefert werden, sind "kommerzielle Computersoftware" oder "kommerzielle technische Daten" gemäß der anwendbaren Beschaffungsverordnung der Vereinigten Staaten (Federal Acquisition Regulation – FAR) und der ergänzenden Bestimmungen der spezifischen Behörde. Damit unterliegen die Nutzung, das Kopieren, die Offenlegung, das Modifizieren und die Anpassung den im anwendbaren Regierungsvertrag gemachten Einschränkungen und Lizenzbedingungen und, soweit im Rahmen der Bedingungen des Regierungsvertrags und der in FAR 52.227-19 aufgeführten Rechte anwendbar, der Lizenz für die kommerzielle Computersoftware.

Informatica, das Informatica-Logo und Big Data Management sind Marken oder eingetragene Marken der Informatica LLC in den Vereinigten Staaten von Amerika und zahlreichen anderen Ländern der Welt. Eine aktuelle Liste der Informatica-Marken ist im Internet auf https://www.informatica.com/trademarks.html verfügbar. Alle weiteren Produkt- und Firmennamen sind möglicherweise Markennamen oder Warenzeichen der jeweiligen Eigentümer.

Teile dieser Software und/oder Dokumentation sind durch die Urheberrechte Dritter geschützt und zwar einschließlich, ohne Einschränkung: Copyright DataDirect Technologies. Alle Rechte vorbehalten. Copyright © Sun Microsystems. Alle Rechte vorbehalten. Copyright © RSA Security Inc. Alle Rechte vorbehalten. Copyright © Ordinal Technology Corp. Alle Rechte vorbehalten. Copyright © Aandacht c.v. Alle Rechte vorbehalten. Copyright Genivia, Inc. Alle Rechte vorbehalten. Copyright Isomorphic Software. Alle Rechte vorbehalten. Copyright © Meta Integration Technology, Inc. Alle Rechte vorbehalten. Copyright © Intalio. Alle Rechte vorbehalten. Copyright © Oracle. Alle Rechte vorbehalten. Copyright © Adobe Systems Incorporated. Alle Rechte vorbehalten. Copyright © DataArt, Inc. Alle Rechte vorbehalten. Copyright © ComponentSource. Alle Rechte vorbehalten. Copyright © Microsoft Corporation. Alle Rechte vorbehalten. Copyright © Rouge Wave Software, Inc. Alle Rechte vorbehalten. Copyright © Teradata Corporation. Alle Rechte vorbehalten. Copyright © Yahoo! Inc. Alle Rechte vorbehalten. Copyright © Glyph & Cog, LLC. Alle Rechte vorbehalten. Copyright © Thinkmap, Inc. Alle Rechte vorbehalten. Copyright © Clearpace Software Limited. Alle Rechte vorbehalten. Copyright © Information Builders, Inc. Alle Rechte vorbehalten. Copyright © OSS Nokalva, Inc. Alle Rechte vorbehalten. Copyright Edifecs, Inc. Alle Rechte vorbehalten. Copyright Cleo Communications, Inc. Alle Rechte vorbehalten. Copyright © International Organization for Standardization 1986. Alle Rechte vorbehalten. Copyright © ej-technologies GmbH . Alle Rechte vorbehalten. Copyright © Jaspersoft Corporation. Alle Rechte vorbehalten. Copyright © International Business Machines Corporation. Alle Rechte vorbehalten. Copyright © yWorks GmbH. Alle Rechte vorbehalten. Copyright © Lucent Technologies. Alle Rechte vorbehalten. Copyright © University of Toronto. Alle Rechte vorbehalten. Copyright © Daniel Veillard. Alle Rechte vorbehalten. Copyright © Unicode, Inc. Copyright IBM Corp. Alle Rechte vorbehalten. Copyright © MicroQuill Software Publishing, Inc. Alle Rechte vorbehalten. Copyright © PassMark Software Pty Ltd. Alle Rechte vorbehalten. Copyright © LogiXML, Inc. Alle Rechte vorbehalten. Copyright © 2003-2010 Lorenzi Davide. Alle Rechte vorbehalten. Copyright © Red Hat, Inc. Alle Rechte vorbehalten. Copyright © The Board of Trustees of the Leland Stanford Junior University. Alle Rechte vorbehalten. Copyright © EMC Corporation. Alle Rechte vorbehalten. Copyright © Flexera Software. Alle Rechte vorbehalten. Copyright © Jinfonet Software. Alle Rechte vorbehalten. Copyright © Apple Inc. Alle Rechte vorbehalten. Copyright © Telerik Inc. Alle Rechte vorbehalten. Copyright © BEA Systems. Alle Rechte vorbehalten. Copyright © PDFlib GmbH. Alle Rechte vorbehalten. Copyright © Orientation in Objects GmbH. Alle Rechte vorbehalten. Copyright © Tanuki Software, Ltd. Alle Rechte vorbehalten. Copyright © Ricebridge. Alle Rechte vorbehalten. Copyright © Sencha, Inc. Alle Rechte vorbehalten. Copyright © Scalable Systems, Inc. Alle Rechte vorbehalten. Copyright © jQWidgets. Alle Rechte vorbehalten. Copyright © Tableau Software, Inc. Alle Rechte vorbehalten. Copyright© MaxMind, Inc. Alle Rechte vorbehalten. Copyright © TMate Software s.r.o. Alle Rechte vorbehalten. Copyright © MapR Technologies Inc. Alle Rechte vorbehalten. Copyright © Amazon Corporate LLC. Alle Rechte vorbehalten. Copyright © Highsoft. Alle Rechte vorbehalten. Copyright © Python Software Foundation. Alle Rechte vorbehalten. Copyright © BeOpen.com. Alle Rechte vorbehalten. Copyright © CNRI. Alle Rechte vorbehalten.

Dieses Produkt enthält Software, die von der Apache Software Foundation (http://www.apache.org/) entwickelt wurde, und andere Software, die unter den Bedingungen des Apache-Lizenzvertrags lizenziert ist („Lizenz“). Eine Kopie dieser Lizenzen finden Sie unter http://www.apache.org/licenses/. Sofern nicht gesetzlich vorgeschrieben oder schriftlich vereinbart, erfolgt der Vertrieb der Software unter der Lizenz auf der BASIS „WIE BESEHEN” OHNE GARANTIEN ODER KONDITIONEN IRGENDEINER ART, weder ausdrücklich noch impliziert. Berechtigungen und Einschränkungen für bestimmte Sprachen finden Sie in der Lizenz.

Dieses Produkt enthält Software, die von Mozilla (http://www.mozilla.org/) entwickelt wurde, Software Copyright The JBoss Group, LLC. Alle Rechte vorbehalten; Software Copyright © 1999-2006 by Bruno Lowagie und Paulo Soares, und andere Software, die gemäß den verschiedenen Versionen des GNU Lesser General Public License Agreement unter http://www.gnu.org/licenses/lgpl.html lizenziert ist. Die Materialien werden „wie besehen“ kostenlos von Informatica bereitgestellt, ohne ausdrückliche oder stillschweigende Gewährleistung, einschließlich, jedoch nicht beschränkt auf die stillschweigenden Gewährleistungen der Handelsüblichkeit und der Eignung für einen bestimmten Zweck.

Das Produkt enthält ACE(TM) und TAO(TM) Software, Copyright Douglas C. Schmidt und seine Forschungsgruppe an der Washington University, University of California, Irvine und Vanderbilt University, Copyright (©) 1993-2006. Alle Rechte vorbehalten.

Dieses Produkt enthält Software, die von OpenSSL Project zur Verwendung im OpenSSL Toolkit entwickelt wurde (Copyright The OpenSSL Project. Alle Rechte vorbehalten). Die erneute Verteilung dieser Software unterliegt den unter „http://www.openssl.org“ und „http://www.openssl.org/source/license.html“ verfügbaren Bedingungen.

Dieses Produkt enthält urheberrechtlich geschützte Curl-Software (Copyright 1996-2013, Daniel Stenberg, <daniel@haxx.se>). Alle Rechte vorbehalten. Die mit dieser Software verbundenen Berechtigungen und Einschränkungen unterliegen den unter „http://curl.haxx.se/docs/copyright.html“ verfügbaren Bedingungen. Die Erlaubnis, diese Software für jeden beliebigen Zweck gegen Gebühr oder kostenlos zu verwenden, zu kopieren, zu ändern und zu verteilen, wird hiermit erteilt, sofern die oben genannten urheberrechtlichen Hinweise und diese Erlaubnis in allen Exemplaren angegeben werden.

Das Produkt enthält urheberrechtlich geschützte Software, Copyright 2001-2005 (©) MetaStuff, Ltd. Alle Rechte vorbehalten. Die mit dieser Software verbundenen Berechtigungen und Einschränkungen unterliegen den unter „http://www.dom4j.org/ license.html“ verfügbaren Bedingungen.

Das Produkt enthält urheberrechtlich geschützte Software, Copyright © 2004-2007, The Dojo Foundation. Alle Rechte vorbehalten. Die mit dieser Software verbundenen Berechtigungen und Einschränkungen unterliegen den unter „http://dojotoolkit.org/license“ verfügbaren Bedingungen.

Dieses Produkt enthält urheberrechtlich geschützte ICU-Software, Copyright International Business Machines Corporation und andere. Alle Rechte vorbehalten. Die mit dieser Software verbundenen Berechtigungen und Einschränkungen unterliegen den unter „http://source.icu-project.org/repos/icu/icu/trunk/license.html“ verfügbaren Bedingungen.

Dieses Produkt enthält urheberrechtlich geschützte Software, Copyright © 1996-2006 Per Bothner. Alle Rechte vorbehalten. Das Ihnen erteilte Recht, diese Materialien zu verwenden, unterliegt den unter „http:// www.gnu.org/software/ kawa/Software-License.html“ verfügbaren Bedingungen.

Dieses Produkt enthält urheberrechtlich geschützte OSSP UUID-Software (Copyright © 2002 Ralf S. Engelschall, Copyright © 2002 The OSSP Project Copyright © 2002 Cable & Wireless Deutschland). Die mit dieser Software verbundenen Berechtigungen und Einschränkungen unterliegen den unter „http://www.opensource.org/licenses/mit-license.php“ verfügbaren Bedingungen.

Dieses Produkt enthält Software, die von Boost (http://www.boost.org/) oder unter der Softwarelizenz von Boost entwickelt wurde. Die mit dieser Software verbundenen Berechtigungen und Einschränkungen unterliegen den unter „http://www.boost.org/LICENSE_1_0.txt“ verfügbaren Bedingungen.

Dieses Produkt enthält urheberrechtlich geschützte Software, Copyright © 1997-2007 University of Cambridge. Die mit dieser Software verbundenen Berechtigungen und Einschränkungen unterliegen den unter http://www.pcre.org/license.txt einsehbaren Bedingungen.

Dieses Produkt enthält urheberrechtlich geschützte Software, Copyright © 2007 The Eclipse Foundation. Alle Rechte vorbehalten. Die mit dieser Software verbundenen Berechtigungen und Einschränkungen unterliegen den unter „http://www.eclipse.org/org/documents/epl-v10.php“ und „http://www.eclipse.org/org/documents/edl-v10.php“ verfügbaren Bedingungen.

Dieses Produkt enthält Software gemäß den Lizenzbedingungen unter 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/hsqlLicense.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.schneier.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/asl/; https://github.com/twbs/bootstrap/blob/master/LICENSE; https://sourceforge.net/p/xmlunit/code/HEAD/tree/trunk/LICENSE.txt.

Dieses Produkt enthält Software, die unter der Academic Free License (http://www.opensource.org/licenses/afl-3.0.php), der Common Development Distribution License (http://www.opensource.org/licenses/cddl1.php), der Common Public License (http://www.opensource.org/licenses/cpl1.0.php), den Sun Binary Code License Agreement Supplemental License Terms, der BSD License (http:// www.opensource.org/licenses/bsd-license.php), der neuen BSD License (http://opensource.org/licenses/BSD-3-Clause), der MIT License (http://www.opensource.org/licenses/mit-license.php), der Artistic License (http://www.opensource.org/licenses/artistic-license-1.0) und der Initial Developer's Public License Version 1.0 (http://www.firebirdsql.org/en/initial-developer-s-public-license-version-1-0/) lizenziert ist.

Dieses Produkt enthält urheberrechtlich geschützte Software, Copyright © 2003-2006 Joe Walnes, 2006-2007 XStream Committers. Alle Rechte vorbehalten. Die mit dieser Software verbundenen Berechtigungen und Einschränkungen unterliegen den unter „http://xstream.codehaus.org/license.html“ verfügbaren Bedingungen. Dieses Produkt enthält Software, die von der Indiana University Extreme! Lab. entwickelt wurde. Weitere Informationen finden Sie unter http://www.extreme.indiana.edu/.

Dieses Produkt enthält Software, Copyright © 2013 Frank Balluffi und Markus Moeller. Alle Rechte vorbehalten. Die mit dieser Software verbundenen Berechtigungen und Einschränkungen unterliegen den Bedingungen der MIT-Lizenz.

Weitere Informationen über die Patente finden Sie unter https://www.informatica.com/legal/patents.html.

HAFTUNGSAUSSCHLUSS: Informatica LLC stellt diese Dokumentation „wie besehen“ bereit, ohne ausdrückliche oder stillschweigende Gewährleistung, einschließlich, jedoch nicht beschränkt auf die Gewährleistungen der Nichtverletzung der Rechte von Dritten, der Handelsüblichkeit oder Eignung für einen bestimmten Zweck. Informatica LLC garantiert nicht die Fehlerfreiheit dieser Software oder Dokumentation. Die in dieser Software oder Dokumentation bereitgestellten Informationen können technische Ungenauigkeiten oder Druckfehler enthalten. Die in dieser Software und in dieser Dokumentation enthaltenen Informationen können jederzeit ohne vorherige Ankündigung geändert werden.

HINWEISE

Dieses Informatica-Produkt (die „Software“) umfasst bestimmte Treiber (die „DataDirect-Treiber“) von DataDirect Technologies, einem Betreiber von Progress Software Corporation („DataDirect“), die folgenden Bedingungen und Bestimmungen unterliegen:

1. DIE DATADIRECT-TREIBER WERDEN „WIE GESEHEN“ OHNE JEGLICHE GEWÄHRLEISTUNG, WEDER AUSDRÜCKLICH NOCH STILLSCHWEIGEND, BEREITGESTELLT, EINSCHLIESSLICH, JEDOCH NICHT BESCHRÄNKT AUF DIE STILLSCHWEIGENDEN GEWÄHRLEISTUNGEN DER HANDELSÜBLICHKEIT, EIGNUNG FÜR EINEN BESTIMMTEN ZWECK UND DER NICHTVERLETZUNG VON RECHTEN DRITTER.
2. IN KEINEM FALL SIND DATADIRECT ODER DRITTANBIETER DEM ENDBENUTZER GEGENÜBER HAFTBAR FÜR UNMITTELBARE, MITTELBARE, KONKRETE, NEBEN-, FOLGE- ODER ANDERE SCHÄDEN, DIE SICH AUS DER VERWENDUNG DER ODBC-TREIBER ERGEBEN, UNABHÄNGIG DAVON, OB SIE IM VORAUS ÜBER DIE MÖGLICHKEIT SOLCHER SCHÄDEN INFORMIERT WORDEN SIND ODER NICHT. DIESE BESCHRÄNKUNGEN GELTEN FÜR ALLE KLAGEGEGENSTÄNDE, EINSCHLIESSLICH, JEDOCH NICHT BESCHRÄNKT AUF VERTRAGSBRUCH, GEWÄHRLEISTUNGSBRUCH, FAHRLÄSSIGKEIT, KAUSALHAFTUNG, TÄUSCHUNG UND ANDERE UNERLAUBTE HANDLUNGEN.

Die in dieser Dokumentation enthaltenen Informationen können jederzeit ohne vorherige Ankündigung geändert werden. Wenn Sie Probleme in dieser Dokumentation finden, zeigen Sie uns diese bitte schriftlich an: Informatica LLC 2100 Seaport Blvd. Redwood City, CA 94063, USA.

Informatica-Produkte unterliegen einer Gewährleistung gemäß den Geschäftsbedingungen der Vereinbarungen, unter denen sie bereitgestellt werden. INFORMATICA STELLT DIE INFORMATIONEN IN DIESEM DOKUMENT OHNE MÄNGELGEWÄHR UND OHNE AUSDRÜCKLICHE ODER STILLSCHWEIGENDE GEWÄHRLEISTUNG JEGLICHER ART ZUR VERFÜGUNG. DIES GILT EINSCHLIESSLICH FÜR GEWÄHRLEISTUNGEN DER MARKTGÄNGIGKEIT, DER EIGNUNG FÜR EINEN BESTIMMTEN ZWECK UND GEWÄHRLEISTUNGEN ODER ZUSICHERUNGEN ÜBER DIE NICHTVERLETZUNG VON RECHTEN DRITTER.

Publikationsdatum: 2018-06-08

Inhalt

Einleitung 8

Informatica-Ressourcen. 8

Informatica-Netzwerk. 8

Informatica-Wissensdatenbank. 8

Informatica-Dokumentation. 8

Informatica-Produktverfügbarkeitsmatrizen. 9

Informatica Velocity. 9

Informatica Marketplace. 9

Globaler Kundensupport von Informatica. 9

Kapitel 1: Einführung in Referenzdaten. 10

Referenzdaten – Übersicht. 10

Informatica Referenzdaten. 11

Benutzerdefinierte Referenzdaten. 11

Referenztabellen. 12

Referenztabellenstruktur. 12

Referenzdaten-Warehouse – Berechtigungen. 13

Parameter und Referenztabellen. 13

Referenzdatenobjekte und Versionskontrolle. 14

Kapitel 2: Referenztabellen im Analyst-Tool. 15

Analyst-Tool-Referenztabellen – Übersicht. 15

Referenztabelleneigenschaften. 15

Allgemeine Eigenschaften der Referenztabelle. 16

Spalteneigenschaften der Referenztabelle. 16

Erstellen einer Referenztabelle im Referenztabellen-Editor. 17

Erstellen einer Referenztabelle aus Profildaten. 18

Erstellen einer Referenztabelle aus Profilspaltendaten. 18

Erstellen einer Referenztabelle aus Wertmustern. 19

Erstellen einer Referenztabelle aus einer Einfachdatei. 21

Eigenschaften der Analysetool-Einfachdatei. 21

Erstellen einer Referenztabelle aus einer Einfachdatei. 21

Erstellen einer Referenztabelle aus einer Datenbanktabelle. 22

Erstellen einer Referenztabelle aus einer Datenbanktabelle. 22

Arbeiten mit Referenztabellen in einem versionierten Modellrepository. 23

Referenztabellenaktualisierung. 24

Verwalten von Spalten. 24

Verwalten von Zeilen. 25

Werte suchen und ersetzen. 26

Exportieren von Referenztabellendaten. 26

Aktivieren und Deaktivieren der Editierbarkeit in nicht verwalteten Referenztabellen. 27
Aktualisieren Sie die Referenztabellenwerte. 27
Audit-Trail-Ereignisse. 28
Audit-Trail-Ereignisse anzeigen. 28
Regeln und Richtlinien für Referenztabellen. 29

Kapitel 3: Referenzdaten im Developer-Tool. 30

Developer-Tool-Referenzdaten – Übersicht. 30
Referenzdaten und Umwandlungen. 31
Arbeiten mit Referenzdatenobjekten in einem versionierten Modellrepository. 31
Auschecken von Referenzdatenobjekten. 31
Einchecken von Referenzdatenobjekten. 32
Referenztabellen. 32
Eigenschaften von Referenztabellendaten. 33
Erstellen eines Referenztabellenobjekts. 33
Erstellen einer Referenztabelle aus einer Einfachdatei. 34
Erstellen einer Referenztabelle aus einer relationalen Quelle. 35
Inhaltssätze. 37
Zeichensätze. 37
Klassifizierermodelle. 38
Mustersätze. 38
Probabilistische Modelle. 38
Reguläre Ausdrücke. 39
Tokensätze. 40
Regeln und Richtlinien für probabilistische Modelle und Klassifizierermodelle. 41
Erstellen eines Inhaltssatzes. 42
Erstellen eines Referenzdatenobjekts in einem Inhaltssatz. 43

Kapitel 4: Klassifizierermodelle. 44

Klassifikatormodelle – Übersicht. 44
Struktur von Klassifikatormodellen. 45
Klassifikator-Scores. 45
Beispiel für eine Klassifikatorumwandlung. 46
Optionen des Klassifizierermodells. 46
Referenzdaten in Klassifikatormodellen. 48
Beschriftungsdaten in Klassifikatormodellen. 48
Beschriftungsverwaltung im Klassifizierermodell. 49
Konfiguration des Klassifizierermodells. 50
Erstellen eines Klassifizierermodells. 50
Anhängen von Daten aus einer Datenquelle an ein Klassifizierermodell 51
Hinzufügen einer Referenzdatenzeile zu einem Klassifizierermodell. 52
Hinzufügen einer Beschriftung zu einem Klassifikatormodell. 52
Zuweisen einer Beschriftung zu Referenzdatenzeilen. 52

Erkennen ungenutzter Beschriftungswerte. . . . 53
Löschen von Zeilen aus einem Klassifizierermodell. . . . 53
Löschen einer Beschriftung aus einem Klassifikatormodell. . . . 53
Kompilieren eines Klassifikatormodells. . . . 54
Filter- und Suchvorgänge. . . . 54
Verwenden eines Datenwerts zum Filtern der Referenzdatenzeilen. . . . 54
Verwenden eines Beschriftungswerts zum Filtern der Referenzdatenzeilen. . . . 55
Suchen eines Werts in einer Referenzdatenzeile. . . . 55
Kopieren und Einfügen. . . . 55
Kopieren eines Klassifikatormodells in einen anderen Inhaltssatz. . . . 55
Importieren eines Klassifikatormodells aus einem anderen Inhaltssatz. . . . 56

Kapitel 5: Probabilistische Modelle. . . . 57

Probabilistische Modelle – Übersicht. . . . 57
Struktur des probalistischen Modells. . . . 58
Beispiel für eine Beschriftungsumwandlung. . . . 58
Beispiel für eine Parser-Transformation. . . . 60
Optionen des probabilistischen Modells. . . . 60
Datenansicht des probabilistischen Modells. . . . 60
Beschriftungsansicht des probabilistischen Modells. . . . 62
Referenzdaten in probabilistischen Modellen. . . . 63
Beschriftungsdaten in probabilistischen Modellen. . . . 63
Overflow-Beschriftung. . . . 64
Eigenschaften des probabilistischen Modells. . . . 65
Konfiguration des probabilistischen Modells. . . . 65
Erstellen eines leeren probabilistischen Modells. . . . 66
Erstellen eines probabilistischen Modells aus einem Datenobjekt. . . . 67
Anhängen von Daten aus einer Datenquelle an ein probabilistisches Modell. . . . 68
Hinzufügen einer Referenzdatenzeile zu einem probabilistischen Modell. . . . 68
Hinzufügen einer Beschriftung zu einem probabilistischen Modell. . . . 69
Zuweisen einer Beschriftung zu einem Referenzdatenwert. . . . 69
Zuweisen einer Beschriftung zu mehreren Datenwerten . . . 70
Löschen von Zeilen aus einem probabilistischen Modell. . . . 70
Löschen einer Beschriftung aus einem probabilistischen Modell. . . . 71
Kompilieren des probabilistischen Modells. . . . 71
Suchen nach Datenzeilen in einem probabilistischen Modell. . . . 71
Filtern von Referenzdatenwerten nach Beschriftungszuweisungen. . . . 71
Suchen nach ungenutzten Beschriftungswerten. . . . 72
Kopieren und Einfügen. . . . 72
Kopieren eines probabilistischen Modells in einen anderen Inhaltssatz. . . . 73
Importieren eines probabilistischen Modells aus einem anderen Inhaltssatz. . . . 73
Kopieren von Referenzdatenzeilen in die Zwischenablage. . . . 73

Anhang A: Referenzdaten und Informatica Big Data Management..... 74

Referenzdaten und Informatica Big Data Management - Übersicht. 74

Referenzdaten für die Adressvalidierung. 74

Installieren der Adressenreferenzdaten-Dateien. 75

Index. 77

Einleitung

Das *Informatica-Referenzdaten-Handbuch* enthält Informationen über die Referenzdatenobjekte und Dateien, die in Informatica Developer und Informatica Analyst verwendet werden können. Es wurde für Datenanalysten, Datenverwalter und andere Personen geschrieben, die Referenzdaten verwenden, um die Genauigkeit und Benutzerfreundlichkeit von Unternehmensdaten zu überprüfen und zu optimieren.

Informatica-Ressourcen

Informatica-Netzwerk

Im Informatica-Netzwerk finden Sie den globalen Kundensupport von Informatica, die Informatica-Wissensdatenbank und andere Produktressourcen. Für den Zugriff auf das Informatica-Netzwerk besuchen Sie https://network.informatica.com.

Als Mitglied können Sie:

- zentral auf alle Ihre Informatica-Ressourcen zugreifen.
- Durchsuchen Sie die Wissensdatenbank nach Produktressourcen, einschließlich Dokumentation, häufig gestellter Fragen und bewährter Methoden.
- Zeigen Sie Informationen zur Produktverfügbarkeit an.
- Ihre Support-Fälle prüfen.
- Ihr lokales Informatica-Netzwerk für Benutzergruppen suchen und mit anderen Benutzern zusammenarbeiten.

Informatica-Wissensdatenbank

Verwenden Sie die Informatica-Wissensdatenbank, um das Informatica-Netzwerk nach Produktressourcen, wie z. B. Dokumentation, Ratgeberartikeln, bewährten Methoden und PAMs, zu durchsuchen.

Für den Zugriff auf die Wissensdatenbank besuchen Sie https://kb.informatica.com. Wenn Sie Fragen, Kommentare oder Ideen zur Wissensdatenbank haben, wenden Sie sich per E-Mail an das Team der Informatica-Wissensdatenbank unter KB_Feedback@informatica.com.

Informatica-Dokumentation

Navigieren Sie zur Informatica-Wissensdatenbank unter https://kb.informatica.com/_layouts/ProductDocumentation/Page/ProductDocumentSearch.aspx, um die aktuelle Dokumentation für Ihr Produkt abzurufen.

Wenn Sie Fragen, Kommentare oder Ideen zu dieser Dokumentation haben, wenden Sie sich per E-Mail an das Informatica-Dokumentationsteam unter infa_documentation@informatica.com.

Informatica-Produktverfügbarkeitsmatrizen

Produktverfügbarkeitsmatrizen (PAMs) geben die Versionen der Betriebssysteme, Datenbanken und anderen Typen von Datenquellen und Zielen an, die in einer Produktversion unterstützt werden. Als Mitglied des Informatica-Netzwerks können Sie unter https://network.informatica.com/community/informatica-network/product-availability-matrices auf PAMs zugreifen.

Informatica Velocity

Bei Informatica Velocity handelt es sich um eine Sammlung von Tipps und bewährten Methoden, die von den professionellen Informatica-Diensten entwickelt wurden. Informatica Velocity basiert auf der Praxiserfahrung aus Hunderten von Datenmanagementprojekten und umfasst das kollektive Wissen unserer Berater, die mit Unternehmen aus der ganzen Welt an der Planung, Entwicklung, Bereitstellung und Wartung erfolgreicher Datenmanagementlösungen gearbeitet haben.

Als Mitglied des Informatica-Netzwerks können Sie unter http://velocity.informatica.com auf Informatica Velocity-Ressourcen zugreifen.

Wenn Sie Fragen, Anregungen oder Ideen zu Informatica Velocity haben, wenden Sie sich an die professionellen Informatica-Dienste unter ips@informatica.com.

Informatica Marketplace

Informatica Marketplace ist ein Forum, das Lösungen zur Erweiterung und Verbesserung Ihrer Informatica-Implementierungen bereitstellt. Indem Sie die zahlreichen Lösungen von Informatica-Entwicklern und -Partnern nutzen, können Sie Ihre Produktivität steigern und die Implementierungsdauer Ihrer Projekte verkürzen. Zugriff auf den Informatica Marketplace erhalten Sie unter https://marketplace.informatica.com.

Globaler Kundensupport von Informatica

Sie können sich telefonisch oder über den Online-Support mit einem globalen Support-Center im Informatica-Netzwerk in Verbindung setzen.

Die Telefonnummer des globalen Kundensupports von Informatica vor Ort finden Sie auf der Informatica-Website unter folgender Verknüpfung: http://www.informatica.com/us/services-and-training/support-services/global-support-centers.

Als Mitglied des Informatica-Netzwerks können Sie den Online-Support unter http://network.informatica.com verwenden.

KAPITEL 1

Einführung in Referenzdaten

Dieses Kapitel umfasst die folgenden Themen:

- Referenzdaten – Übersicht, 10
- Informatica Referenzdaten, 11
- Benutzerdefinierte Referenzdaten, 11
- Referenztabellen, 12
- Referenzdatenobjekte und Versionskontrolle, 14

Referenzdaten – Übersicht

Informatica-Umwandlungen können Referenzdaten verwenden, um Daten zu analysieren und zu aktualisieren. Sie können Referenzdatenobjekte im Developer tool und im Analyst-Tool erstellen. Sie können Referenzdatenobjekte und Dateien auch in das Modellrepository und das Dateisystem importieren. Sie können das Installationsprogramm für Data Quality-Inhalte verwenden, um Referenzdatenobjekte zu importieren und Referenzdatendateien zu installieren.

Sie können die folgenden Typen von Referenzdaten erstellen und bearbeiten:

Referenztabellen

Eine Referenztabelle enthält die Standardversion und alternative Versionen einer Gruppe von Datenwerten. Im Developer tool fügen Sie eine Referenztabelle zu einer Umwandlung hinzu, um sicherzustellen, dass Quelldatenwerte genau und richtig formatiert sind.

Die meisten Referenztabellen enthalten mindestens zwei Spalten. Eine Spalte enthält die Standardversion oder bevorzugte Version eines Werts und andere Spalten enthalten alternative Versionen. Wenn Sie eine Referenztabelle zu einer Umwandlung hinzufügen, durchsucht die Umwandlung die Eingabeportdaten nach Werten, die auch in der Tabelle erscheinen. Sie können Tabellen mit allen Daten erzeugen, die für das Datenprojekt, an dem Sie arbeiten, nützlich sind.

Inhaltssätze

Ein Inhaltssatz ist ein Modellrepository-Objekt, das Referenzdatenwerte im Repository oder in einer Datei angibt. Wenn Sie einer Umwandlung einen Inhaltssatz hinzufügen, durchsucht die Umwandlung die Eingabedaten nach Werten, die den Datenmustern im Inhaltssatz entsprechen.

Das Data Quality Content-Installationsprogramm kann die folgenden Referenzdatentypen installieren:

Informatica-Referenztabellen

Von Informatica entwickelte Repository-Objekte und Datendateien. Sie importieren Informatica-Referenztabellen beim Importieren von Beschleunigerobjekten in das Modellrepository. Die Typen von Referenzinformationen umfassen Ortsvorwahlen, Postleitzahlformate, Vornamen, Positionen und Akronyme. Sie können die Informatica-Referenztabellen bearbeiten.

Informatica-Inhaltssätze

Von Informatica entwickelte Repository-Objekte und Datendateien. Sie importieren Inhaltssätze beim Importieren von Beschleunigerobjekten in das Modellrepository. Ein Inhaltssatz enthält verschiedene Referenzdatentypen, die zum Durchführen von Suchläufen mit Datenqualitätsumwandlungen verwendet werden können.

Adressenreferenzdaten-Dateien

Referenzdatendateien, die Daten für die zustellungsfähigen Adressen in einem Land enthalten. Die Adress-Validiererumwandlung liest die Referenzdaten. Sie können Adressreferenzdaten-Dateien nicht erzeugen oder bearbeiten.

Adressreferenzdaten sind für einen festgelegten Zeitraum aktuell und müssen regelmäßig aktualisiert werden, zum Beispiel jedes Quartal.

Identitätspopulationsdateien

Referenzdatendateien, die Informationen zu Personen, Haushalt und Unternehmensidentitäten enthalten. Die Match- und die Vergleichsumwandlung verwenden Populationsdateien, um nach potenziellen Identitäten in Eingabedaten zu suchen. Sie können Identitätspopulationsdateien nicht erzeugen und bearbeiten.

Informatica Referenzdaten

Sie können von Informatica Adressreferenzdaten und Identitätspopulationsdaten erwerben und herunterladen.

Sie erwerben für Adressdaten für ein Land eine jährliche Subskription und können dann jederzeit während der Laufzeit der Subskription die aktuellsten Adressdaten von Informatica herunterladen.

Der Benutzer des Inhalts-Installationsprogramms lädt Referenzdaten herunter und installiert diese getrennt von den Anwendungen. Setzen Sie sich mit einem Administrator in Verbindung, wenn Sie weitere Informationen zu den auf Ihrem System installierten Referenzdaten erhalten möchten.

Benutzerdefinierte Referenzdaten

Sie können die Werte in einem Datenobjekt zur Erzeugung eines Referenzdatenobjekts benutzen.

Zum Beispiel können Sie ein Datenobjekt oder eine Profilspalte auswählen, die Werte enthalten, die für ein Projekt oder eine Organisation spezifisch sind. Erstellen Sie benutzerdefinierte Referenzdatenobjekte aus den Spaltenwerten.

Sie können ein Referenzdatenobjekt aus einer Datenspalte erstellen, um Folgendes zu überprüfen:

- Die Datenzeilen in der Spalte enthalten den selben Typ von Informationen.

- Ein Quellwert ist gültig. Das Referenzobjekt enthält möglicherweise eine Liste der gültigen Werte oder das Referenzobjekt enthält möglicherweise eine Liste von Werten, die nicht gültig sind.

Die folgende Tabelle enthält häufige Beispiele für Projektdatenspalten, die Referenzdaten enthalten können:

Informationen	Referenzdaten-Beispiel
Lagermengeneinheits (SKU)-Codes	Benutzen Sie eine SKU-Spalte, um eine Referenztabelle von gültigen SKU-Codes für eine Organisation zu erzeugen. Benutzen Sie die Referenztabelle, um richtige oder falsche SKU-Codes in einem Datensatz zu finden.
Mitarbeiter-Codes	Mit Hilfe eines Mitarbeiter-Codes oder einer Mitarbeiter-ID-Spalte können Sie eine Referenztabelle mit gültigen Mitarbeiter-Codes erzeugen. Benutzen Sie die Referenztabelle zum Auffinden von Fehlern in Mitarbeiterdaten.
Kundenkonten-Nummern	Lassen Sie ein Profil über eine Kundenkonten-Nummernspalte laufen, um Muster in den Kontennummern zu identifizieren. Erzeugen Sie mit Hilfe des Profils ein Token-Set mit falschen Datenmustern. Mit Hilfe des Token-Sets können Sie Kontennummern herausfinden, die nicht der richtigen Kontennummernstruktur entsprechen.
Kundennamen	Wenn eine Spalte für Kundennamen Vor-, zweiten Vor- sowie Nachnamen enthält, können Sie ein probabilistisches Modell erzeugen, das die erwartete String-Struktur in der Spalte definiert. Benutzen Sie das probabilistische Modell zum Auffinden von Datenstrings, die nicht in die Spalte gehören.

Referenztabellen

Erstellen und aktualisieren Sie Referenztabellen im Analyst-Tool und im Developer tool.

Referenztabellen speichern Metadaten im Modellrepository. Referenztabellen können Spaltendaten im Referenzdaten-Warehouse oder in einer anderen Datenbank speichern. Wenn das Referenzdaten-Warehouse die Spaltendaten speichert, identifizieren die Informatica-Dienste die Tabelle als verwaltete Referenztabelle. Wenn eine andere Datenbank die Spaltendaten speichert, identifizieren die Informatica-Dienste die Tabelle als nicht verwaltete Referenztabelle.

Der Content-Managementdienst speichert die Datenbankverbindung zum Referenzdaten-Warehouse. Sie können eine IBM DB2-Datenbank, eine Microsoft SQL Server-Datenbank oder eine Oracle-Datenbank als Referenzdaten-Warehouse angeben.

Wenn Sie Daten aus einer anderen Datenbank in das Referenzdaten-Warehouse importieren, verwenden Sie eine native Verbindung oder eine ODBC-Verbindung für den Import der Daten. Wenn Sie eine nicht verwaltete Datenbank als Datenquelle für eine Referenztabelle angeben, verwenden Sie eine native Verbindung zum Herstellen einer Verbindung zur Datenbank.

Referenztabellenstruktur

Die meisten Referenztabellen enthalten mindestens zwei Spalten. Eine Spalte enthält die richtigen oder erforderlichen Versionen der Datenwerte. Andere Spalten enthalten verschiedene Versionen der Werte, einschließlich alternative Versionen, die möglicherweise in der Datenquelle angezeigt werden.

Die Spalte, in der die richtigen oder erforderlichen Werte enthalten sind, wird als gültige Spalte bezeichnet. Wenn eine Umwandlung eine Referenztabelle in einem Mapping liest, sucht die Umwandlung nach Werten in den ungültigen Spalten. Wenn die Umwandlung einen ungültigen Wert findet, gibt sie den entsprechenden

Wert aus der gültigen Spalte zurück. Sie können eine Umwandlung auch so konfigurieren, dass nicht die gültigen Werte zurückgegeben werden, sondern ein einzelner allgemeiner Wert.

Die gültige Spalte kann Daten enthalten, die formal richtig sind, z. B. Postleitzahlen. Sie kann Daten enthalten, die für ein Projekt relevant sind, zum Beispiel SKU (Lagermengeneinheit)-Nummern, die für eine Organisation eindeutig sind. Sie können eine gültige Spalte auch aus falschen Daten erstellen, wie zum Beispiel aus Werten, die bekannte Datenfehler enthalten, nach denen Sie suchen möchten.

Beispiel: Sie erstellen eine Referenztabelle, die eine Liste gültiger SKU-Nummern in einer Einzelhandelsorganisation enthält. Fügen Sie die Referenztabelle zu einer Beschriftungsumwandlung hinzu und erstellen Sie ein Mapping mit der Umwandlung. Führen Sie das Mapping mit einer Produktdatenbanktabelle durch. Wenn das Mapping ausgeführt wird, erstellt die Beschriftungsumwandlung eine Spalte, die die Produktdatensätze angibt, die keine gültigen SKU-Nummern enthalten.

Referenztabellen und die Parserumwandlung

Erstellen Sie eine Referenztabelle mit einer einzelnen Spalte, um die Tabellendaten in einem musterbasierten Parsing-Vorgang zu verwenden. Konfigurieren Sie die Parser-Umwandlung, um musterbasiertes Parsing durchzuführen, und importieren Sie die Daten in die Umwandlungskonfiguration.

Referenzdaten-Warehouse – Berechtigungen

Der Content-Managementdienst verwendet Berechtigungen zum Einschränken der Benutzeraktionen für Referenztabellen. Überprüfen oder aktualisieren Sie die Rechte des Service mithilfe der Sicherheitsoptionen im Administrator-Tool.

Zum Arbeiten mit den Referenztabellen müssen Sie über die folgenden Berechtigungen im Content-Managementdienst verfügen:

- Referenztabellen erstellen
- Referenztabellendaten bearbeiten
- Metadata von Referenztabellen bearbeiten

Um Daten in einer nicht verwalteten Referenztabelle zu bearbeiten, stellen Sie außerdem sicher, dass Sie das Referenztabellenobjekt so konfiguriert haben, dass es Bearbeitungen zulässt.

Hinweis: Wenn Sie die Metadaten einer nicht verwalteten Referenztabelle in einer Datenbankanwendung bearbeiten, synchronisieren Sie im Analyst-Tool das Modellrepository mit der Tabelle. Das Modellrepository und die Tabelle müssen synchronisiert werden, bevor die nicht verwaltete Referenztabelle im Developer tool verwendet werden kann.

Parameter und Referenztabellen

Sie können Parameter verwenden, um Referenztabellen im Modellrepository anzugeben. Sie können einen Parameter im Developer Tool erstellen, der die Referenztabelle angibt. Sie können aber auch den Speicherort der Referenztabelle zu einer Parameterdatei hinzufügen.

Wenn Sie einen Parameter im Developer Tool erstellen, fügen Sie ihn einer Umwandlung in einer Zuordnung hinzu. Wenn Sie den Speicherort der Referenztabelle zu einer Parameterdatei hinzufügen, geben Sie die Datei an, wenn Sie eine Zuordnung an der Eingabeaufforderung ausführen. In jedem Fall liest der Datenintegrationsdienst die Referenztabelle, die vom Parameter beim Ausführen der Zuordnung angegeben wird.

Sie können einen Parameter, der eine Referenztabelle angibt, zu folgenden Umwandlungen hinzufügen:

- Groß-/Kleinschreibungsumwandlung
- Bezeichnungsumwandlung

- Parser-Umwandlung im Token-Parsing-Modus
- Standardisierer-Umwandlung

Hinweis: Verwenden Sie den Befehl *infacmd ms runMapping*, um eine Zuordnung an der Eingabeaufforderung auszuführen.

Referenzdatenobjekte und Versionskontrolle

Wenn das Modellrepository, in dem die Referenzdatenobjekte gespeichert werden, mit einer Versionskontrollanwendung integriert werden kann, können Sie die Versionskontrolle auf die Objekte anwenden. Sie können die Versionskontrolle auf Referenztabellen und Inhaltssätze anwenden.

Sie können Referenzdatenobjekte einchecken oder aus einem Modellrepository auschecken, das die Versionskontrolle unterstützt. Sie können das Auschecken rückgängig machen, eine frühere Version eines Objekts abrufen und die frühere Version des Objekts wiederherstellen. Wenn die Versionskontrolle für die Referenzdatenobjekte nicht aktiviert ist, sperrt das Modellrepository ein Referenzdatenobjekt, das Sie bearbeiten. Andere Benutzer können ein gesperrtes Objekt, an dem Sie arbeiten, nicht bearbeiten. Wenn Sie das Objekt schließen, wird die Sperre vom Modellrepository aufgehoben und andere Benutzer können das Objekt bearbeiten.

Hinweis: Die Versionskontrolle wird auf die Metadaten angewendet, die vom Modellrepository für ein nicht verwaltetes Referenztabellenobjekt gespeichert werden. Die Versionskontrolle wird auf Daten in einer nicht verwalteten Referenztabelle nicht angewendet. Sie können die Referenzdaten aus einer früheren Version einer nicht verwalteten Referenztabelle weder anzeigen noch wiederherstellen.

KAPITEL 2

Referenztabellen im Analyst-Tool

Dieses Kapitel umfasst die folgenden Themen:

- Analyst-Tool-Referenztabellen – Übersicht, 15
- Referenztabelleneigenschaften, 15
- Erstellen einer Referenztabelle im Referenztabellen-Editor, 17
- Erstellen einer Referenztabelle aus Profildaten, 18
- Erstellen einer Referenztabelle aus einer Einfachdatei, 21
- Erstellen einer Referenztabelle aus einer Datenbanktabelle, 22
- Arbeiten mit Referenztabellen in einem versionierten Modellrepository, 23
- Referenztabellenaktualisierung, 24
- Audit-Trail-Ereignisse, 28
- Regeln und Richtlinien für Referenztabellen, 29

Analyst-Tool-Referenztabellen – Übersicht

Erstellen Sie Referenztabellen im Designarbeitsbereich des Analyst-Tools.

Sie können eine Referenztabelle aus einer Einfachdatei, einer Datenquelle im Modellrepository und aus einer Tabelle in einer anderen Datenbank erstellen.

Sie können eine Referenztabelle aus einer Profilspalte oder einer Teilmenge der Daten in einer Profilspalte erstellen. Sie können auch eine Referenztabelle aus Spaltenmustern erstellen, die Sie aus einem Profil auswählen.

Beim Erstellen oder Aktualisieren einer Referenztabelle konfigurieren Sie die Eigenschaften in der Tabelle und den enthaltenen Datenspalten.

Referenztabelleneigenschaften

Sie können Referenztabelleneigenschaften im Analyst Tool anzeigen und aktualisieren. Eine Referenztabelle zeigt allgemeine Eigenschaften und Spalteneigenschaften an. Die allgemeinen Eigenschaften umfassen den Namen der Referenztabelle, das Erstellungsdatum, den Verbindungsnamen für die Datenbank und den

gültigen Spaltennamen. Die Spalteneigenschaften umfassen die Spaltennamen, die Werte der Gesamtstellenanzahl und die Werte der Größenordnung.

Sie können die Eigenschaften im schreibgeschützten Modus anzeigen. Bearbeiten Sie die Referenztabelle oder checken Sie diese aus, um die Eigenschaften zu aktualisieren.

Allgemeine Eigenschaften der Referenztabelle

Die allgemeinen Eigenschaften enthalten Informationen zum Referenztabellenobjekt.

In der folgenden Tabelle werden die allgemeinen Eigenschaften beschrieben:

Eigenschaft	Beschreibung
Name	Der Name der Referenztabelle.
Beschreibung	Alle Beschreibungen, die ein Benutzer für die Referenztabelle eingegeben hat.
Speicherort	Der Speicherort des Referenztabellenobjekts im Modellrepository.
Gültige Spalte	Der Name der gültigen Spalte in der Referenztabelle.
Erstellt am	Das Erstellungsdatum und die Erstellungszeit für den Referenztabellennamen.
Erstellt von	Der Anmeldename des Benutzers, der die Referenztabelle erstellt hat.
Zuletzt geändert	Das Datum und die Uhrzeit der letzten Aktualisierung der Referenztabelle.
Zuletzt geändert von	Der Anmeldename des Benutzers, der die letzte Aktualisierung durchgeführt hat.
Verbindungsname	Der Verbindungsname für die Datenbank, die die Referenzdatenwerte speichert.
Typ	Der Typ der Referenztabelle. Die Referenztabelle kann verwaltet oder nicht verwaltet sein.

Spalteneigenschaften der Referenztabelle

Die Spalteneigenschaften enthalten Informationen über die Spaltenmetadaten.

In der folgenden Tabelle werden die Spalteneigenschaften beschrieben:

Eigenschaft	Beschreibung
Name	Der Spaltenname.
Datentyp	Der Datentyp für die Daten in jeder Spalte. Sie können einen der folgenden Datentypen auswählen: - Bigint - Datum/Uhrzeit - Dezimal - Doppel - Ganzzahl - Zeichenfolge Beim Erstellen einer leeren Referenztabelle oder einer Referenztabelle aus einer Einfachdatei können Sie keinen doppelten Datentyp auswählen.

Eigenschaft	Beschreibung
Gesamtstellenanzahl	Die Gesamtstellenanzahl für jede Spalte. Gesamtstellenanzahl ist die maximale Anzahl an Zahlen oder die maximale Anzahl an Zeichen, die die Spalte aufnehmen kann. Die Gesamtstellenanzahl, die Sie konfigurieren können, hängt vom Datentyp ab.
Größenordnung	Die Größenordnung für jede Spalte. Größenordnung ist die maximale Anzahl an Zahlen, die eine Spalte rechts neben dem Dezimalkomma aufnehmen kann. Gilt für dezimale Spalten. Die Größenordnungswerte, die Sie konfigurieren können, hängen vom Datentyp ab.
Beschreibung	Eine optionale Beschreibung für jede Spalte.
Nullwerte zulassen	Gibt an, ob die Spalte Nullwerte enthalten darf.
Schlüssel	Erkennt eine Schlüsselspalte. Das Analyst Tool kann eine Schlüsselspalte erkennen, wenn Sie die Referenzdaten aus einer Tabelle importieren, die eine Schlüsselspalte angibt.

Erstellen einer Referenztabelle im Referenztabellen-Editor

Definieren Sie die Struktur der Tabelle und fügen Sie Daten zu einer Referenztabelle im Referenztabellen-Editor hinzu.

1. Klicken Sie auf **Neu** > **Referenztabelle** .

   Der Assistent **Neue Referenztabelle** wird geöffnet.
2. Wählen Sie die Option **Referenztabellen-Editor verwenden** und klicken Sie auf **Weiter**.
3. Verwenden Sie die Option **Neue Spalte hinzufügen**, um Spalten zu der Tabelle hinzuzufügen.
4. Konfigurieren Sie die Eigenschaften für jede Spalte.

   Die Eigenschaften umfassen Spaltenname, Datentyp, Gesamtstellenanzahl und Größenordnung.

   Wenn die Spalte Daten enthält, die eine Umwandlung bei einer Referenzdatensuche zurückgeben kann, wählen Sie die Option „Gültig“ aus.
5. Optional können Sie eine Spalte zum Einbeziehen von Beschreibungen auf niedriger Ebene als Metadaten in der Referenztabelle hinzufügen.
6. Optional können Sie eine Audit-Bemerkung für die Tabelle eingeben.

   Die Audit-Bemerkung wird im Audit-Trail-Protokoll angezeigt.
7. Klicken Sie auf **Weiter**.
8. Geben Sie einen Namen für die Referenztabelle ein und wählen Sie einen Speicherort für das Referenztabellenobjekt im Modellrepository aus.
9. Klicken Sie auf **Fertig stellen**.

Erstellen einer Referenztabelle aus Profildaten

Sie können Profildaten zum Erstellen von Referenztabellen verwenden, die sich auf die Quelldaten im Profil beziehen. Verwenden Sie die Referenztabellen, um unterschiedliche Typen von Informationen in der Datenquelle zu suchen.

Sie können ein Profil verwenden, um eine Referenztabelle wie folgt zu erstellen oder zu aktualisieren:

- Wählen Sie im Profil eine Spalte aus und fügen Sie sie einer Referenztabelle hinzu.
- Durchsuchen Sie eine Profilspalte und fügen Sie eine Teilmenge der Spaltendaten zu einer Referenztabelle hinzu.
- Wählen Sie im Profil eine Spalte aus und fügen Sie die Musterwerte für diese Spalte zu einer Referenztabelle hinzu.

Erstellen einer Referenztabelle aus Profilspaltendaten

Sie können eine Referenztabelle aus einem oder mehreren Spaltenwerten in einer Profildatenspalte erstellen. Wählen Sie eine Spalte in einem Profil und wählen Sie die Spaltenwerte aus, die zur Referenztabelle hinzugefügt werden sollen.

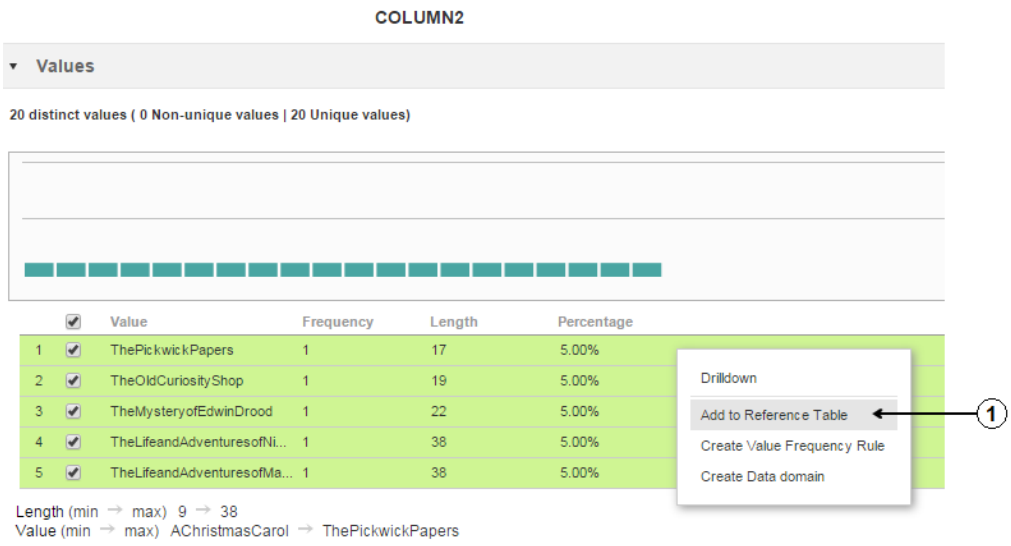
1. Öffnen Sie den Arbeitsbereich **Bibliothek** im Analyst Tool.
2. Wählen Sie die Objektkategorie **Profile** aus.

   Die Inhaltsbibliothek zeigt eine Liste der Profile im Modellrepository an.
3. Öffnen Sie das Profil, das die Spalte enthält, die zur Referenztabelle hinzugefügt werden soll.

   In der Profilübersicht werden die Namen der Profilspalten aufgelistet.
4. Überprüfen Sie die Spaltendaten.

   Klicken Sie auf den Namen der Spalte, um die Spaltendaten anzuzeigen.
5. Wählen Sie in der detaillierten Profilansicht die Datenwerte aus, die zur Referenztabelle hinzugefügt werden sollen. Sie können Werte einzeln nacheinander oder alle Werte auswählen.
6. Klicken Sie mit der rechten Maustaste auf den Namen der Spalte und wählen Sie **Hinzufügen zu Referenztabelle** aus.

Die folgende Abbildung zeigt eine Datenspalte in der detaillierten Profilansicht:



Die Nummer 1 gibt die Option **Hinzufügen zu Referenztabelle** in der Abbildung an.

7. Der Assistent **Hinzufügen zu Referenztabelle** wird geöffnet.

   Wählen Sie die Option **Erstellen einer Referenztabelle**.

   Hinweis: Sie können auch eine Option auswählen, um die Daten zu einer aktuellen Referenztabelle hinzuzufügen.

8. Klicken Sie auf **Weiter**.

   Der Spaltenname wird standardmäßig als Referenztabellenname angezeigt. Optional können Sie den Namen aktualisieren.

9. Geben Sie optional eine Beschreibung und einen Standardwert ein.

   Das Analyst Tool verwendet den Standardwert für jeden Tabellendatensatz, der keinen Wert enthält.

10. Klicken Sie auf **Weiter**.
11. Überprüfen Sie die Spalteneigenschaften.

    Optional können Sie ein Spalte für beschreibende Metadaten aus niedriger Ebene erstellen.

12. Klicken Sie auf **Weiter**.
13. Überprüfen Sie Referenztabellennamen und -beschreibung.

    Optional können Sie eine Audit-Bemerkung eingeben.

14. Wählen Sie einen Modellrepository-Speicherort für das Referenztabellenobjekt aus.
15. Klicken Sie auf **Fertig stellen**.

Erstellen einer Referenztabelle aus Wertmustern

Sie können eine Referenztabelle aus den Spaltenmustern in einer Profilspalte erstellen. Die Muster stellen die Zusammensetzung der Datenwerte in mindestens einem Spaltenfeld dar. Wählen Sie eine Spalte im Profil und wählen Sie die Muster aus, die Sie zur erstellten Referenztabelle hinzufügen möchten.

1. Öffnen Sie den Arbeitsbereich **Bibliothek** im Analyst Tool.
2. Wählen Sie die Objektkategorie **Profile** aus.

Die Inhaltsbibliothek zeigt eine Liste der Profile im Modellrepository an.

3. Öffnen Sie das Profil, das die Wertmuster enthält, die zur Referenztabelle hinzugefügt werden soll.

   In der Profilübersicht werden die Namen der Profilspalten aufgelistet.

4. Wählen Sie die Spalte aus, die die Musterdaten definiert, die zur Referenztabelle hinzugefügt werden sollen.

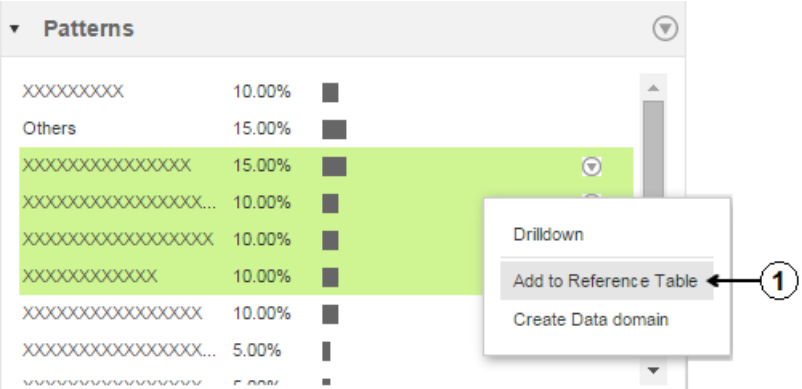
5. Überprüfen Sie die Spaltendatenmuster.

   Klicken Sie auf den Namen der Spalte, um die Spaltendaten anzuzeigen.

6. Wählen Sie in der detaillierten Profilansicht die Spaltenmuster aus, die Sie hinzufügen möchten.

7. Klicken Sie mit der rechten Maustaste auf die ausgewählten Muster und wählen Sie **Hinzufügen zu Referenztabelle** aus.

   Die folgende Abbildung zeigt ein Datenmuster für eine Spalte in der detaillierten Profilansicht:



   Die Nummer 1 gibt die Option **Hinzufügen zu Referenztabelle** in der Abbildung an.

8. Der Assistent **Hinzufügen zu Referenztabelle** wird geöffnet.

   Wählen Sie die Option **Erstellen einer Referenztabelle**.

   Hinweis: Sie können auch eine Option auswählen, um die Daten zu einer aktuellen Referenztabelle hinzuzufügen.

9. Klicken Sie auf **Weiter**.

   Der Spaltenname wird standardmäßig als Referenztabellenname angezeigt. Optional können Sie den Namen aktualisieren.

10. Geben Sie optional eine Beschreibung und einen Standardwert ein.

    Das Analyst Tool verwendet den Standardwert für jeden Tabellendatensatz, der keinen Wert enthält.

11. Klicken Sie auf **Weiter**.

12. Überprüfen Sie die Spalteneigenschaften.

    Optional können Sie ein Spalte für beschreibende Metadaten aus niedriger Ebene erstellen.

13. Klicken Sie auf **Weiter**.

14. Überprüfen Sie Referenztabellennamen und -beschreibung.

    Optional können Sie eine Audit-Bemerkung eingeben.

15. Wählen Sie einen Modellrepository-Speicherort für das Referenztabellenobjekt aus.

16. Klicken Sie auf **Fertig stellen**.

Erstellen einer Referenztabelle aus einer Einfachdatei

Sie können Referenzdaten aus einer CSV -Datei importieren. Verwenden Sie die **Neue Referenztabelle**, um die Dateidaten zu importieren.

Sie müssen die Eigenschaften für jede verwendete Einfachdatei konfigurieren, um eine Referenztabelle zu erstellen.

Eigenschaften der Analysetool-Einfachdatei

Beim Importieren einer Einfachdatei als Referenztabelle müssen Sie die Eigenschaften für jede Spalte in der Datei konfigurieren. Mit den Optionen, die Sie konfigurieren können, legen Sie fest, wie das Analysetool die Daten aus der Datei liest.

In der folgenden Tabelle werden die Eigenschaften erläutert, die Sie beim Importieren von Daten für eine Referenztabelle konfigurieren können:

Eigenschaften	Beschreibung
Delimiter	Verwendetes Zeichen, um Datenspalten zu trennen. Verwenden Sie das Feld "Sonstige", um einen anderen Delimiter einzugeben. Delimiter müssen druckfähige Zeichen sein und sich vom Escape-Zeichen und ggf. gewählten Anführungszeichen unterscheiden. Sie können keine nichtdruckenden Multibyte-Zeichen als Delimiter verwenden.
Textqualifikator	Anführungszeichen, das die Grenzen von Textstrings definiert. Wählen Sie "Keine Anführungszeichen", "Einfache Anführungszeichen" oder "Doppelte Anführungszeichen". Wenn Sie ein Anführungszeichen auswählen, ignoriert der Assistent die Delimiter innerhalb von Anführungspaaren.
Spaltennamen	Importiert Spaltennamen aus der ersten Zeile. Wählen Sie diese Option aus, wenn die Spaltennamen in der ersten Zeile angezeigt werden sollen. Der Assistent verwendet Daten in der ersten Zeile in der Vorschau für Spaltennamen. Standardeinstellung ist nicht aktiviert.
Werte	Option, den Werteimport von einer Zeile aus zu starten. Zeigt die Zeilennummer in der Vorschau an, in der der Assistent beim Importieren der Datei anfängt zu lesen.

Erstellen einer Referenztabelle aus einer Einfachdatei

Wenn Sie eine Referenztabelle aus einer Einfachdatei erstellen, verwendet die Tabelle die Spaltenstruktur der Datei und importiert die Dateidaten.

1. Klicken Sie auf **Neu** > **Referenztabelle**.

   Der Assistent **Neue Referenztabelle** wird angezeigt.
2. Wählen Sie die Option **Einfachdatei importieren**.
3. Klicken Sie auf **Weiter**.
4. Klicken Sie auf **Datei auswählen**, um die Einfachdatei auszuwählen.
5. Wählen Sie eine Codeseite, die zu den Daten in der Einfachdatei passt.
6. Klicken Sie auf **Hochladen**, um die Datendatei hochzuladen.

7. Klicken Sie auf **Weiter**.
8. Konfigurieren Sie die Eigenschaften der Einfachdatei.

   Die Eigenschaften identifizieren den von der Datei verwendeten Delimiter und geben an, ob die erste Zeile der Datei Spaltennamen enthält.
9. Um eine Vorschau der von Ihnen konfigurierten Eigenschaften zu erhalten, aktualisieren Sie den Bereich **Vorschau**.
10. Klicken Sie auf **Weiter**.
11. Konfigurieren Sie die Eigenschaften für jede Spalte.

    Die Eigenschaften umfassen Spaltenname, Datentyp, Gesamtstellenanzahl und Größenordnung.

    Wenn die Spalte Daten enthält, die eine Umwandlung bei einer Referenzdatensuche zurückgeben kann, wählen Sie die Option „Gültig“ aus.
12. Optional können Sie eine Spalte zum Einbeziehen von Beschreibungen auf niedriger Ebene als Metadaten in der Referenztabelle hinzufügen.
13. Optional können Sie eine Audit-Bemerkung für die Tabelle eingeben.

    Die Audit-Bemerkung wird im Audit-Trail-Protokoll angezeigt.
14. Klicken Sie auf **Weiter**.
15. Geben Sie einen Namen für die Referenztabelle ein und wählen Sie einen Speicherort für das Referenztabellenobjekt im Modellrepository aus.
16. Geben Sie optional eine Beschreibung der Tabelle ein.
17. Klicken Sie auf **Fertig stellen**.

Erstellen einer Referenztabelle aus einer Datenbanktabelle

Beim Erstellen einer Referenztabelle aus einer Datenbanktabelle erstellen Sie ein Metadatenobjekt im Modellrepository. Sie können die Tabellendaten auch in das Referenzdaten-Warehouse importieren.

Wenn Sie eine verwaltete Referenztabelle erstellen, importieren Sie die Spaltendaten in das Referenzdaten-Warehouse. Beim Erstellen einer nicht verwalteten Referenztabelle geben Sie die Datenbanktabelle an, die die Spaltendaten speichert. Sie können eine verwaltete Referenztabelle aus einer ODBC-Verbindung oder einer nativen Verbindung erstellen. Sie können eine nicht verwaltete Referenztabelle aus einer nativen Verbindung erstellen.

Bevor Sie die Referenztabelle erstellen, stellen Sie sicher, dass die Informatica-Domäne eine Verbindung zur Datenbank aufweist, die die Referenzdaten enthält. Wenn die Domäne noch keine Verbindung zur Datenbank enthält, können Sie sie im Analyst-Tool definieren.

Um eine Datenbankverbindung zu erstellen, klicken Sie auf **Verwalten** > **Verbindungen**.

Erstellen einer Referenztabelle aus einer Datenbanktabelle

Um die Referenztabelle zu erstellen, stellen Sie eine Verbindung zu einer Datenbank her und wählen Sie die Tabelle aus, die die Referenzdaten enthält.

1. Wählen Sie **Neu** > **Referenztabelle**.

   Der Assistent **Neue Referenztabelle** wird angezeigt.

2. Wählen Sie die Option **Verbindung zu einer relationalen Tabelle herstellen**.

   Zum Erstellen einer Referenztabelle, die keine Daten im Referenzdaten-Warehouse speichert, wählen Sie **Nicht verwaltete Tabelle** aus.

   Um Benutzer zum Bearbeiten einer nicht verwalteten Referenztabelle zu aktivieren, wählen Sie die Option **Editierbar** aus.

   Klicken Sie auf **Weiter**.
3. Wählen Sie die Datenbankverbindung aus der Liste der Verbindungen aus.

   Klicken Sie auf **Weiter**.
4. Wählen Sie im Bereich **Tabellen** eine Tabelle.
5. Überprüfen Sie die Tabelleneigenschaften im Bereich **Eigenschaften**.

   Optional können Sie durch einen Klick auf **Datenvorschau** die Tabellendaten anzeigen.

   Klicken Sie auf **Weiter**.
6. Wählen Sie im Bereich **Spaltenattribute** die gültige Spalte.

   Wenn Sie eine verwaltete Referenztabelle erstellen, können Sie die folgenden Aktionen im Bereich **Spaltenattribute** durchführen:

   - Bearbeiten der Spaltennamen der Referenztabelle
   - Hinzufügen einer Metadatenspalte für Zeilenbeschreibungen
7. Optional können Sie eine Spalte zum Einbeziehen von Beschreibungen auf niedriger Ebene als Metadaten in der Referenztabelle hinzufügen.
8. Optional können Sie eine Audit-Bemerkung für die Tabelle eingeben.

   Die Audit-Bemerkung wird im Audit-Trail-Protokoll angezeigt.
9. Klicken Sie auf **Weiter**.
10. Geben Sie einen Namen für die Referenztabelle ein und wählen Sie einen Speicherort für das Referenztabellenobjekt im Modellrepository aus.
11. Geben Sie optional eine Beschreibung für die Referenztabelle ein.
12. Klicken Sie auf **Fertig stellen**.

Arbeiten mit Referenztabellen in einem versionierten Modellrepository

Eine Referenztabelle wird im schreibgeschützten Modus geöffnet. Sie müssen in den Bearbeitungsmodus wechseln oder die Referenztabelle aus dem Modellrepository auschecken, um an ihr arbeiten zu können.

1. Klicken Sie auf der Informatica-Symbolleiste auf **Öffnen**.

   Die Objektbibliothek wird geöffnet.
2. Wählen Sie die Objektkategorie **Referenztabellen** und wählen Sie einen Referenztabellennamen aus.

   Die Referenztabelle wird im schreibgeschützten Modus geöffnet.
3. Klicken Sie zum Bearbeiten der aktuellen Version der Referenztabelle auf **Bearbeiten**.

   Checken Sie die Referenztabelle aus, um sie in einem versionierten Modellrepository zu bearbeiten.

4. Wenn Sie die Arbeit an einer Referenztabelle abgeschlossen haben, klicken Sie auf **Fertig stellen**. Das Analyst Tool speichert Ihre Änderungen an der Referenztabelle.

   Wenn Sie die Referenztabelle aus einem versionierten Modellrepository ausgecheckt haben, checken Sie das Objekt ein. Ein versioniertes Modellrepository aktualisiert die Version der Referenztabelle erst dann, wenn Sie das Objekt eingecheckt haben.

Referenztabellenaktualisierung

Die in einer Referenztabelle enthaltenen Geschäftsdaten können sich im Laufe der Zeit ändern. Überprüfen und aktualisieren Sie die Daten und Metadaten in Referenztabellen und stellen Sie sicher, dass die Tabellen korrekte Daten enthalten. Referenztabellen werden im Analyst-Tool aktualisiert. Sie können die Daten und Metadaten sowohl von verwalteten als auch von nicht verwalteten Referenztabellen aktualisieren.

Für die Daten und Metadaten in Referenztabellen können Sie die folgenden Operationen durchführen:

Verwalten von Spalten

Sie können Spalten hinzufügen, löschen und Spalteneigenschaften bearbeiten.

Verwalten von Zeilen

Sie können Datenzeilen zu einer Referenztabelle hinzufügen.

Bearbeiten von Referenzdatenwerten

Sie können einen Referenzdatenwert bearbeiten.

Ersetzen von Datenwerten

Nutzen Sie die Option **Suchen und ersetzen**, um Datenwerte zu ersetzen, die inzwischen inkorrekt oder irrelevant für das Unternehmen geworden sind. Sie können einen Wert in einer Spalte suchen und ihn mit einem anderen Wert ersetzen. Sie können auch alle Werte in einer Spalte durch einen einzigen Wert ersetzen.

Exportieren einer Referenztabelle

Eine Referenztabelle können Sie entweder als Datei mit kommagetrennten Werten (CSV), als Wörterbuchdatei oder als Excel-Datei exportieren.

Aktivieren oder Deaktivieren der Editierbarkeit einer nicht verwalteten Tabelle

Sie können eine nicht verwaltete Referenztabelle aktualisieren, um Änderungen an Tabellendaten und Metadaten zuzulassen oder zu verhindern.

Aktualisieren Sie die Referenztabellendaten

Laden Sie die Referenztabellendaten erneut in das Analyst-Tool zum Anzeigen der letzten Änderungen an den Daten.

Verwalten von Spalten

Sie können Spalten zu einer Referenztabelle hinzufügen und die Spalteneigenschaften aktualisieren. Sie können auch die Editierbarkeit einer nicht verwalteten Referenztabelle aktualisieren.

1. Klicken Sie auf **Öffnen**.

   Die Objektbibliothek wird geöffnet.

2. Wählen Sie die Objektkategorie **Referenztabellen** und wählen Sie einen Referenztabellennamen aus.

Die Referenztabelle wird im schreibgeschützten Modus geöffnet.

3. Klicken Sie zum Bearbeiten der aktuellen Version der Referenztabelle auf **Bearbeiten**.

   Checken Sie die Referenztabelle aus, um sie in einem versionierten Modellrepository zu bearbeiten.

4. Öffnen Sie das Menü **Aktionen** und wählen Sie **Spalteneigenschaften ändern** aus.

   Das Dialogfeld **Spalteneigenschaften ändern** wird geöffnet. Mithilfe der Dialogfeldoptionen können Sie die folgenden Operationen durchführen:

   - Hinzufügen einer Spalte.
   - Ändern der gültigen Spalte in der Tabelle
   - Ändern eines Spaltennamens
   - Aktualisieren einer Spaltenbeschreibung
   - Aktualisieren der Editierbarkeit einer nicht verwalteten Referenztabelle
   - Aktualisieren der Audit-Bemerkung der Tabelle

5. Wenn Sie die gewünschten Operationen durchgeführt haben, klicken Sie auf **OK**.

Verwalten von Zeilen

Sie können Zeilen in einer Referenztabelle hinzufügen, bearbeiten oder löschen.

1. Klicken Sie auf **Öffnen**.

   Die Objektbibliothek wird geöffnet.

2. Wählen Sie die Objektkategorie **Referenztabellen** und wählen Sie einen Referenztabellennamen aus.

   Die Referenztabelle wird im schreibgeschützten Modus geöffnet.

3. Klicken Sie zum Bearbeiten der aktuellen Version der Referenztabelle auf **Bearbeiten**.

   Checken Sie die Referenztabelle aus, um sie in einem versionierten Modellrepository zu bearbeiten.

4. Sie können die Datenzeilen nun wie folgt bearbeiten:

   - Um eine Zeile hinzuzufügen, wählen Sie **Aktionen** > **Zeile hinzufügen**.

     Geben Sie im Dialogfeld **Zeile hinzufügen** einen Wert in die gültige Spalte und in mindestens eine weitere Spalte ein. Optional können Sie eine Audit-Bemerkung eingeben.

     Klicken Sie auf **OK**, um die Zeile hinzuzufügen.

   - Klicken Sie zum Aktualisieren eines einzelnen Datenwerts auf den Wert und aktualisieren Sie die Daten.

     Mithilfe der Zeilenoptionen können Sie die Daten nach dem Aktualisieren übernehmen oder verwerfen. Sie können keine Audit-Bemerkung eingeben, wenn Sie Daten direkt in die Datenzeile eingeben.

   - Wählen Sie zur Aktualisierung der Datenwerte in einer Zeile **Aktionen** > **Zeile bearbeiten**.

     Geben Sie im Dialogfeld **Zeile bearbeiten** einen Wert in mindestens eine Spalte ein. Optional können Sie eine Audit-Bemerkung eingeben.

     Klicken Sie auf **Anwenden**, um die Daten in den ausgewählten Spalten zu aktualisieren.

   - Wählen Sie zur Aktualisierung der Werte in mehreren Zeilen die zu bearbeitenden Zeilen und wählen Sie **Aktionen** > **Zeile bearbeiten** aus.

     Geben Sie im Dialogfeld **Mehrere Zeilen bearbeiten** einen Wert in mindestens eine Spalte ein. Optional können Sie eine Audit-Bemerkung eingeben.

     Klicken Sie auf **OK**, um die Daten in den ausgewählten Spalten zu aktualisieren.

- Um Zeilen zu löschen, wählen Sie die zu löschenden Zeilen und wählen Sie dann **Aktionen** > **Löschen**.

  Im Dialogfeld **Zeilen löschen** können Sie optional eine Audit-Bemerkung eingeben.

  Klicken Sie auf **OK**, um die Zeilen zu löschen.

Hinweis: Nutzen Sie das Developer Tool, um Zeilendaten in großen Referenztabellen zu bearbeiten. Wenn eine Referenztabelle beispielsweise über 500 Zeilen umfasst, sollten Sie diese im Developer Tool bearbeiten.

Werte suchen und ersetzen

Sie können Datenwerte in einer Referenztabelle suchen und ersetzen. Nutzen Sie die Such- und Ersetzfunktion, wenn eine Tabelle eine oder mehrere Instanzen eines zu aktualisierenden Datenwerts enthält.

1. Klicken Sie auf **Öffnen**.

   Die Objektbibliothek wird geöffnet.
2. Wählen Sie die Objektkategorie **Referenztabellen** und wählen Sie einen Referenztabellennamen aus.

   Die Referenztabelle wird im schreibgeschützten Modus geöffnet.
3. Klicken Sie zum Bearbeiten der aktuellen Version der Referenztabelle auf **Bearbeiten**.

   Checken Sie die Referenztabelle aus, um sie in einem versionierten Modellrepository zu bearbeiten.
4. Klicken Sie auf **Aktionen** > **Suchen und ersetzen**.

   Die Symbolleiste **Suchen und ersetzen** erscheint.
5. Geben Sie die Suchkriterien über die Symbolleiste ein:
   - Geben Sie einen Datenwert im Feld **Suchen** ein.
   - Wählen Sie die zu durchsuchenden Spalten aus. Standardmäßig durchsucht der Vorgang alle Spalten.
   - Geben Sie einen Datenwert im Feld **Ersetzen durch** ein.
6. Nutzen Sie die folgenden Optionen, um einzelne Werte oder alle Werte zu ersetzen:
   - Nutzen Sie die Optionen **Nächster** und **Vorheriger**, um Werte einzeln nacheinander zu suchen.
   - Um einen Wert zu ersetzen, wählen Sie **Ersetzen**.
   - Um alle Instanzen des Werts zu anzuzeigen, wählen Sie **Alle markieren**.
   - Um alle Instanzen des Werts zu ersetzen, wählen Sie **Alle ersetzen**.

Exportieren von Referenztabellendaten

Die Daten in einer Referenztabelle können Sie entweder als CSV-Datei, als Wörterbuchdatei oder als Microsoft-Excel-Datei exportieren. Sie können die Daten im schreibgeschützten Modus exportieren.

1. Klicken Sie auf **Öffnen**.

   Die Objektbibliothek wird geöffnet.
2. Wählen Sie die Objektkategorie **Referenztabellen** und wählen Sie einen Referenztabellennamen aus.

   Die Referenztabelle wird im schreibgeschützten Modus geöffnet.
3. Klicken Sie auf **Aktionen** > **Daten exportieren.**

   Das Dialogfenster **Daten an Datei exportieren** wird geöffnet.

In der folgenden Tabelle werden die Optionen für das Dialogfeld beschrieben:

Option	Beschreibung
Dateiname	Name der Datei, die die Daten enthalten soll. Die Datei wird beim Exportvorgang erstellt.
Dateiformat	Format der Datei, die die Daten enthalten soll. Wählen Sie eines der folgenden Formate aus: • csv. Kommagetrennte Datei. Standard-Format. • xls. Microsoft Excel-Datei. • dic. Informatica-Wörterbuchdatei.
Feldnamen als erste Zeile exportieren	Spaltennamenoption. Wählen Sie diese Option, wenn die erste Zeile der Datei die Spaltennamen enthält.
Codeseite	Codeseite der Referenzdaten. Die Standardcodeseite ist UTF-8.

4. Klicken Sie auf **OK**, um die Datei zu exportieren.

Aktivieren und Deaktivieren der Editierbarkeit in nicht verwalteten Referenztabellen

Sie können Aktualisierungen von Datenwerten und Spalten in einer nicht verwalteten Referenztabelle aktivieren oder deaktivieren.

Speichern Sie die Tabelle, bevor Sie ihre Editierbarkeit ändern.

1. Klicken Sie auf **Öffnen**.

   Die Objektbibliothek wird geöffnet.
2. Wählen Sie die Objektkategorie **Referenztabellen** und wählen Sie einen Referenztabellennamen aus.

   Die Referenztabelle wird im schreibgeschützten Modus geöffnet.
3. Klicken Sie zum Bearbeiten der aktuellen Version der Referenztabelle auf **Bearbeiten**.

   Checken Sie die Referenztabelle aus, um sie in einem versionierten Modellrepository zu bearbeiten.
4. Öffnen Sie das Menü **Aktionen** und wählen Sie **Spalteneigenschaften ändern** aus.

   Das Dialogfeld **Spalteneigenschaften ändern** wird geöffnet.
5. Aktivieren oder deaktivieren Sie die Option **Editierbar**.

Hinweis: Die folgenden Bedingungen gelten für eine nicht verwaltete Referenztabelle, die Aktualisierungen von Benutzern zulässt:

- Die Referenztabelle muss einfache Datentypen wie „string“ und „number“ verwenden.
- Definieren Sie keine Einschränkung für die Metadaten von Referenztabellen oder geben Sie einen Standardwert für eine Spalte an.

Aktualisieren Sie die Referenztabellenwerte

Möglicherweise müssen Sie die im Analyst Tool für die Referenztabelle angezeigten Werte aktualisieren.

Zum erneuten Laden der Referenztabellenwerte klicken Sie auf **Aktionen** > **Aktualisieren**. Das Analyst Tool ruft die aktuellen Versionen der Datenwerte aus der Datenbank ab.

Audit-Trail-Ereignisse

Sie können einen Audit-Trail der Änderungen anzeigen, die Benutzer an einer Referenztabelle vornehmen. Verwenden Sie die Audit-Trail-Ansicht in der Referenztabelle, um die Audit-Trail-Ereignisse anzuzeigen. Sie können die im Analyst Tool angezeigten Audit-Trail-Ereignisse filtern.

In der folgenden Tabelle werden die Filteroptionen beschrieben, die Sie festlegen können:

Option	Beschreibung
Datum	Start- und Enddatum für die anzuzeigenden Aktionen. Verwenden Sie die Kalenderoptionen, um die Zeitpunkte festzulegen.
Typ	Typ des Audit-Trail-Ereignisses. Sie können die folgenden Ereignistypen anzeigen: - Daten. Ereignisse, die sich auf die Datenwerte in der Referenztabelle beziehen. Protokollereignisse beinhalten Vorgänge zum Hinzufügen, Löschen und Aktualisieren einer Zeile. - Metadaten. Ereignisse, die sich auf die Referenztabellen-Metadaten beziehen. Ereignisse beinhalten Vorgänge zum Erstellen der Referenztabelle, zum Hinzufügen oder Löschen einer Spalte oder zum Einchecken der Referenztabelle. **Hinweis:** Sie können Daten- und Metadatenereignisse nicht gleichzeitig anzeigen.
Benutzer	Benutzer, der die Referenztabelle bearbeitet hat. Der Filter zeigt den vollständigen Namen und den Anmeldenamen des Benutzers an.
Status	Status der Audit-Trail-Protokollereignisse. Der Status entspricht der Aktion, die Sie im Referenztabellen-Editor ausgeführt haben. Der Status kann beispielsweise anzeigen, dass ein Benutzer die Referenztabelle erstellt oder eine Zeile hinzugefügt hat.

Audit-Trail-Protokollereignisse beinhalten außerdem die Audit-Trail-Bemerkungen und die Spaltenwerte, die Sie eingefügt, aktualisiert oder gelöscht haben.

Audit-Trail-Ereignisse anzeigen

Zeigen Sie Audit-Trail-Ereignisse an, um Details über die Aktualisierungen zu erhalten, die Benutzer an einer Referenztabelle vorgenommen haben. Sie können die Audit-Trail-Ereignisse im schreibgeschützten Modus anzeigen.

1. Klicken Sie auf **Öffnen**.

   Die Objektbibliothek wird geöffnet.
2. Wählen Sie die Objektkategorie **Referenztabellen** und wählen Sie einen Referenztabellennamen aus.

   Die Referenztabelle wird im schreibgeschützten Modus geöffnet.
3. Klicken Sie auf **Audit-Trail**.
4. Konfigurieren Sie die Filteroptionen.

   Sie können nach dem Datum der Aktualisierung, dem Aktualisierungstyp, dem Aktualisierungsstatus und dem Namen des Benutzers filtern, der die Aktualisierung durchgeführt hat.
5. Klicken Sie auf **Anzeigen**.

   Die Protokollereignisse werden für die von Ihnen angegebenen Filteroptionen angezeigt.

Regeln und Richtlinien für Referenztabellen

Verwenden Sie die folgenden Regeln und Richtlinien beim Arbeiten mit Referenztabellen im Analyst Tool:

- Wenn Sie eine Referenztabelle aus einer Datenbank von Oracle, IBM DB2 oder Microsoft SQL Server importieren, kann das Analyst Tool die Vorschau nicht anzeigen, wenn die Tabellen-, Ansichts-, Schema-, Synonym- oder Spaltennamen gemischte Groß- und Kleinschreibung oder Kleinbuchstaben enthalten.

  Um eine Vorschau von Tabellendaten anzuzeigen, die in Datenbanken mit Erkennung von Groß-/Kleinschreibung enthalten sind, legen Sie das Attribut „Bezeichner mit gemischter Groß-/Kleinschreibung unterstützen“ auf TRUE fest.

- Wenn Sie eine Referenztabelle aus abgeleiteten Spaltenmustern in einem Format erstellen, füllt das Analyst Tool die Referenztabelle mit Spaltenmustern in einem anderen Format.

  Wenn Sie zum Beispiel eine Referenztabelle für das Spaltenmuster X(5) erstellen, zeigt das Analyst Tool das folgende Format für das Spaltenmuster in der Referenztabelle an: `XXXXX`.

- Überprüfen Sie beim Importieren einer Oracle-Datenbanktabelle die Länge der VARCHAR2-Spalten in der Tabelle. Das Analyst Tool kann keine Oracle-Datenbanktabelle importieren, die eine VARCHAR2-Spalte enthält, die länger als 1000 ist.

- Um eine Referenztabelle zu lesen, brauchen Sie Ausführungsberechtigungen für die Verbindung mit der Datenbank, in der die Tabellendatenwerte gespeichert sind. Wenn beispielsweise das Referenz-Data-Warehouse die Datenwerte speichert, benötigen Sie für die Verbindung zum Referenz-Data-Warehouse Ausführungsberechtigungen. Sie benötigen Ausführungsberechtigungen, um auf die Referenztabelle im Lese- oder Schreibmodus zuzugreifen. Die Berechtigungen für die Verbindung mit der Datenbank gelten für alle Referenzdaten in der Datenbank.

- Bei Ausführung eines Mappings mit einer Umwandlung, bei dem eine Referenztabelle angegeben ist, wird die aktuelle Version der Referenztabelle im Modellrepository verwendet. Sie können keine frühere Version der Referenztabelle auswählen, wenn Sie die Umwandlung konfigurieren.

  Wenn ein anderer Benutzer eine frühere Version der Referenztabelle in einer gleichzeitigen Developer Tool-Sitzung wiederherstellt, sind die Versionen der Referenztabelle in den Sitzungen nicht mehr identisch. Wenn Sie ein Mapping konfigurieren und ausführen, das eine Referenztabelle verwendet, könnte das Mapping fehlschlagen, weil die aktuelle Sitzung nicht die aktuelle Version der Referenztabelle erkennt. Um sicherzustellen, dass das Mapping die aktuelle Referenztabelle verwendet, aktualisieren Sie das Modellrepository, bevor Sie das Mapping ausführen.

- Wenn Sie eine nicht verwaltete Referenztabelle zum Zulassen von Bearbeitungen konfigurieren, stellen Sie sicher, dass in der Referenztabelle einfache Datentypen wie „string“ und „number“ verwendet werden. Stellen Sie außerdem sicher, dass die Referenztabelle weder eine Einschränkung für die Metadaten der Referenztabelle definiert noch Standardwerte für Spalten verwendet.

KAPITEL 3

Referenzdaten im Developer-Tool

Dieses Kapitel umfasst die folgenden Themen:

- Developer-Tool-Referenzdaten – Übersicht, 30
- Referenzdaten und Umwandlungen, 31
- Arbeiten mit Referenzdatenobjekten in einem versionierten Modellrepository, 31
- Referenztabellen, 32
- Inhaltssätze, 37

Developer-Tool-Referenzdaten – Übersicht

Sie können die Konfigurationseigenschaften für Referenzdatenobjekte im Developer-Tool erstellen, aktualisieren und anzeigen.

Verwenden Sie das Developer-Tool, um die folgenden Objekttypen zu erstellen und zu aktualisieren:

Referenztabellen

Eine Referenztabelle enthält die Standardversion und alternative Versionen einer Gruppe von Datenwerten. Im Developer-Tool fügen Sie eine Referenztabelle zu einer Umwandlung hinzu, um sicherzustellen, dass Quelldatenwerte genau und richtig formatiert sind.

Inhaltssätze

Ein Inhaltssatz ist ein Modellrepository-Objekt, das Referenzdatenwerte im Repository oder in einer Datei angibt. Ein Inhaltssatz enthält verschiedene Referenzdatentypen, die Sie zum Durchführen von Suchläufen in Datenqualitätsumwandlungen verwenden können.

Darüber hinaus können Sie im Developer-Tool auch mit Adressreferenzdaten- und Identitätspopulationsdateien arbeiten. Sie wählen Adressreferenzdatendateien aus, wenn Sie eine Adress-Validiererumwandlung konfigurieren. Sie wählen Identitätspopulationsdateien aus, wenn Sie eine Match-Umwandlung für Identitätsvergleichsanalysen konfigurieren.

Referenzdaten und Umwandlungen

Mehrere Umwandlungen lesen Referenzdaten, um Datenqualitätsaufgaben durchzuführen.

Die folgenden Umwandlungen können Referenzdaten lesen:

- Adressvalidator Liest Adressreferenzdaten, um die Genauigkeit von Adressen zu überprüfen.
- Fallkonverter Liest Referenzdatentabellen, um Strings zu finden, bei denen die Groß-/Kleinschreibung geändert werden muss.
- Klassifizierer. Liest Inhaltssatzdaten, um den Typ der Informationen in einer Zeichenfolge zu identifizieren.
- Vergleich Liest Identitätspopulationsdaten während der Doppelanalyse.
- Labeler Liest Inhaltssatzdaten, um Zeichenfolgen zu identifizieren und zu beschriften.
- Match Liest Identitätspopulationsdaten während der Doppelanalyse.
- Parser Liest Inhaltssatzdaten, um Zeichenfolgen basierend auf den enthaltenen Informationen zu analysieren.
- Standardisierer Liest Referenzdatentabellen zum Standardisieren von Strings in ein gemeinsames Format.

Der Data Quality Content Installer-Dateisatz enthält Informatica-Referenzdatenobjekte, die Sie importieren können.

Arbeiten mit Referenzdatenobjekten in einem versionierten Modellrepository

Wenn Sie mit Referenztabellen oder Inhaltssätzen in einem versionierten Modellrepository arbeiten, wendet das Repository möglicherweise die Versionskontrolle auf die Objekte an. Um die Versionskontrolle auf ein Objekte anwenden zu können, checkt der Benutzer das Objekt in das Modellrepository ein.

Wenn die Versionskontrolle für eine Referenztabelle oder einen Inhaltssatz nicht aktiviert ist, können Sie das Objekt außerhalb des Versionsverwaltungssystems öffnen und aktualisieren. Wenn Sie das Objekt öffnen, wird es vom Modellrepository gesperrt, sodass kein anderer Benutzer daran arbeiten kann.

Wenn die Versionskontrolle für eine Referenztabelle oder einen Inhaltssatz aktiviert ist, wird das Objekt im schreibgeschützten Modus geöffnet. Checken Sie das Objekt aus dem Modellrepository aus, um daran zu arbeiten. Alternativ checken Sie das Objekt aus und öffnen es anschließend. Checken Sie das Objekt ein, um eine Version des Objekts zu erstellen, in dem Ihre letzten Änderungen enthalten sind.

Auschecken von Referenzdatenobjekten

Wenn Sie an einer Referenztabelle oder einem Inhaltssatz arbeiten möchten, die/der ein Benutzer in das Modellrepository eingecheckt hat, checken Sie das Objekt aus dem Repository aus.

1. Navigieren Sie im Objekt-Explorer zu einer Referenztabelle oder einem Inhaltssatz.
2. Klicken Sie mit der rechten Maustaste auf den Namen des Objekts und klicken Sie auf **Öffnen**.

   Das Objekt wird im schreibgeschützten Modus geöffnet.
3. Klicken Sie mit der rechten Maustaste auf den Namen des Objekts und klicken Sie auf **Auschecken**.

   Sie können das Objekt bearbeiten.

Einchecken von Referenzdatenobjekten

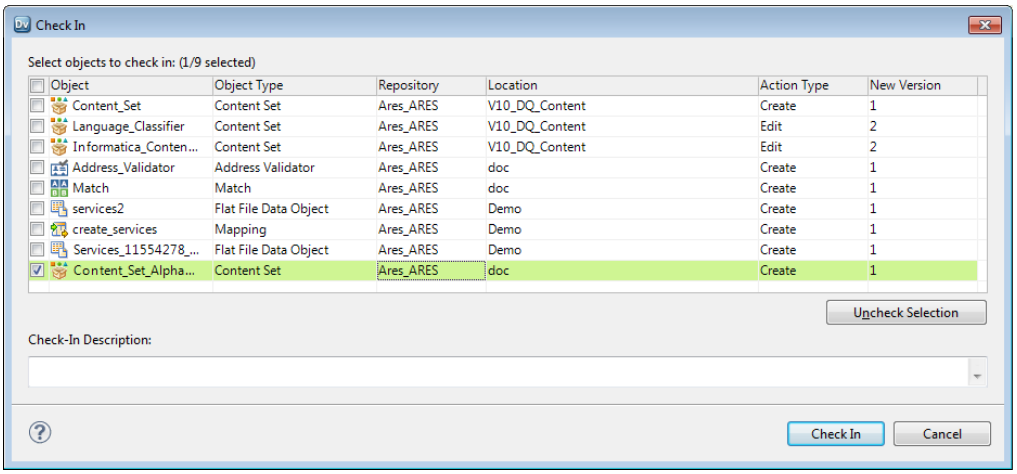
Checken Sie das Objekt ein, wenn Sie eine Referenztabelle oder einen Inhaltssatz, die/der aus dem Modellrepository ausgecheckt wurde, fertig bearbeitet haben.

Öffnen Sie zum Anzeigen der Liste mit den aktuell ausgecheckten Objekten die Registerkarte **Ausgecheckte Objekte** unter dem Referenztabellen-Editor.

1. Speichern Sie alle Änderungen, die Sie an der Referenztabelle oder dem Inhaltssatz vorgenommen haben.
2. Navigieren Sie im Objekt-Explorer zur Referenztabelle oder dem Inhaltssatz.
3. Klicken Sie mit der rechten Maustaste auf den Namen des Objekts und klicken Sie auf **Einchecken**.

   Das Dialogfeld **Einchecken** wird geöffnet.

   Die folgende Abbildung zeigt das Dialogfeld:



4. Wählen Sie mindestens ein Objekt aus, die in das Repository eingecheckt werden soll.

   Hinweis: Sie können ein Objekt einchecken, das in der aktuellen Sitzung nicht geöffnet ist. Sie können jedes Objekt einchecken, das sich im ausgecheckten Status befindet.
5. Optional können Sie eine Beschreibung für den Vorgang eingeben.
6. Klicken Sie auf **Einchecken**.

   Beim Eincheck-Vorgang wird die Versionsnummer des Objekts aktualisiert. Wenn Sie das Objekt zum ersten Mal einchecken, erstellt das Modellrepository die Version eins (1) des Objekts.

Referenztabellen

Verwenden Sie das Developer tool, um eine Referenztabelle einer Umwandlung hinzuzufügen. Sie konfigurieren die Umwandlung, um nach Referenztabellenwerten in Eingabedaten zu suchen und die entsprechenden gültigen Werte aus der Referenztabelle als Ausgabe zu schreiben.

Zum Erstellen einer Referenztabelle im Developer tool verwenden Sie eine der folgenden Methoden:

- Erstellen Sie eine leere Referenztabelle und geben Sie die Datenwerte ein.
- Erstellen Sie eine Referenztabelle aus Daten in einer Einfachdatei.
- Erstellen Sie eine Referenztabelle aus Daten in einer Datenbanktabelle, einem Synonym oder einer Ansicht.

Eigenschaften von Referenztabellendaten

Sie können die Eigenschaften für Referenztabellendaten und Metadaten im Developer-Tool anzeigen. Das Developer-Tool zeigt die Eigenschaften beim Öffnen der Referenztabelle aus dem Modellrepository an.

Eine Referenztabelle zeigt allgemeine Eigenschaften und Spalteneigenschaften an. Sie können Referenztabelleneneigenschaften im Developer-Tool anzeigen. Sie können Referenztabelleneigenschaften im Analyst-Tool anzeigen und bearbeiten.

In der folgenden Tabelle werden die allgemeinen Eigenschaften einer Referenztabelle beschrieben:

Eigenschaft	Beschreibung
Name	Name der Referenztabelle.
Beschreibung	Optionale Beschreibung der Referenztabelle.

In der folgenden Tabelle werden die Spalteneigenschaften einer Referenztabelle beschrieben:

Eigenschaft	Beschreibung
Gültig	Identifiziert die Spalte mit den gültigen Referenzdaten.
Name	Name jeder Spalte.
Datentyp	Datentyp der Daten in jeder Spalte.
Gesamtstellenanzahl	Gesamtstellenanzahl für jede Spalte.
Größenordnung	Größenordnung für jede Spalte.
Beschreibung	Beschreibung der Spalteninhalte. Sie können eine Beschreibung hinzufügen, wenn Sie die Referenztabelle erstellen.
Einbeziehen einer Spalte für Beschreibungen auf niedriger Ebene	Gibt an, dass die Referenztabelle eine Spalte für Beschreibungen der Spaltendaten enthält.
Standardwert	Standardwert für die Felder in der Spalte. Sie können einen Standardwert hinzufügen, wenn Sie die Referenztabelle erstellen.
Verbindungsname	Name der Verbindung zur Datenbank, die die Datenwerte der Referenztabelle enthält.

Erstellen eines Referenztabellenobjekts

Wählen Sie diese Option, wenn Sie eine leere Referenztabelle erzeugen und Daten von Hand hinzufügen wollen.

1. Wählen Sie im Menü des Developer-Tools die Option **Datei** > **Neu** > **Referenztabelle**.
2. Im Eingabeassistenten für neue Tabellen wählen Sie **Referenztabelle als leer**.
3. Geben Sie einen Namen für die Tabelle ein.
4. Wählen Sie ein Projekt, in dem die Tabellenmetadaten gespeichert werden sollen.

Im Feld für den Speicherort klicken Sie auf **Durchsuchen**. Das Dialogfenster **Speicherort wählen** wird geöffnet und zeigt die Projekte im Repository an. Wählen Sie das benötigte Projekt.

Klicken Sie auf **Weiter**.

5. Fügen Sie der Tabelle mindestens zwei Spalten hinzu. Klicken Sie auf die Option **Neu**, um eine Spalte zu erstellen.

   Die folgende Tabelle beschreibt die Eigenschaften für jede Spalte:

Eigenschaft	Standardwert
Name	Spalte
Datentyp	String
Präzision	10
Größenordnung	0
Beschreibung	Leer. Optionale Eigenschaft.

6. Wählen Sie die Spalte aus, in der die gültigen Werte enthalten sind. Sie können die Reihenfolge der Spalten ändern, die Sie erstellen.
7. Die folgende Tabelle beschreibt optionale Eigenschaften:

Eigenschaft	Standardwert
Schließen Sie eine Spalte für Beschreibungen auf Zeilenebene ein.	Gelöscht
Audit-Hinweis	Leer
Standardwert	Leer

   Klicken Sie auf **Fertig stellen.**

Die Referenztabelle wird im Developer Tool-Workspace geöffnet.

Erstellen einer Referenztabelle aus einer Einfachdatei

Sie können eine Referenztabelle aus Daten in einer Einfachdatei erstellen.

1. Wählen Sie im Menü des Developer-Tools die Option **Datei** > **Neu** > **Referenztabelle**.
2. Wählen Sie im Eingabeassistenten für neue Tabellen **Referenztabelle aus einer Einfachdatei**.
3. Blättern Sie zu der Datei, die Sie als Datenquelle für die Tabelle verwenden möchten.
4. Geben Sie einen Namen für die Tabelle ein.
5. Wählen Sie ein Projekt, in dem die Tabellenmetadaten gespeichert werden sollen.

   Im Feld für den Speicherort klicken Sie auf **Durchsuchen**. Das Dialogfenster **Speicherort wählen** wird geöffnet und zeigt die Projekte im Repository an. Wählen Sie das benötigte Projekt.

   Klicken Sie auf **Weiter**.
6. Wählen Sie UTF-8 als Codepage.
7. Geben Sie das in der Einfachdatei verwendete Trennzeichen ein.

8. Wenn die Einfachdatei Spaltennamen enthält, wählen Sie die Option zum Importieren von Spaltennamen aus der ersten Zeile der Datei.
9. Die folgende Tabelle beschreibt optionale Tabelleneigenschaften:

Eigenschaft	Standardwert
Textqualifikator	Keine Anführungszeichen
Import beginnen in Zeile	Zeile 1
Zeilen-Delimiter	\012 LF (\n)
Aufeinander folgende Delimiter wie einen behandeln	Gelöscht
Escape-Zeichen	Leer
Escape-Zeichen in Daten beibehalten	Gelöscht
Max. Anzahl Zeilen in der Vorschau	500

Klicken Sie auf **Weiter.**

10. Wählen Sie die Spalte, welche die gültigen Werte enthält.
11. Die folgende Tabelle beschreibt optionale Eigenschaften:

Eigenschaft	Standardwert
Schließen Sie eine Spalte für Beschreibungen auf Zeilenebene ein.	Gelöscht
Audit-Hinweis	Leer
Standardwert	Leer
Max. Anzahl Zeilen in der Vorschau	500

Klicken Sie auf **Fertig stellen**.

Die Referenztabelle wird im Developer Tool-Workspace geöffnet.

Erstellen einer Referenztabelle aus einer relationalen Quelle

Sie können eine Referenztabelle aus einer relationalen Tabelle, einem Synonym oder einer Ansicht erstellen.

Wenn Sie eine verwaltete Referenztabelle erstellen, importieren Sie die Spaltendaten in das Referenzdaten-Warehouse. Beim Erstellen einer nicht verwalteten Referenztabelle geben Sie die Datenbanktabelle an, die die Spaltendaten speichert. Sie können eine verwaltete Referenztabelle aus einer ODBC-Verbindung oder einer nativen Verbindung erstellen. Sie können eine nicht verwaltete Referenztabelle aus einer nativen Verbindung erstellen.

Bevor Sie die Referenztabelle erstellen, stellen Sie sicher, dass die Informatica-Domäne eine Verbindung zur Datenbank aufweist, die die Referenzdaten enthält.

Sie können eine Datenbankverbindung im Verbindungs-Explorer konfigurieren. Wenn das Developer tool den Verbindungs-Explorer nicht anzeigt, wählen Sie **Fenster** > **Ansicht anzeigen** > **Verbindungs-Explorer** im Developer tool-Menü aus.

Erstellen einer Referenztabelle aus einer relationalen Quelle

Um die Referenztabelle zu erstellen, stellen Sie eine Verbindung zu einer Datenbank her und wählen Sie die Tabelle aus, die die Referenzdaten enthält.

1. Wählen Sie im Menü des Developer tool die Option **Datei** > **Neu** > **Referenztabelle**.
2. Wählen Sie im Assistenten zur Erstellung von Referenztabellen **Referenztabelle aus einer relationalen Quelle** aus.

   Klicken Sie auf **Weiter**.
3. Wählen Sie eine Datenbankverbindung.

   Im Feld für die Verbindung klicken Sie auf **Durchsuchen**. Das Dialogfeld **Verbindung wählen** öffnet sich und zeigt die verfügbaren Datenbankverbindungen an.

   Klicken Sie auf **OK**, wenn Sie eine Verbindung gewählt haben.
4. Wählen Sie eine Datenbank-Ressource aus.

   Im Feld für die Ressource klicken Sie auf **Durchsuchen**. Das Dialogfeld **Ressource wählen** öffnet sich und zeigt die Ressourcen der Datenbankverbindung an. Durchsuchen Sie die Datenbank und wählen Sie eine Datenbanktabelle, ein Synonym oder eine Ansicht aus.

   Optional können Sie eine Vorschau der Entitätsinformationen für die Ressource anzeigen.
5. Geben Sie einen Namen für die Tabelle ein.
6. Wählen Sie einen Speicherort für das Referenztabellenobjekt aus.

   Im Feld für den Speicherort klicken Sie auf **Durchsuchen**. Das Dialogfeld **Speicherort wählen** öffnet sich und zeigt die Projekte im Repository an.

   Wählen Sie einen Speicherort aus und klicken Sie auf **Weiter**.
7. Zum Erstellen einer Referenztabelle, die keine Daten im Referenzdaten-Warehouse speichert, wählen Sie **Nicht verwaltete Tabelle** aus.

   Um Benutzer zum Bearbeiten einer nicht verwalteten Referenztabelle zu aktivieren, wählen Sie die Option **Editierbar** aus.

   Klicken Sie auf **Weiter**.
8. Wählen Sie die Spalte aus, in der die gültigen Werte enthalten sind.
9. In der folgenden Tabelle werden die optionalen Eigenschaften beschrieben, die Sie festlegen können:

Eigenschaft	Standardwert
Schließen Sie eine Spalte für Beschreibungen auf Zeilenebene ein.	Gelöscht
Beschreibung	Gelöscht
Standardwert	Leer
Audit-Hinweis	Leer
Max. Anzahl Zeilen in der Vorschau	500

10. Klicken Sie auf **Fertig stellen**.

Inhaltssätze

Ein Inhaltssatz ist ein Modellrepository-Objekt, in dem Daten oder Metadaten für andere Referenzdatenobjekte gespeichert werden. Ein Inhaltssatz kann Zeichensätze, Muster- und Tokensätze, reguläre Ausdrücke, probabilistische Modelle und Klassifikatormodelle enthalten. Verwenden Sie einen Inhaltssatz zur Definition und Verwaltung von Referenzdatenobjekten, die sich auf ein einzelnes Projekt, einen einzelnen Informationstyp oder Geschäftszweck beziehen.

Das Developer-Tool enthält systemdefinierte Zeichen- und Tokensätze, die im Modellrepository nicht angezeigt werden. Konfigurieren Sie zum Anzeigen und Verwenden der systemdefinierten Objekte eine Strategie in der Beschriftungs-, Parser- oder Standardisierer-Umwandlung.

Zeichensätze

Ein Zeichensatz enthält Expressions, die bestimmte Zeichen und Zeichenbereiche kennzeichnen. Sie können Zeichensätze in Labeler-Umwandlungen verwenden, die den Zeichenbeschriftungmodus verwenden.

Zeichenbereiche geben Sie einen sequentielle Bereich von Zeichencodes an. Der Zeichenbereich "[A-C]" entspricht beispielsweise den Großbuchstaben "A," "B" und "C." Dieser Zeichenbereich stimmt nicht mit den Kleinbuchstaben "a," "b" oder "c" überein.

Verwenden Sie Zeichensätze zur Angabe eines Sonderzeichens oder von Zeichenbereichen als Teil von Beschriftungsoperationen. Sie können beispielsweise alle Nummernzeichen in einer Spalte kennzeichnen, die Telefonnummern beinhalten. Nach der Beschriftung der Zahlen können Sie Muster mit einer Parser-Umwandlung angeben und problematische Muster in separate Ausgabeports schreiben.

Zeichensatz-Eigenschaften

Konfigurieren Sie die Eigenschaften, die Zeichenbeschriftungsoperationen für einen Zeichensatz festlegen.

In der folgenden Tabelle werden die Eigenschaften für einen benutzerdefinierten Zeichensatz beschrieben:

Eigenschaft	Beschreibung
Beschriftung	Definiert die Beschriftung, die eine Labeler-Umwandlung auf Daten anwendet, die mit dem Zeichensatz übereinstimmen.
Standardmodus	Aktiviert eine einfache Bearbeitungsansicht mit Feldern für den Anfangs- und Endbereich.
Startbereich	Gibt das erste Zeichen in einem Zeichenbereich an.
Endbereich	Gibt das letzte Zeichen in einem Zeichenbereich an. Lassen Sie dieses Feld für einen Bereich mit einem einzelnen Zeichen leer.
Erweiterter Modus	Ermöglicht eine erweiterte Bearbeitungsansicht, in der Zeichenbereich unter Verwendung von Bereichszeichen und Delimiter-Zeichen manuell eingeben können.
Bereichszeichen	Temporäre Änderungen des Symbols, das einen Zeichenbereich kennzeichnet. Der Zeichenbereich wird auf das Standardzeichen zurückgesetzt, wenn Sie den Zeichensatz schließen.
Delimiter-Zeichen	Temporäre Änderungen des Symbols, das Zeichenbereiche trennt. Das Delimiter-Zeichen wird auf das Standardzeichen zurückgesetzt, wenn Sie den Zeichensatz schließen.

Klassifizierermodelle

Ein Klassifizierermodell analysiert Eingabezeichenfolgen und legt die Informationstypen fest, die aller Wahrscheinlichkeit nach in den Zeichenfolgen enthalten sind. Sie verwenden ein Klassifiziermodell in einer Klassifizierumwandlung.

Ein Klassifizierermodell enthält Referenzdatenzeilen und Beschriftungswerte. Die Zeilen stellen die Eingabedaten auf dem Port dar, den Sie unter Umständen mit der Klassifiziererumwandlung verbinden. Die Beschriftungswerte beschreiben die Informationstypen, die in den Datenzeilen enthalten sind. Wenn Sie ein Klassifizierermodell konfigurieren, weisen Sie jeder Referenzdatenzeile im Modell eine Beschriftung zu.

Sie kompilieren das Modell, um die Referenzdatenzeilen mit den Beschriftungen in einem Klassifizierermodell zu verknüpfen. Der Kompilierungsprozess erzeugt eine Reihe von logischen Verbindungen zwischen den Datenzeilen und den Beschriftungswerten. Wenn Sie eine Zuordnung ausführen, die das Modell liest, wendet der Datenintegrationsdienst die Logik des Modells auf die Eingabedaten der Klassifiziererumwandlung an. Der Datenintegrationsdienst gibt die Beschriftungen zurück, die die Informationen in den Eingabedatenfeldern am genauesten beschreiben.

Sie erstellen ein Klassifizierermodell im Developer-Tool. Das Modellrepository speichert das Klassifizierermodellobjekt. Das Developer-Tool schreibt die Datenzeilen, die Beschriftungen und die Kompilierungsdaten in eine Datei in der Informatica-Verzeichnisstruktur.

Mustersätze

Ein Mustersatz enthält Expressions, die Datensätze in der Ausgabe einer Token-Bezeichnung kennzeichnen. Sie können Mustersätze verwenden, um den Datenausgabeport zu analysieren, der mit einem Token versehen ist, und um Matching-Strings in einen oder mehrere Ausgabeports zu schreiben. Verwenden Sie Mustersätze in Parser-Umwandlungen, die den Muster-Parsing-Modus verwenden.

Sie können beispielsweise eine Parser-Umwandlung konfigurieren, um Mustersätze zu verwenden, die Namen und Anfangsbuchstaben angeben. Diese Umwandlung verwendet die Mustersätze, um die Ausgabe einer Labeler-Umwandlung im Tokenbeschriftungsmodus zu analysieren. Sie können die Parser-Umwandlung konfigurieren, um Namen und Anfangsbuchstaben in der Ausgabe an verschiedene Ports zu schreiben.

Mustersatzeigenschaften

Konfigurieren Sie die Eigenschaften, mit denen die Muster in einem Mustersatz festgelegt werden.

In der folgenden Tabelle wird die Eigenschaft für einen benutzerdefinierten Mustersatz beschrieben:

Eigenschaft	Beschreibung
Muster	Definiert die Muster, nach denen der Muster-Parser sucht. Sie können mehrere Muster für einen Mustersatz eingeben. Sie können Muster eingeben, die aus einer Kombination aus Platzhaltern, Zeichen und Strings erstellt wurden.

Probabilistische Modelle

Ein probabilistisches Modell analysiert die Eingabedatenwerte und legt die Informationstypen fest, die aller Wahrscheinlichkeit nach in den Werten enthalten sind. Verwenden Sie ein probabilistisches Modell in einer Beschriftungsumwandlung und in einer Parser-Umwandlung.

Ein probabilistisches Modell enthält Referenzdatenwerte und Beschriftungswerte. Die Referenzdatenwerte stellen die Daten auf einem Eingabeport dar, den Sie mit der Umwandlung verbinden. Die Beschriftungswerte

beschreiben die Typen von Informationen, die die Referenzdatenwerte enthalten. Sie weisen jedem Referenzdatenwert im Modell eine Beschriftung zu.

Sie kompilieren das Modell, um die Referenzdatenzeilen mit den Beschriftungen in einem probabilistischen Modell zu verknüpfen. Der Kompilierungprozess generiert eine Reihe von logischen Assoziationen zwischen den Datenwerten und den Beschriftungen. Wenn Sie ein Mapping ausführen, das das Modell liest, wendet der Datenintegrationsdienst die Logik des Modells auf die Eingabedaten der Umwandlung an. Der Datenintegrationsdienst gibt die Beschriftung zurück, die die Eingabedatenwerte am genauesten beschreibt.

Sie erstellen ein probabilistisches Modell im Developer Tool. Das Modellrepository speichert das probabilistische Modellobjekt. Das Developer-Tool schreibt die Datenwerte, die Beschriftungen und die Kompilierungsdaten in eine Datei in der Informatica-Verzeichnisstruktur.

Reguläre Ausdrücke

Im Zusammenhang mit Inhaltssätzen ist eine reguläre Expression eine Expression, die Sie beim Parsen und Kennzeichnen von Operationen verwenden können. Verwenden Sie reguläre Ausdrücke, um eine oder mehrere Strings in Eingabedaten anzugeben. Sie können reguläre Ausdrücke in Parser-Umwandlungen im Token-Parsing-Modus verwenden. Sie können reguläre Ausdrücke ebenfalls in Labeler-Umwandlungen im Token-Beschriftungsmodus verwenden.

Parser-Umwandlungen verwenden reguläre Ausdrücke, um Muster in Eingabedaten abzugleichen und alle Matching-Strings in einen oder mehrere Ausgabeports zu parsen. Beispielsweise können Sie mit einem regulären Ausdruck alle E-Mail-Adressen in Eingabedaten kennzeichnen und jede E-Mail-Adresskomponente in einen anderen Ausgang parsen.

Labeler-Umwandlungen verwenden reguläre Ausdrücke, um ein Eingabemuster abzugleichen und eine einzelne Beschriftung zu erstellen. Reguläre Ausdrücke mit mehreren Ausgaben generieren keine Mehrfachbeschriftungen.

Reguläre Ausdruckseigenschaften

Konfigurieren Sie Eigenschaften, die bestimmen, wie ein regulärer Ausdruck Ausgabestrings identifiziert schreibt.

In der folgenden Tabelle werden die Eigenschaften für einen benutzerdefinierten regulären Ausdruck beschrieben:

Eigenschaft	Beschreibung
Anzahl der Ausgaben	Definiert die Anzahl der Ausgabeports, die der reguläre Ausdruck schreibt.
Regulärer Ausdruck	Definiert ein Muster, das die Parser-Umwandlung für den Abgleich von Strings verwendet.
Ausdruck testen	Enthält Daten, die Sie zum Testen des regulären Ausdrucks eingeben. Wenn Sie Daten in diesem Feld eingeben, markiert das Feld Strings, die mit dem regulären Ausdruck übereinstimmen.
Nächster Ausdruck	Navigiert zum nächsten String, der mit dem regulären Ausdruck übereinstimmt, und ändert die Schriftart dieses Strings in fett.
Vorheriger Ausdruck	Navigiert zum vorherigen String, der mit dem regulären Ausdruck übereinstimmt, und ändert die Schriftart dieses Strings in fett.

Tokensätze

Ein Tokensatz enthält Ausdrücke, die bestimmte Token kennzeichnen. Sie können Tokensätze in Labeler-Umwandlungen verwenden, die im Modus "Tokenbeschriftung" ausgeführt werden. Sie können Tokensätze auch in Parser-Umwandlungen verwenden, die den Modus "Token-Parsing" verwenden.

Verwenden Sie Tokensätze, um bestimmte Token als Teil der Bezeichnungs- und Parsing-Operationen zu kennzeichnen. Sie können einen Tokensatz beispielsweise zum Beschriften aller E-Mail Adressen verwenden, die das Format "Kontoname@Domänenname" aufweisen. Nach dem Beschriften der Token können Sie die Parser-Umwandlung verwenden, um E-Mail-Adressen auf angegebene Ausgabeports zu schreiben.

Eigenschaften von Tokensätzen

Konfigurieren Sie Eigenschaften, die die Beschriftungsoperationen für einen Tokensatz angeben.

In der folgenden Tabelle werden die Eigenschaften für einen benutzerdefinierten Zeichensatz beschrieben:

Eigenschaft	Tokensatz-Modus	Beschreibung
Name	N/V	Definiert den Namen des Tokensatzes.
Beschreibung	N/V	Beschreibt den Tokensatz.
Tokensatz-Optionen	N/V	Gibt an, ob der Tokensatz den Modus "Regulärer Ausdruck" oder den Modus "Zeichen" verwendet.
Beschriftung	Regulärer Ausdruck	Definiert die Beschriftung, die eine Labeler-Umwandlung auf Daten anwendet, die mit dem Tokensatz übereinstimmen.
Regulärer Ausdruck	Regulärer Ausdruck	Definiert ein Muster, das von der Labeler-Umwandlung zum Zuordnen von Strings verwendet wird.
Ausdruck testen	Regulärer Ausdruck	Enthält Daten, die Sie zum Testen des regulären Ausdrucks eingeben. Bei der Eingabe von Daten in diesem Feld werden Strings hervorgehoben, die mit dem regulären Ausdruck übereinstimmen.
Nächster Ausdruck	Regulärer Ausdruck	Navigiert zum nächsten String, der mit dem regulären Ausdruck übereinstimmt, und ändert die Schriftart dieses Strings in fett.
Vorheriger Ausdruck	Regulärer Ausdruck	Navigiert zum vorherigen String, der mit dem regulären Ausdruck übereinstimmt, und ändert die Schriftart dieses Strings in fett.

Eigenschaft	Tokensatz-Modus	Beschreibung
Beschriften	Zeichen	Definiert die Beschriftung, die eine Labeler-Umwandlung auf Daten anwendet, die mit dem Zeichensatz übereinstimmen.
Standardmodus	Zeichen	Aktiviert eine einfache Bearbeitungsansicht mit Feldern für den Anfangs- und Endbereich.
Startbereich	Zeichen	Gibt das erste Zeichen in einem Zeichenbereich an.
Endbereich	Zeichen	Gibt das letzte Zeichen in einem Zeichenbereich an. Lassen Sie dieses Feld für Einzelzeichenbereiche leer.
Erweiterter Modus	Zeichen	Ermöglicht eine erweiterte Bearbeitungsansicht, in der Zeichenbereich unter Verwendung von Bereichszeichen und Delimiter-Zeichen manuell eingeben können.
Bereichszeichen	Zeichen	Temporäre Änderungen des Symbols, das einen Zeichenbereich kennzeichnet. Der Zeichenbereich wird auf das Standardzeichen zurückgesetzt, wenn Sie den Zeichensatz schließen.
Delimiter-Zeichen	Zeichen	Temporäre Änderungen des Symbols, das Zeichenbereiche trennt. Das Delimiter-Zeichen wird auf das Standardzeichen zurückgesetzt, wenn Sie den Zeichensatz schließen.

Regeln und Richtlinien für probabilistische Modelle und Klassifizierermodelle

Jedes probabilistische Modell und Klassifizierermodell im Modellrepository gibt eine Datei in der Informatica-Verzeichnisstruktur an. Die Dateien enthalten die Datenwerte und Beschriftungen, die Sie zum Modell im Developer-Tool hinzufügen. Die Dateien enthalten auch die Kompilierungslogik, die die Zuordnungen zwischen den Datenwerten und den Beschriftungen definiert.

Berücksichtigen Sie die folgenden Regeln und Richtlinien, wenn Sie mit probabilistischen Modellen oder Klassifizierermodellen arbeiten:

- Beim Ausführen eines Mappings, das ein Modell enthält, wendet der Datenintegrationsdienst die Logik des kompilierten Modells auf die Eingabedaten der Umwandlung an. Der Datenintegrationsdienst liest die Datenwerte bzw. die Beschriftungen im Modell nicht, wenn das Mapping ausgeführt wird.

- Optional können Sie die Datenwerte und die Beschriftungen aus einem probabilistischen Modell oder einem Klassifizierermodell entfernen. Sie können beispielsweise die vertraulichen oder proprietären Daten aus einem Modell entfernen. Sie können einzelne Datenwerte und Beschriftungen im Developer-Tool entfernen. Sie können alle Datenwerte und Beschriftungen entfernen, wenn Sie ein Modell aus dem Modellrepository exportieren.

  Hinweis: Wenn Sie alle Datenwerte und Beschriftungen aus einem Modell entfernen, können Sie das Modell nicht kompilieren.
- Wenn Sie einen oder mehrere Datenwerte oder Beschriftungen aus einem Modell entfernen, stellt die Logik des kompilierten Modells nicht mehr die aktuellen Daten in der Modelldatei dar. Um die Logik des Modells, die Datenwerte und Beschriftungen zu synchronisieren, kompilieren Sie das Modell erneut. Kompilieren Sie das Modell nicht, wenn Sie die aktuelle Logik des Modells beibehalten möchten.
- Um die Daten in einem Klassifizierermodell oder einem probabilistischen Modell zu schützen, sichern Sie die Modelldatei in der Informatica-Verzeichnisstruktur. Sichern Sie die Datei, bevor Sie alle Datenwerte und Beschriftungen aus einem Modell entfernen.
- Suchen Sie die Modelldateien auf dem Computer, der den Content-Management-Dienst hostet.

  Probabilistische Modelldateien weisen den folgenden Standardspeicherort und die folgende Dateinamenerweiterung auf:

  ```
  <Informatica_Installation_Directory>/tomcat/bin/ner/<filename>.ner
  ```

  Klassifizierermodelldateien weisen den folgenden Standardspeicherort und die folgende Dateinamenerweiterung auf:

  ```
  <Informatica_Installation_Directory>/tomcat/bin/classifier/<filename>.classifier
  ```
- Wenn Sie ein Upgrade der Informatica-Installation durchführen, müssen Sie die probabilistischen Modelle und Klassifizierermodelle möglicherweise kompilieren, bevor Sie die Modelle in einem Mapping verwenden. Wenn ein Modell keine Daten enthält, ersetzen Sie die aktuelle Datei in der Informatica-Verzeichnisstruktur mit der Sicherungsdatei, die die Daten enthält.

Verwalten von Beschriftungen in Klassifizierer- und probabilistischen Modellen

Verwenden Sie zum Überprüfen und Aktualisieren der Beschriftungen in einem probabilistischen oder einem Klassifizierermodell das Dialogfeld **Beschriftungen verwalten**.

1. Öffnen Sie den Inhaltssatz, der das Klassifizierermodell enthält.
2. Wählen Sie den Modellnamen aus und klicken Sie auf **Bearbeiten**.
3. Öffnen Sie das Dialogfeld **Beschriftungen verwalten**.

   Im Dialogfeld werden die Beschriftungen im Modell aufgelistet.

Verwandte Themen:

- "Beschriftungsverwaltung im Klassifizierermodell" auf Seite 49
- "Konfiguration des probabilistischen Modells" auf Seite 65

Erstellen eines Inhaltssatzes

Erstellen Sie einen Inhaltssatz zur Verwaltung von Referenzdatenobjekten, die auf ein einzelnes Projekt, einen einzelnen Informationstyp oder Geschäftszweck verweisen.

1. Wählen Sie in der Ansicht **Objekt-Explorer** ein Projekt oder einen Ordner aus, um den Inhaltssatz zu speichern.

2. Klicken Sie auf **Datei** > **Neu** > **Inhaltssatz**.
3. Geben Sie einen Namen für den Inhaltssatz ein.
4. Wählen Sie optional **Durchsuchen** aus, um den Modellrepository-Speicherort für den Inhaltssatz zu ändern.
5. Klicken Sie auf **Fertig stellen**.

Erstellen eines Referenzdatenobjekts in einem Inhaltssatz

Sie können Zeichen-, Muster- und Tokensätze, reguläre Ausdrücke sowie probabilistische und Klassifikatormodelle in einem Inhaltssatz erstellen.

1. Öffnen Sie einen Inhaltssatz im Editor und wählen Sie die Ansicht **Inhalt** aus.
2. Wählen Sie einen Typ für das Referenzdatenobjekt aus.
3. Klicken Sie auf **Hinzufügen**.
4. Geben Sie einen Namen für das Referenzdatenobjekt ein.

   Geben Sie optional eine Beschreibung des Objekts ein.
5. Konfigurieren Sie die Eigenschaften des Referenzdatenobjekts.
6. Klicken Sie auf **Fertig stellen**.

Tipp: Sie können Referenzdatenobjekte aus einem Inhaltssatz in einen anderen Inhaltssatz kopieren. Verwenden Sie die Optionen **Kopieren** und **Einfügen aus**, um eine Kopie eines Objekts in einem Inhaltssatz zu erstellen. Verwenden Sie die **STRG**-Taste, um mehrere Inhaltssatzobjekte auszuwählen.

KAPITEL 4

Klassifizierermodelle

Dieses Kapitel umfasst die folgenden Themen:

- Klassifikatormodelle – Übersicht, 44
- Struktur von Klassifikatormodellen, 45
- Klassifikator-Scores, 45
- Beispiel für eine Klassifikatorumwandlung, 46
- Optionen des Klassifizierermodells, 46
- Referenzdaten in Klassifikatormodellen, 48
- Beschriftungsdaten in Klassifikatormodellen, 48
- Konfiguration des Klassifizierermodells, 50
- Filter- und Suchvorgänge, 54
- Kopieren und Einfügen, 55

Klassifikatormodelle – Übersicht

Ein Klassifikatormodell ist ein Referenzdatenobjekt in einem Inhaltssatz. Verwenden Sie ein Klassifikatormodell, um lange Zeichenfolgen zu analysieren, die mehrere Werte enthalten. Ein Klassifikatormodell identifiziert den häufigsten Informationstyp in jeder Zeichenfolge.

Fügen Sie ein Klassifikatormodell zu einer Klassifikatorumwandlung hinzu. Die Umwandlung sucht nach gemeinsamen Werten zwischen den Daten des Klassifikatormodells und den Daten in jeder Eingabezeile. Die Umwandlung verwendet die gemeinsamen Werte, um die durch jede Zeile dargestellten Informationstypen zu kategorisieren.

Verwenden Sie ein Klassifikatormodell, wenn die Eingabedaten folgende Eigenschaften aufweisen:

- Die Eingabedaten enthalten Text. Klassifikatormodelle wenden natürliche Sprachprozesse auf Textdaten an, um die Informationstypen im Text zu identifizieren. Die natürlichen Sprachprozesse erkennen relevante Wörter in der Eingabezeichenfolge. Die natürlichen Sprachprozesse ignorieren irrelevante Wörter.
- Die Eingabezeichenfolgen enthalten mehrere Werte. Sie können beispielsweise eine Datenspalte erstellen, die in jedem Feld die Inhalte einer E-Mail-Nachricht enthält.

Die Klassifikatorumwandlung liest Zeichenfolgen. Die Umwandlung beschränkt dabei nicht die Länge der Eingabezeichenfolgen.

Kompilieren Sie Klassifikatormodelle im Developer-Tool. Wenn Sie ein Modell kompilieren, erstellen Sie Assoziationen zwischen ähnlichen Datenwerten im Modell. Die Klassifikatorumwandlung verwendet die kompilierten Daten zur Suche nach Informationen in den Eingabedaten.

Struktur von Klassifikatormodellen

Ein Klassifizierermodell enthält Referenzdaten- und Beschriftungswerte. Bei den Referenzdatenwerten handelt es sich um die zu klassifizierenden Daten. Mit den Beschriftungswerten werden die Informationstypen festgelegt, die von einer Klassifiziererumwandlung in den Daten erkannt werden können.

Ein Klassifikatormodell umfasst außerdem Kompilierungsdaten. Die Klassifiziererumwandlung verwendet die Kompilierungsdaten, um die Ähnlichkeiten zwischen den Referenzdaten im Modell und den Eingabedaten der Umwandlung zu messen. Wenn Sie ein Klassifizierermodell kompilieren, erstellen oder aktualisieren Sie die Kompilierungsdaten. Wenn eine Klassifiziererumwandlung die Eingabedaten mit den Modelldaten vergleicht, gibt die Umwandlung Beschriftungswerte zurück, die jede Eingabedatenzeile beschreiben.

Das Developer Tool schreibt die Referenzdatenwerte, die Beschriftungswerte und die Kompilierungsdaten in eine Datei in der Informatica-Verzeichnisstruktur. Der Dateiname wird im Objekt des Klassifizierermodells im Modellrepository gespeichert. Wenn Sie ein Klassifizierermodell speichern, schreiben Sie die aktuellen Referenzdatenwerte und die Beschriftungswerte in die Datei. Wenn Sie das Modell kompilieren, aktualisieren Sie die Kompilierungsdaten in der Datei. Sie können den Dateinamen aus den Modelleigenschaften im Developer Tool lesen.

Klassifikator-Scores

Eine Klassifikatorumwandlung vergleicht jede Zeile an Eingabedaten mit jeder Zeile an Referenzdaten in einem Klassifikatormodell. Die Umwandlung berechnet einen Score für jeden Vergleich. Die Scores geben den Ähnlichkeitsgrad zwischen den Eingabezeilen und den Referenzdatenzeilen an.

Wenn Sie ein Mapping ausführen, das eine Klassifikatorumwandlung beinhaltet, gibt das Mapping die Beschriftung aus, welche die Referenzdatenzeile mit dem höchsten Score identifiziert. Scores liegen im Bereich von 0 bis 1. Ein hoher Score steht für eine starke Übereinstimmung zwischen den Eingabedaten und den Modelldaten.

Überprüfen Sie die Klassifikator-Scores und stellen Sie sicher, dass die ausgegebenen Beschriftungen die jeweiligen Zeilen an Eingabedaten akkurat beschreiben. Sie können die Scores außerdem überprüfen, um sicherzustellen, dass das Klassifikatormodell den Eingabedaten angemessen ist. Wenn die Umwandlungsausgabe sehr viele niedrige Scores enthält, ist das Klassifikatormodell möglicherweise ungeeignet. Kompilieren Sie das Modell erneut, um die Vergleiche zu verbessern. Wenn das kompilierte Modell die Scores nicht verbessert, ersetzen Sie das Modell in der Umwandlung.

Beispiel für eine Klassifikatorumwandlung

Sie können ein Klassifikatormodell und eine Klassifikatorumwandlung nutzen, um E-Mail-Nachrichten basierend auf ihrem enthaltenen Text zu kategorisieren.

Sie arbeiten beispielsweise als Datenverwalter im Supportcenter eines Software-Herstellers. Sie überprüfen die E-Mail-Nachrichten, die das Supportcenter von Kunden erhält. Da das Unternehmen Kunden in zahlreichen Ländern hat, erhält das Supportcenter E-Mails in unterschiedlichen Sprachen. Sie möchten die E-Mails nach Sprache sortieren, damit Sie die Nachrichten jeweils an die entsprechende Abteilung weiterleiten können.

Führen Sie zum Sortieren der E-Mails die folgenden Schritte durch:

1. Schreiben Sie die E-Mail-Nachrichten in eine einzelne Datei oder Datenbanktabelle.
2. Erstellen Sie ein Datenobjekt in dem Modellrepository, das die Datei oder die Datenbanktabelle liest.
3. Erstellen Sie im Modellrepository Datenobjekte für jede Sprache, die in einer Nachricht verwendet wird.
4. Erstellen Sie ein Klassifizierermodell, das Beispieltexte für jede Sprache enthält.

   Hinweis: Als Quelldaten für das Modell können Sie Auszüge aus den E-Mail-Nachrichten verwenden.
5. Fügen Sie das Klassifizierermodell einer wiederverwendbaren Klassifiziererumwandlung hinzu.
6. Konfigurieren Sie eine Zuordnung, um die Klassifiziererumwandlung auf die Nachrichtendaten anzuwenden.

   Führen Sie zum Konfigurieren der Zuordnung die folgenden Schritte durch:
   - Fügen Sie der Zuordnung die Klassifiziererumwandlung und die Datenobjekte hinzu.
   - Verbinden Sie den Eingabeport einer Klassifiziererumwandlung mit dem Quelldatenobjekt.
   - Verbinden Sie die Ausgabeports einer Klassifiziererumwandlung mit den Zieldatenobjekten.

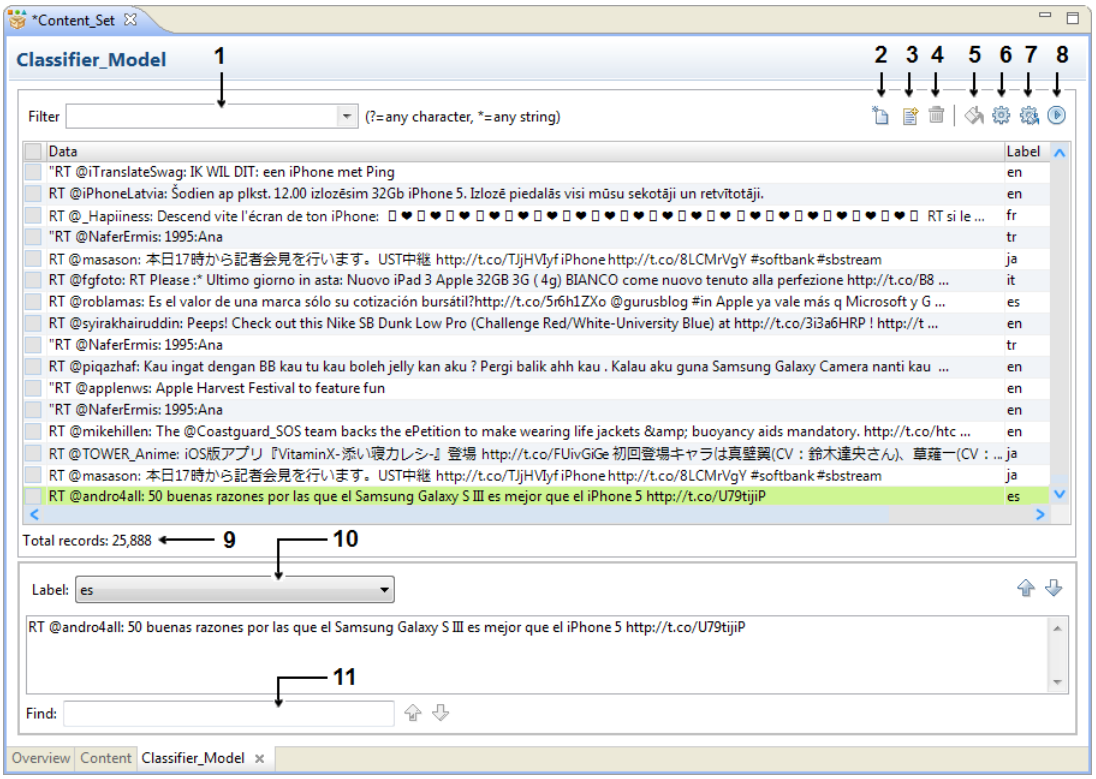
Beim Ausführen des Mappings analysiert die Klassifikatorumwandlung die E-Mail-Nachrichten und schreibt den E-Mail-Text in das korrekte Datenziel. Sie können die Datenziele mit den Teammitgliedern in jeder Abteilung gemeinsam verwenden.

Optionen des Klassifizierermodells

Das Developer Tool zeigt die Daten des Klassifizierermodells in einem Editor an, der sich aus einem unteren und einem oberen Bereich zusammensetzt. Im oberen Bereich wird jede Referenzdatenzeile und Beschriftung angezeigt, die Sie den Daten zuweisen. Im unteren Bereich werden die Inhalte einer von Ihnen ausgewählten Zeile angezeigt.

Verwenden Sie den oberen Bereich, um die Referenzdatenzeilen zu überprüfen und alle Zeilen zu ermitteln, die keine Beschriftungen verwenden. Verwenden Sie den unteren Bereich, um die Inhalte einer Zeile zu überprüfen und der Zeile eine Beschriftung zuzuweisen. Im oberen Bereich werden maximal 100 Zeichen pro Zeile angezeigt. Im unteren Bereich werden alle Daten in der von Ihnen ausgewählten Zeile angezeigt.

Die folgende Abbildung zeigt den Editor des Klassifizierermodells:



Der Editor enthält die folgenden Optionen:

1. Feld „Filter“

   Filtert die Liste der Referenzdatenzeilen basierend auf dem Datenwert oder der Beschriftung, die Sie angeben.

2. Zeile hinzufügen

   Fügt ein leere Referenzdatenzeile ein.

3. Daten anhängen

   Importiert Daten aus einem Datenobjekt in das Modellrepository.

4. Löschen

   Löscht die von Ihnen ausgewählten Referenzdatenzeilen. Verwenden Sie die Kontrollkästchen zum Auswählen der Zeilen.

5. Beschriftung zuweisen

   Weist mindestens einer von Ihnen ausgewählten Referenzdatenzeile eine Beschriftung zu. Verwenden Sie die Kontrollkästchen zum Auswählen der Zeilen.

6. Eigenschaften bearbeiten

   Zeigt die Eigenschaften des Klassifizierermodells an.

7. Beschriftungen verwalten

   Öffnet das Dialogfeld **Beschriftungen verwalten**. Verwenden Sie das Dialogfeld, um dem Klassifizierermodell Beschriftungswerte hinzuzufügen bzw. aus diesem zu löschen.

8. Kompilieren

   Kompiliert das Klassifizierermodell.

9. Datensätze insgesamt

   Gibt die Anzahl der Referenzdatenzeilen im Klassifizierermodell an.

10. Feld „Beschriftung“

    Zeigt einen Beschriftungswert an, den Sie auf die aktuelle Referenzdatenzeile anwenden können.

11. Feld „Suchen“

    Sucht nach einem Datenwert, den Sie in der aktuellen Referenzdatenzeile festlegen.

Referenzdaten in Klassifikatormodellen

Ein Klassifikatormodell enthält eine Referenzdatenspalte, die Sätze, Absätze oder ganze Seiten mit Text beinhalten kann. Die Referenzdaten stellen die verschiedenen Arten von Texteingaben dar, die eine Klassifikatorumwandlung in einem Mapping lesen kann. Achten Sie beim Erstellen eines Modells darauf, dass die Referenzdaten auch die Textarten enthalten, die beim Ausführen des Mappings zu erwarten sind.

Zum Erstellen eines Klassifikatormodells können Sie die Quelldaten des Mappings verwenden. Wählen Sie einen Auszug der Quelldaten und kopieren Sie diese Datenstichprobe in das Modell.

Berücksichtigen Sie die folgenden Regeln und Richtlinien bei der Arbeit mit Referenzdaten in Klassifikatormodellen:

- Ein Referenzdatenfeld kann beliebig lang sein. Sie können in jedes Datenfeld mehrere Seiten Text eingeben.
- Referenzdaten werden aus einem Datenobjekt importiert.
- Referenzdatenwerte können nicht bearbeitet werden. Allerdings können Sie eine Datenzeile löschen.
- Wenn Sie ein Klassifikatormodell kompilieren, werden beim Kompilierungsprozess keine Zahlenwerte in den Referenzdaten berücksichtigt.

Beschriftungsdaten in Klassifikatormodellen

Ein Klassifizierermodell enthält eine oder mehrere beschreibende Beschriftungen, die die Informationstypen in den Referenzdatenzeilen zusammenfassen. Weisen Sie jeder Referenzdatenzeile eine Beschriftung zu.

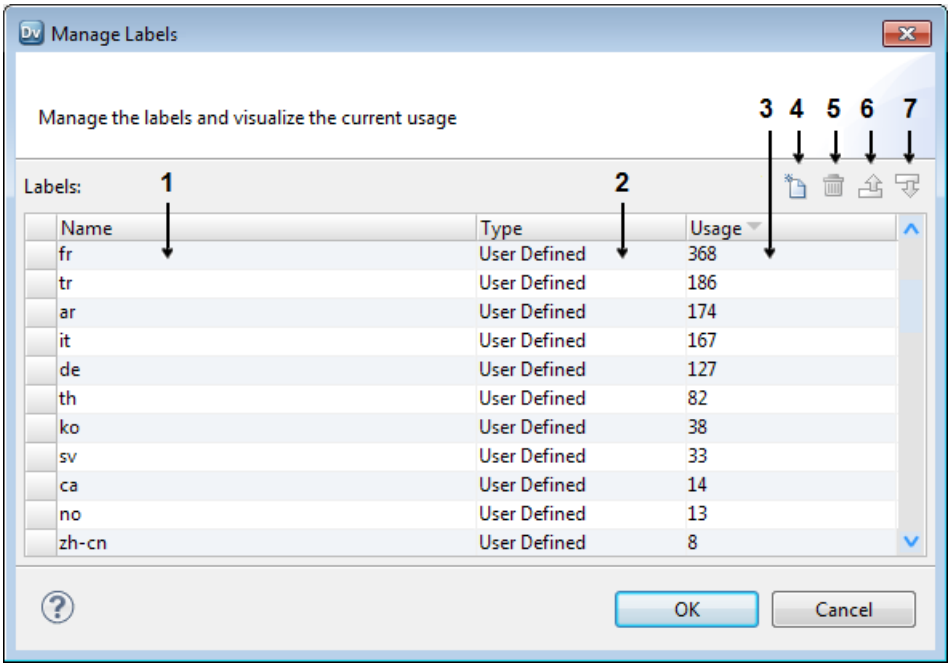
Wenn Sie einem Klassifizierermodell Daten aus einer Datenquelle hinzufügen, können Sie eine Spalte als Beschriftungsdatenspalte festlegen. Darüber hinaus können Sie Beschriftungen im Modell erstellen.

Beschriftungen existieren unabhängig von den Referenzdatenwerten, die sie beschreiben. Wenn Sie die Referenzdatenzeilen löschen, die eine Beschriftung verwenden, hat dies keine Auswirkungen auf die Beschriftung im Modell. Und wenn Sie eine Beschriftung löschen, werden dadurch nicht die mit der Beschriftung assoziierten Referenzdatenwerte gelöscht.

Beschriftungsverwaltung im Klassifizierermodell

Verwenden Sie das Dialogfeld **Beschriftungen verwalten**, um die Beschriftungswerte in einem Klassifizierermodell zu überprüfen und zu aktualisieren. Sie können die Beschriftungswerte auch sortieren und aktualisieren.

Die folgende Abbildung zeigt das Dialogfeld **Beschriftungen verwalten**:



Das Dialogfeld **Beschriftungen verwalten** enthält die folgenden Elemente:

1. Spalte „Name“.

   Enthält die Beschriftungswerte, die die Klassifiziererumwandlung auf die Eingabedatenzeilen anwenden kann. Sie können die Beschriftungen nach Name sortieren.

2. Spalte „Typ“.

   Gibt die Quelle der Beschriftungswerte an. Das Klassifizierermodell gibt alle Beschriftungen als benutzerdefinierte Werte an.

3. Spalte „Verwendung“.

   Gibt die Anzahl der Referenzdatenzeilen an, die jede Beschriftung verwenden. Sie können die Beschriftungen nach der Anzahl der Zeilen sortieren.

4. Schaltfläche „Hinzufügen“.

   Fügt dem Klassifizierermodell eine Beschriftung hinzu. Geben Sie einen Beschriftungswert in der Spalte „Name“ in der Zeile ein.

   Hinweis: Doppelklicken Sie zum Aktualisieren eines Beschriftungswerts auf den Wert und geben Sie den gewünschten Wert ein.

5. Schaltfläche „Löschen“.

   Löscht eine Beschriftung aus dem Klassifizierermodell.

6. Nach-oben-Pfeil.

   Verschiebt die Beschriftung im Dialogfeld eine Zeile nach oben.

7. Nach-unten-Pfeil.

   Verschiebt die Beschriftung im Dialogfeld eine Zeile nach unten.

Konfiguration des Klassifizierermodells

Die Schritte zum Konfigurieren eines Klassifizierermodells beginnen mit den zu klassifizierenden Daten. Der Inhalt der Referenzdaten, die zum Modell hinzugefügt werden, muss die Daten widerspiegeln, die Sie mit der Klassifiziererumwandlung verbunden haben. Die Umwandlung vergleicht die Datenwerte und Muster in den Eingabedaten mit den Datenwerten und Mustern im Klassifizierermodell.

Führen Sie zum Erstellen eines Klassifizierermodells, das in einer Klassifiziererumwandlung verwendet werden kann, die folgenden Aufgaben durch:

1. Geben Sie die Referenzdaten- und Beschriftungswerte an, die dem Modell hinzugefügt werden sollen.

   Sie können ein Fragment der zu klassifizierenden Daten verwenden. Erstellen Sie ein Datenobjekt in dem Modellrepository, das das Datenfragment liest.
2. Erstellen Sie einen Inhaltssatz und fügen Sie diesem ein Klassifizierermodell hinzu.
3. Fügen Sie dem Modell die Referenzdatenwerte hinzu.
4. Fügen Sie dem Modell die Beschriftungswerte hinzu.

   Sie können die Daten aus dem Datenobjekt im Modellrepository importieren. Sie können auch eine einzelne Referenzdatenzeile oder eine einzelne Beschriftung eingeben.
5. Weisen Sie jeder Referenzdatenzeile eine Beschriftung zu.

   Sie können mehreren Zeilen in einem einzelnen Vorgang eine Beschriftung zuweisen.
6. Kompilieren Sie das Modell.

Nach dem Kompilieren des Klassifizierermodells können Sie das Modell in einer Klassifiziererumwandlung verwenden.

Erstellen eines Klassifizierermodells

Verwenden Sie ein Datenobjekt als Quelle für die Daten des Klassifizierermodells.

Ein Klassifizierermodell arbeitet optimal, wenn Sie die Eingabedaten aus der Klassifiziererumwandlung als Quelle für die Modellreferenzdaten verwenden.

1. Öffnen oder erstellen Sie im Objekt-Explorer einen Inhaltssatz.
2. Wählen Sie die Ansicht **Inhalt**.
3. Wählen Sie **Klassifikatormodelle** und klicken Sie auf **Hinzufügen**.

   Der Klassifikatormodell-Assistent wird geöffnet.
4. Geben Sie einen Namen für das Klassifikatormodell ein.

   Geben Sie optional eine Textbeschreibung des Modells ein.
5. Durchsuchen Sie das Modellrepository und wählen Sie das Datenobjekt aus, das die zu importierenden Daten enthält.

   Wählen Sie kein Datenobjekt der sozialen Medien aus.

   Klicken Sie auf **Weiter**.

6. Überprüfen Sie die Spalten im Datenobjekt und wählen Sie mindestens eine Spalte zum Hinzufügen zum Modell aus. Sie können Referenzdatenspalten und eine Beschriftungsspalte im selben Vorgang hinzufügen.
   - Wählen Sie zum Importieren einer Datenspalte als Referenzdaten den Namen der Spalte aus und klicken Sie auf **Daten**.

     Sie können mehrere Datenspalten auswählen. Das Developer Tool führt die Inhalte der von Ihnen ausgewählten Spalten in einer Spalte zusammen.
   - Wählen Sie zum Importieren einer Datenspalte als Beschriftungswerte den Namen der Spalte aus und klicken Sie auf **Beschriftung**.

   Wenn Sie Referenzdaten- und Beschriftungswerte importieren, weist das Developer Tool die Beschriftung in jeder Zeile zur Referenzdatenzeichenfolge in derselben Zeile zu. Sie können die Daten in der Vorschau anzeigen, bevor Sie die Spalten auswählen. Sie können die Beschriftungszuweisungen nach dem Erstellen des Modells ändern.

   Klicken Sie auf **Weiter**.
7. Wählen Sie die Anzahl der Zeilen aus, die Sie aus der Datenquelle importieren möchten.

   Das Developer Tool importiert standardmäßig alle Zeilen aus der Datenquelle. Wenn Sie eine Anzahl eingeben, zählt das Modell die Zeilen vom Anfang des Datensatzes.
8. Klicken Sie auf **Fertig stellen** und speichern Sie das Modell.

Überprüfen Sie nach dem Erstellen des Modells die Beschriftungszuweisungen und kompilieren Sie das Modell.

Anhängen von Daten aus einer Datenquelle an ein Klassifizierermodell

Sie können mehrere Zeilen mit Referenzdatenwerten oder Beschriftungswerten in einem einzelnen Vorgang in ein Klassifizierermodell importieren.

1. Öffnen Sie den Inhaltssatz, der das Klassifizierermodell enthält.
2. Wählen Sie den Modellnamen aus und klicken Sie auf **Bearbeiten**.
3. Klicken Sie auf **Daten anfügen**.

   Der Klassifikatormodell-Assistent wird geöffnet.
4. Durchsuchen Sie das Modellrepository und wählen Sie das Datenobjekt aus, das die zu importierenden Daten enthält.

   Wählen Sie kein Datenobjekt der sozialen Medien aus.

   Klicken Sie auf **Weiter**.
5. Überprüfen Sie die Spalten im Datenobjekt und wählen Sie mindestens eine Spalte zum Hinzufügen zum Modell aus. Sie können Referenzdatenspalten und eine Beschriftungsspalte im selben Vorgang hinzufügen.
   - Wählen Sie zum Importieren einer Datenspalte als Referenzdaten den Namen der Spalte aus und klicken Sie auf **Daten**.

     Sie können mehrere Datenspalten auswählen. Das Developer Tool führt die Inhalte der von Ihnen ausgewählten Spalten in einer Spalte zusammen.
   - Wählen Sie zum Importieren einer Datenspalte als Beschriftungswerte den Namen der Spalte aus und klicken Sie auf **Beschriftung**.

Wenn Sie Referenzdaten- und Beschriftungswerte importieren, weist das Developer Tool die Beschriftung in jeder Zeile zur Referenzdatenzeichenfolge in derselben Zeile zu. Sie können die Daten in der Vorschau anzeigen, bevor Sie die Spalten auswählen. Sie können die Beschriftungszuweisungen nach dem Erstellen des Modells ändern.

Klicken Sie auf **Weiter**.

6. Wählen Sie die Anzahl der Zeilen aus, die Sie aus der Datenquelle importieren möchten.

   Das Developer Tool importiert standardmäßig alle Zeilen aus der Datenquelle. Wenn Sie eine Anzahl eingeben, zählt das Modell die Zeilen vom Anfang des Datensatzes.

7. Klicken Sie auf **Fertig stellen** und speichern Sie das Modell.

Hinzufügen einer Referenzdatenzeile zu einem Klassifizierermodell

Sie können einem Klassifizierermodell eine einzelne Zeile mit Referenzdaten hinzufügen.

1. Öffnen Sie den Inhaltssatz, der das Klassifizierermodell enthält.
2. Wählen Sie den Modellnamen aus und klicken Sie auf **Bearbeiten**.
3. Klicken Sie auf **Zeile hinzufügen**.

   Das Developer Tool fügt unterhalb der aktuellen Zeile eine neue Zeile in den Referenzdaten ein.

4. Geben Sie die Referenzdatenwerte in die Zeile ein.

   Sie können Windows-Tastaturkürzel verwenden, um Daten in die Zeile einzufügen.

Hinzufügen einer Beschriftung zu einem Klassifikatormodell

Sie können einem Klassifizierermodell eine einzelne Beschriftung hinzufügen.

1. Öffnen Sie den Inhaltssatz, der das Klassifizierermodell enthält.
2. Wählen Sie den Modellnamen aus und klicken Sie auf **Bearbeiten**.
3. Öffnen Sie das Dialogfeld **Beschriftungen verwalten**.

   Im Dialogfeld werden die Beschriftungen im Modell aufgelistet.

4. Klicken Sie auf **Neu**.

   Das Developer Tool fügt unten in der Liste der Beschriftungen eine Zeile ein.

5. Doppelklicken Sie auf den Standardwert in der Spalte „Name“ und geben Sie einen Beschriftungsnamen ein.
6. Klicken Sie auf **OK**.

Nach dem Erstellen der Beschriftung können Sie sie einer oder mehreren Referenzdatenzeilen zuweisen. In der Spalte „Verwendung“ im Dialogfeld **Beschriftungen verwalten** wird die Anzahl der Zeilen angegeben, die die Beschriftung verwenden.

Zuweisen einer Beschriftung zu Referenzdatenzeilen

Sie können einer oder mehreren Referenzdatenzeilen in einem einzelnen Vorgang eine Beschriftung zuweisen.

1. Öffnen Sie den Inhaltssatz, der das Modell enthält.
2. Wählen Sie den Modellnamen aus und klicken Sie auf **Bearbeiten**.
3. Wählen Sie mindestens eine Referenzdatenzeile aus. Verwenden Sie die Kontrollkästchenoptionen zum Auswählen der Zeilen.

Hinweis: Sie können die Filteroption zum Anzeigen aller Zeilen verwenden, die einen von Ihnen angegebenen Datenwert enthalten. Verwenden Sie die Kontrollkästchenoption „Alle auswählen“, um alle Zeilen mit dem Wert auszuwählen.

4. Klicken Sie auf **Beschriftung zuweisen**.

   Das Developer Tool zeigt die Liste der Beschriftungen im Klassifizierermodell an.

5. Wählen Sie einen Beschriftungswert aus und klicken Sie auf **Zuweisen**.

   Das Developer Tool aktualisiert die von Ihnen ausgewählten Referenzdatenzeilen mit dem Beschriftungswert.

Kompilieren Sie optional das Modell, um der Logik des Klassifizierermodells die Beschriftungsnamen hinzuzufügen.

Erkennen ungenutzter Beschriftungswerte

Verwenden Sie das Dialogfeld **Beschriftungen verwalten**, um nach allen Beschriftungswerten zu suchen, die im Klassifizierermodell ungenutzt bleiben. Im Dialogfeld **Beschriftungen verwalten** werden Nutzungsdaten für die Beschriftungswerte im Klassifizierermodell angezeigt. Verwenden Sie die Nutzungsdaten, um die Anzahl der Referenzdatenzeilen zu überprüfen, die einen Beschriftungswert verwenden, und um nach ungenutzten Beschriftungswerten zu suchen.

1. Öffnen Sie den Inhaltssatz, der das Klassifizierermodell enthält.
2. Wählen Sie den Modellnamen aus und klicken Sie auf **Bearbeiten**.
3. Öffnen Sie das Dialogfeld **Beschriftungen verwalten**. Das Dialogfeld listet die Beschriftungen im Klassifizierermodell auf.
4. Überprüfen Sie die Daten in der Spalte „Nutzung“ für jede Beschriftung.

   In der Spalte „Nutzung“ wird die Anzahl der Referenzdatenzeilen aufgelistet, die die Beschriftung verwenden. Ist ein Beschriftungswert ungenutzt, weist die Spalte „Nutzung“ den Wert null auf.

Löschen von Zeilen aus einem Klassifizierermodell

Sie können eine oder mehrere Referenzdatenzeilen in einer einzelnen Aktion aus einem Klassifizierermodell löschen.

1. Öffnen Sie den Inhaltssatz, der das Modell enthält.
2. Wählen Sie den Modellnamen aus und klicken Sie auf **Bearbeiten**.
3. Wählen Sie mindestens eine Referenzdatenzeile aus. Verwenden Sie die Kontrollkästchenoptionen zum Auswählen der Zeilen.
4. Klicken Sie auf **Löschen**.

   Das Developer Tool entfernt die von Ihnen ausgewählten Zeilen aus dem Klassifizierermodell.

Verwenden Sie die Tastenkombination Strg + Z, um den Vorgang rückgängig zu machen.

Löschen einer Beschriftung aus einem Klassifikatormodell

Verwenden Sie das Dialogfeld **Beschriftungen verwalten**, um eine Beschriftung aus einem Klassifizierermodell zu löschen.

1. Öffnen Sie den Inhaltssatz, der das Modell enthält.
2. Wählen Sie den Modellnamen aus und klicken Sie auf **Bearbeiten**.
3. Öffnen Sie das Dialogfeld **Beschriftungen verwalten**.

4. Klicken Sie auf **Löschen**.
5. Klicken Sie auf **Ja**, um den Vorgang zu bestätigen.

   Das Developer Tool löscht die Beschriftung aus dem Modell. Das Developer Tool löscht keine Referenzdatenzeilen, die die Beschriftung verwenden.
6. Klicken Sie auf **OK**, um das Dialogfeld zu schließen.

Verwenden Sie die Tastenkombination Strg + Z, um den Vorgang rückgängig zu machen.

Kompilieren eines Klassifikatormodells

Jedes Mal, wenn Sie einen Beschriftungswert oder einen Referenzdatenwert in einem Klassifikatormodell bearbeiten, müssen Sie das Modell danach kompilieren. Wenn Sie das Modell kompilieren, aktualisieren Sie die Kompilierungsdaten im Modell.

▶ Um die Kompilierungsdaten zu aktualisieren, öffnen Sie das Modell im Developer-Tool und klicken Sie auf **Kompilieren**.

Filter- und Suchvorgänge

Verwenden Sie die Filteroptionen zum Ausblenden oder Anzeigen der Referenzdatenzeilen, die die von Ihnen angegebenen Kriterien erfüllen. Wenn Sie einen Filter anwenden, können Sie zusätzliche Aktionen in den Datenzeilen durchführen, die im Klassifizierermodell angezeigt werden. Sie können beispielsweise einen Beschriftungswert auf alle Datenzeilen anwenden.

Verwenden Sie die Filteroption, um die folgenden Aufgaben durchzuführen:

- Suchen Sie nach den Referenzdatenzeilen, die einen von Ihnen eingegebenen Wert enthalten.
- Suchen Sie nach den Referenzdatenzeilen, die eine von Ihnen ausgewählte Beschriftung enthalten.
- Suchen Sie nach den Referenzdatenzeilen, die keine Beschriftung verwenden.

Sie können auch nach einem Datenwert in einer Referenzdatenzeile suchen.

Verwenden eines Datenwerts zum Filtern der Referenzdatenzeilen

Verwenden Sie den Filter, um sicherzustellen, dass eine oder mehrere Referenzdatenzeilen die erwarteten Datenwerte enthalten.

1. Öffnen Sie den Inhaltssatz, der das Klassifizierermodell enthält.
2. Wählen Sie den Modellnamen aus und klicken Sie auf **Bearbeiten**.
3. Geben Sie einen Wert im Feld „Filter“ ein.

   Sie können Platzhalterzeichen in dem Wert verwenden, den Sie eingeben.

   Das Developer Tool zeigt die Referenzdatenzeilen an, die den Filtertext enthalten.

Verwenden eines Beschriftungswerts zum Filtern der Referenzdatenzeilen

Verwenden Sie den Filter zum Anzeigen oder Ausblenden der Referenzdatenzeilen, die eine von Ihnen ausgewählte Beschriftung verwenden.

1. Öffnen Sie den Inhaltssatz, der das Modell enthält.
2. Wählen Sie den Modellnamen aus und klicken Sie auf **Bearbeiten**.
3. Wählen Sie einen Beschriftungswert im Menü „Filter“ aus.

   Im Developer Tool werden die Referenzdatenzeilen angezeigt, die den Beschriftungswert verwenden.

   Hinweis: Zum Auffinden aller Referenzdatenzeilen, die keine Beschriftung verwenden, wählen Sie im Menü „Filter“ die Option **Keine Beschriftung** aus.

Suchen eines Werts in einer Referenzdatenzeile

Verwenden Sie das Feld „Suchen“, um nach einem Datenwert in einer von Ihnen ausgewählten Zeile zu suchen.

1. Öffnen Sie den Inhaltssatz, der das Modell enthält.
2. Wählen Sie den Modellnamen aus und klicken Sie auf **Bearbeiten**.
3. Wählen Sie eine Referenzdatenzeile aus.
4. Geben Sie einen Wert im Feld „Suchen“ ein.

   Das Modell markiert die Instanzen des Werts in der Referenzdatenzeile.
5. Verwenden Sie den nach oben oder nach unten weisenden Pfeil, um nach weiteren Instanzen des Werts in der Zeile zu suchen.

Kopieren und Einfügen

Sie können ein Klassifikatormodell aus einem Inhaltssatz in einen anderen in einem Modell-Repository kopieren. Kopieren Sie ein Klassifikatormodell, um Ressourcen mit anderen Benutzern des Developer-Tools zu teilen.

Sie können ein Modell in einen anderen Inhaltssatz kopieren oder in den aktuellen Inhaltssatz importieren. Sie können mehrere Modelle aus mehreren Inhaltssätzen im Repository in einem einzigen Vorgang importieren.

Wenn Sie ein Modell kopieren, erstellt der Content Management Service eine Kopie der Modelldatendatei auf dem Servicerechner. Jedes Modell verwendet eine andere Datendatei.

Kopieren eines Klassifikatormodells in einen anderen Inhaltssatz

Sie können ein Klassifikatormodell aus einem Inhaltssatz in einen anderen in einem Modell-Repository kopieren. Beim Kopieren eines Klassifikatormodells geben Sie das Modellobjekt sowie die Quell- und Zielinhaltssätze an.

1. Öffnen Sie den Inhaltssatz, der das Klassifikatormodell enthält.
2. Wählen Sie ein Klassifikatormodell und klicken Sie auf **Kopieren nach**.
3. Durchsuchen Sie das Modell-Repository und wählen Sie einen Inhaltssatz.

Sie können das Klassifikatormodell in einen Inhaltssatz des aktuellen oder eines anderen Projekts kopieren.

4. Klicken Sie auf **OK**.

   Das Developer-Tool kopiert das Klassifikatormodell in den gewählten Inhaltssatz.

Importieren eines Klassifikatormodells aus einem anderen Inhaltssatz

Sie können ein Klassifikatormodell aus einem Inhaltssatz in einen anderen in einem Modell-Repository importieren. Beim Importieren eines Klassifikatormodells geben Sie ein oder mehrere Modellobjekte sowie die Quell- und Zielinhaltssätze an.

1. Öffnen Sie den Inhaltssatz, der das Klassifikatormodell enthalten soll.
2. Wählen Sie ein Klassifikatormodell und klicken Sie auf **Einfügen aus**.
3. Durchsuchen Sie das Modell-Repository und wählen Sie ein Klassifikatormodell.

   Sie können das Klassifikatormodell aus einem Inhaltssatz des aktuellen oder eines anderen Projekts einfügen.

4. Klicken Sie auf **OK**.

   Das Developer-Tool fügt das Klassifikatormodell in den aktuellen Inhaltssatz ein.

KAPITEL 5

Probabilistische Modelle

Dieses Kapitel umfasst die folgenden Themen:

- Probabilistische Modelle – Übersicht, 57
- Struktur des probalistischen Modells, 58
- Beispiel für eine Beschriftungsumwandlung, 58
- Beispiel für eine Parser-Transformation, 60
- Optionen des probabilistischen Modells, 60
- Referenzdaten in probabilistischen Modellen, 63
- Beschriftungsdaten in probabilistischen Modellen, 63
- Eigenschaften des probabilistischen Modells, 65
- Konfiguration des probabilistischen Modells, 65
- Kopieren und Einfügen, 72

Probabilistische Modelle – Übersicht

Bei einem probabilistischen Modell handelt es sich um ein Referenzdatenobjekt, das in einem Inhaltssatz erstellt wurde. Verwenden Sie ein probabilistisches Modell zur Analyse einer Datenzeichenfolge, die mehrere Datenwerte enthält. Ein probabilistisches Modell gibt die Informationstypen in jedem Wert in der Zeichenfolge an. Sie können einer Beschriftungs- oder Parser-Umwandlung ein probabilistisches Modell hinzufügen:

Verwenden Sie ein probabilistisches Modell in einer Beschriftungsumwandlung, um jedem Wert in einer Eingabezeichenfolge eine beschreibende Beschriftung hinzuzufügen. Die Beschriftungsumwandlung schreibt die Beschriftungen in einen einzelnen Ausgabeport. Verwenden Sie ein probabilistisches Modell in einer Parser-Umwandlung, um jeden Wert in einer Eingabezeichenfolge auf einen Port zu schreiben, der die Informationen im Wert darstellt. Die Parser-Umwandlung erstellt einen Ausgabeport für jeden Informationstyp.

Sie entwerfen und kompilieren ein probabilistisches Modell im Developer-Tool. Wenn Sie ein probabilistisches Modell definieren, fügen Sie eine Reihe von Datenzeilen zum Modell hinzu und weisen jedem Wert in jeder Zeile eine Beschriftung zu. Wenn Sie ein probabilistisches Modell kompilieren, erstellt das Developer-Tool Verknüpfungen zwischen den Datenwerten und den hinzugefügten Beschriftungen. Die Beschriftungs- und Parser-Umwandlung verwenden natürliche Sprachprozesse, um die Daten des probabilistischen Modells mit den Daten des Eingabeports zu vergleichen.

Natürliche Sprachprozesse verwenden folgende Methoden, um die Informationstypen in Datenwerten anzugeben:

- Natürliche Sprachprozesse können ähnliche Datenwerte erkennen und diesen Werten dieselbe Beschriftung zuweisen.
- Natürliche Sprachprozesse können einen Datenwert mit den angrenzenden Werten in der Zeichenfolge vergleichen. Natürliche Sprachprozesse analysieren die Reihenfolge der Werte, um die Verwendung jeder Zeichenfolge zu verstehen und die Informationstypen zu überprüfen, die von den Zeichenfolgen dargestellt werden.

Struktur des probalistischen Modells

Ein probabilistisches Modell enthält Zeilen mit Referenzdaten- und Beschriftungswerten. Die Referenzdatenwerte stellen die unterschiedlichen Werte dar, die in den Eingabedaten der Umwandlung angezeigt werden können. Die Beschriftungswerte geben die Typen von Informationen an, die die Eingabedaten enthalten sollen.

Ein probalistisches Modell enthält auch Kompilierungsdaten. Die Beschriftungs- und Parser-Umwandlung verwenden die Kompilierungsdaten, um die Ähnlichkeiten zwischen den Referenzdaten im Modell und den Eingabedaten der Umwandlung zu messen. Wenn Sie ein probabilistisches Modell kompilieren, erstellen oder aktualisieren Sie die Kompilierungsdaten.

Eine Datenzeile kann einen einzelnen Wert oder mehrere Werte enthalten. Jede Datenzeile kann eine andere Struktur aufweisen. Sie können verschiedenen Werten in einer Datenzeile dieselbe Beschriftung zuweisen. Alternativ können Sie identische Werte, die an verschiedenen Positionen in einer Zeile angezeigt werden, mit einer anderen Beschriftung versehen. Der Datenintegrationsdienst berücksichtigt die relativen Positionen der Werte in der Eingabezeichenfolge, wenn die Zuordnung ausgeführt wird. Weisen Sie jede Beschriftung mindestens einem Datenwert zu, bevor Sie das probabilistische Modell kompilieren.

Das Developer Tool schreibt die Referenzdatenwerte, die Beschriftungswerte und die Kompilierungsdaten in eine Datei in der Informatica-Verzeichnisstruktur. Der Dateiname wird im Objekt des probabilistischen Modells im Modellrepository gespeichert. Wenn Sie ein probabilistisches Modell speichern, schreiben Sie die aktuellen Referenzdatenwerte und die Beschriftungswerte in die Datei. Wenn Sie das Modell kompilieren, aktualisieren Sie die Kompilierungsdaten in der Datei. Sie können den Dateinamen aus den Modelleigenschaften im Developer Tool lesen.

Hinweis: Stellen Sie zur Optimierung der Funktionen des probabilistischen Modells sicher, dass jede Datenzeile mehrere Referenzdatenwerte enthält. Die Reihenfolge der Werte in jeder Zeile muss so weit wie möglich mit der Reihenfolge der Werte in den Eingabedaten der Umwandlung übereinstimmen. Wenn die Datenzeilen einzelne Referenzdatenwerte enthalten, kann weder die Beschriftungs- noch die Parser-Umwandlung natürliche Sprachprozesse während der probabilistischen Analyse anwenden.

Beispiel für eine Beschriftungsumwandlung

Angenommen, die Kundendatenbank eines Versicherungsunternehmens enthält mehrere Dateneingabefehler. Sie sind einer der Datenverantwortlichen in diesem Versicherungsunternehmen. Sie konfigurieren ein

Mapping mit einer Beschriftungsumwandlung, um die in jeder Spalte enthaltenen unterschiedlichen Datentypen zu bestimmen.

Die folgende Tabelle beschreibt Beispieldaten aus der Kundendatenbank:

Zeilen-ID	Feld 1	Feld 2	Feld 3
1	19132954	AIM SECURITIES	PETRIE TAYBRO
2	10110169	JASE TRAPANI	BANK OF NEW YORK
3	10111786	WANGER ASSET MANAGEMENT, LLP	JAN SEEDORF
4	10112299	FELIX LEVENGER	HARVARD MAGAZINE
5	10112036	DESCHÊNES & FILS LTÉE (QUEBEC)	RICHARD TREMBLAY
6	BERGER ASSOCIATES	10111101	DAREEN HULSMAN
7	19131385	EAGLE FINANCIAL GROUP INC	PATRICK MCKINNIE
8	LAKENYA PASKETT	WHITEHALL FINANCIAL GROUP	15954710

Wenn Sie das Mapping ausführen, vergleicht die Beschriftungsumwandlung die Eingabedaten mit den Referenzdaten des probabilistischen Modells. Die Beschriftungsumwandlung wählt für die Daten auf jedem Eingabeport eine Beschriftung aus. Die Umwandlung schreibt die Beschriftungen in einen Ausgabeport. Jede Ausgabezeile enthält einen Satz an Beschriftungen, der die Datenstruktur der entsprechenden Eingabezeile beschreibt.

Die folgende Tabelle beschreibt die Beschriftungen, die die Beschriftungsumwandlung zum Ausgabeport hinzufügt:

Zeilen-ID	Ausgabebeschriftungen
1	Nummer Unternehmen Kontakt
2	Nummer Kontakt Unternehmen
3	Nummer Unternehmen Kontakt
4	Nummer Kontakt Unternehmen
5	Nummer Unternehmen Kontakt
6	Unternehmen Nummer Kontakt
7	Unternehmen Nummer Kontakt
8	Kontakt Unternehmen Nummer

Beispiel für eine Parser-Transformation

Angenommen, ein Supermarkt speichert Produktbeschreibungen in einer einzigen Spalte einer Datenbanktabelle. Die Produktbeschreibungen enthalten mehrere Datenwerte, die unterschiedliche Informationstypen darstellen. Sie sind einer der Datenverantwortlichen in diesem Supermarkt. Sie möchten Spalten für die unterschiedlichen Informationstypen in den Produktbeschreibungen erstellen.

Sie konfigurieren ein Mapping mit einer Parser-Transformation, um die Datenwerte in die korrekten Felder einzusortieren.

Das folgende Datenfragment enthält die Produkt-Beschreibung für Orangensaft:

```
Sunnydream Orangensaft Ungesüßt 12 oz
```

Die folgende Tabelle beschreibt die Ausgabedaten, die die Parser-Umwandlung aus den Eingabedaten erstellt:

Produktname	Produkttyp	Produktdetails	Produktgröße
Sunnydream	Orangensaft	Ungesüßt	12 oz

Optionen des probabilistischen Modells

Wenn Sie ein probabilistisches Modell bearbeiten, können Sie hierzu die Daten- oder Beschriftungsansicht verwenden.

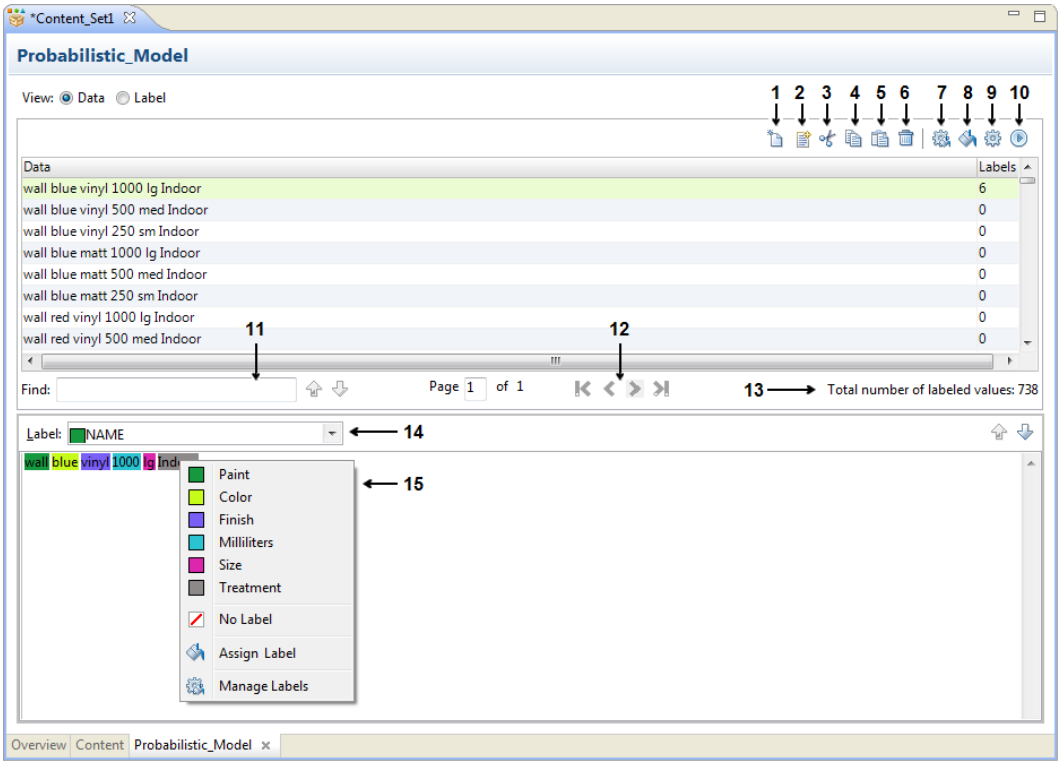
Verwenden Sie die Datenansicht zum Hinzufügen von Referenzdatenzeilen zum Modell sowie zum Zuweisen der Beschriftungen zu den Datenwerten in jeder Zeile. Verwenden Sie die Beschriftungsansicht zum Überprüfen detaillierter Informationen über die Verwendung der Beschriftungswerte im Modell. Sie können dem probabilistischen Modell in der Daten- und der Beschriftungsansicht Beschriftungen hinzufügen.

Datenansicht des probabilistischen Modells

In der Datenansicht werden die Referenzdatenzeilen im probabilistischen Modell sowie die Anzahl der Beschriftungswerte angezeigt, die jeder Zeile zugewiesen werden. In der Datenansicht wird darüber hinaus die Gesamtzahl der Beschriftungen angezeigt, die Sie den Werten im aktuellen Modell zugewiesen haben.

Wenn Sie eine Referenzdatenzeile auswählen, werden die Werte in der Zeile in einem Editor unterhalb des Felds „Suchen“ angezeigt. Klicken Sie zum Zuweisen einer Beschriftung zu einem Referenzdatenwert in einer Zeile mit der rechten Maustaste auf den Wert im Editor und wählen Sie einen Beschriftungswert aus.

Die folgende Abbildung zeigt die Optionen des probabilistischen Modells, die Sie bei Auswahl der Datenansicht verwenden können:



Die Datenansicht enthält die folgenden Optionen:

1. Zeile hinzufügen

   Fügt ein leere Datenzeile ein.

2. Daten anhängen.

   Importiert Daten aus einem Datenobjekt in das Modellrepository.

3. Ausschneiden

   Entfernt eine Datenzeile aus dem probabilistischen Modell und fügt sie in die Zwischenablage ein.

4. Kopieren

   Kopiert eine Datenzeile in die Zwischenablage.

5. Einfügen

   Fügt eine Zeile aus der Zwischenablage in das probabilistische Modell ein.

6. Löschen

   Löscht eine Datenzeile aus dem probabilistischen Modell.

7. Beschriftungen verwalten

   Öffnet das Dialogfeld **Beschriftungen verwalten**. Verwenden Sie das Dialogfeld, um dem probabilistischen Modell Beschriftungswerte hinzuzufügen bzw. aus diesem zu löschen.

8. Beschriftung zuweisen

   Weist mindestens einem von Ihnen ausgewählten Referenzdatenwert eine Beschriftung zu. Sie können die Option verwenden, um allen Instanzen eines Referenzdatenwerts im Modell eine Beschriftung zuzuweisen.

9. Eigenschaften bearbeiten

   Zeigt die Eigenschaften des probabilistischen Modells an.

10. Kompilieren

    Kompiliert das probabilistische Modell.

11. Feld „Suchen“

    Sucht nach Zeilen im Modell, die den von Ihnen eingegebenen Referenzdatenwert enthalten. Verwenden Sie den nach oben und nach unten weisenden Pfeil, um in die Zeilen zu wechseln, die den Wert enthalten.

12. Vor- und Zurück-Pfeile

    Springt vorwärts oder rückwärts durch die Datenzeilen des Modells.

13. Gesamtzahl der beschrifteten Werte

    Gibt die Anzahl der Referenzdatenwerte an, die eine Beschriftung verwenden.

14. Feld „Beschriftung“

    Zeigt einen Beschriftungswert an, den Sie auf den von Ihnen ausgewählten Referenzdatenwert anwenden können.

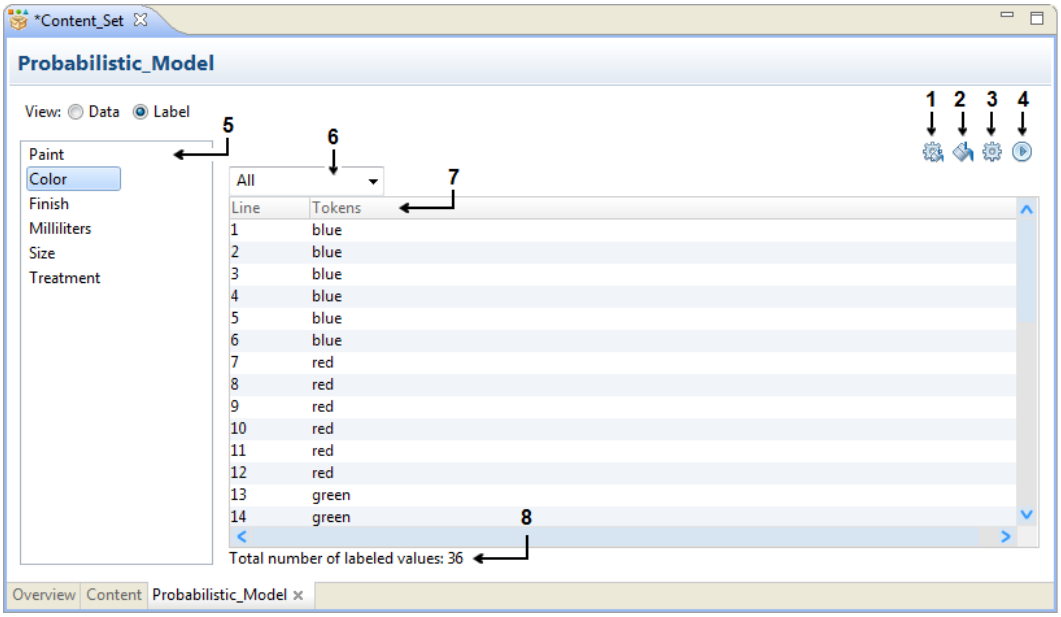
15. Menü „Beschriftung“

    Zeigt eine Liste mit Optionen an, die Sie zum Zuweisen einer Beschriftung zu einem oder mehreren Referenzdatenwerten verwenden können. Klicken Sie zum Öffnen des Menüs mit der rechten Maustaste auf einen Referenzdatenwert im Referenzdateneditor.

Beschriftungsansicht des probabilistischen Modells

In der Beschriftungsansicht werden die Beschriftungen aufgelistet, die im probabilistischen Modell definiert werden. Wenn Sie eine Beschriftung auswählen, werden in der Beschriftungsansicht die Datenwerte angezeigt, die Sie der Beschriftung in jeder Zeile zugewiesen haben.

Die folgende Abbildung zeigt die Optionen des probabilistischen Modells, die Sie bei Auswahl der Datenansicht verwenden können:



Die Beschriftungsansicht umfasst die folgenden Optionen:

1. Beschriftungen verwalten

   Öffnet das Dialogfeld **Beschriftungen verwalten**. Verwenden Sie das Dialogfeld, um dem probabilistischen Modell Beschriftungswerte hinzuzufügen bzw. aus diesem zu löschen.

2. Beschriftung zuweisen

   Weist mindestens einem von Ihnen ausgewählten Referenzdatenwert eine Beschriftung zu.

   Sie können einem einzelnen Datenwert eine Beschriftung zuweisen. Sie können aber auch in einem Vorgang mehreren Werten eine Beschriftung zuweisen.

3. Eigenschaften bearbeiten

   Zeigt die Eigenschaften des probabilistischen Modells an.

4. Kompilieren

   Kompiliert das probabilistische Modell.

5. Liste der Beschriftungswerte

   Listet die Beschriftungen auf, die Sie den Referenzdatenwerten im Modell zuweisen können.

6. Zuweisungsfilter

   Filtert die Liste mit den Referenzdatenwerten, die die von Ihnen ausgewählte Beschriftung verwendet. Basierend auf der Methode, die zum Zuweisen der Beschriftung zu den Datenwerten verwendet wurde, zeigen die Filteroptionen die Referenzdatenwerte an oder blenden sie aus.

   Wenn Sie einen Filter anwenden, steht die Gesamtanzahl der beschrifteten Werte in der Beschriftungsansicht für die Anzahl der Werte, die den Filterbedingungen entsprechen.

7. Spalte mit Referenzdatenwerten

   Listet die Referenzdatenwerte auf, die die aktuelle Beschriftung verwenden.

8. Gesamtzahl der beschrifteten Werte

   Gibt die Anzahl der Referenzdatenwerte an, die die aktuelle Beschriftung verwenden.

Referenzdaten in probabilistischen Modellen

Bei Referenzdatenwerten in einem probabilistischen Modell handelt es sich um die Typen von Eingabedaten, die Sie unter Umständen mit einer Umwandlung in einer Zuordnung verbinden.

Sie können Referenzdatenzeilen im Developer-Tool hinzufügen, bearbeiten und löschen. Sie können Daten aus der Zwischenablage einfügen und Daten aus einer Datenquelle importieren. Weisen Sie nach dem Hinzufügen der Referenzdatenwerte jedem Datenwert in jeder Zeile eine Beschriftung zu.

Beschriftungsdaten in probabilistischen Modellen

Bei den Beschriftungswerten in einem probabilistischen Modell handelt es sich um die Informationstypen, die unter Umständen in den Referenzdatenwerten enthalten sind. Weisen Sie beim Hinzufügen von Referenzdatenzeilen zu einem Modell jedem Wert in jeder Zeile eine Beschriftung zu. Die zum Modell

hinzugefügten Beschriftungen werden in der Beschriftungsansicht und in den Menüoptionen in der Datenansicht angezeigt.

Jedem beliebigen Referenzdatenwert im Modell kann jede beliebige Beschriftung zugewiesen werden. Wenn derselbe Wert verschiedene Bedeutungen in unterschiedlichen Referenzdatenzeilen aufweist, können Sie jedem Wert in jeder Zeile eine andere Beschriftung zuweisen.

Der Bereich der Beschriftungswerte kann mit dem Bereich der Eingabeports übereinstimmen, den die Beschriftungs- oder die Parser-Umwandlung während der probabilistischen Analyse liest. Das probabilistische Modell muss mindestens einen Beschriftungswert enthalten, den die Umwandlung auf die Datenwerte auf jedem Eingabeport anwenden kann.

In einem Warehouse werden beispielsweise Bestandsdaten in einer kommagetrennten Datei mit acht Spalten gespeichert. Sie entwerfen eine Zuordnung, die die Bestandsdaten in eine Datenbanktabelle parst. Sie erstellen ein probabilistisches Modell mit einem Beschriftungswert für jede Datenspalte. Wenn Sie die Zuordnung ausführen, schreibt die Parser-Umwandlung jeden Wert in den Eingabedaten in die entsprechende Spalte in der Zieltabelle.

In der folgenden Tabelle werden die Spalten mit Bestandsdaten sowie die Beschriftungswerte angezeigt, die Sie unter Umständen in einem probabilistischen Modell erstellen:

Name der Bestandsspalte	Beschriftungsname
Product_Name	Product_Name
Menge	Menge
Speicherort	Speicherort
Strichcode	Strichcode
SKU	Stock_Keeping_Unit
Arrival_Date	Arrival_Date
Cost_Price	Cost_Price

Hinweis: Sie können die Namen der Eingabespalten, aber auch andere Namen verwenden. Die Namen müssen nicht übereinstimmen.

Overflow-Beschriftung

Wenn eine Umwandlung keine Beschriftung auf einen Eingabedatenwert anwenden kann, behandelt die Umwandlung die Datenwerte als Überlaufdaten. Die Beschriftungsumwandlung weist jedem Datenwert, der nicht identifiziert werden kann, eine Überlaufbeschriftung zu. Die Parser-Umwandlung schreibt jeden Datenwert, der nicht identifiziert werden kann, auf einen Überlaufport.

In der folgenden Tabelle wird dargestellt, wie eine Parser-Umwandlung einen Überlaufport verwenden kann, um die Adressdatenelemente zu parsen, die von einem probabilistischen Modell nicht erkannt werden:

Eingabedaten	Street_Name-Port	Street_Descriptor-Port	Überlaufport
Park Place	Park	Place	Keine Überlaufdaten
Park Avenue	Park	Avenue	Keine Überlaufdaten

Eingabedaten	Street_Name-Port	Street_Descriptor-Port	Überlaufport
Madison Avenue	Madison	Avenue	Keine Überlaufdaten
Central Park	Central	Park	Keine Überlaufdaten
Washington Square Park	Washington	Square	Park
Madison Square Garden	Madison	Square	Garden

Die Parser-Umwandlung schreibt Werte auch auf einen Überlaufport, wenn die Anzahl der Eingabewerte die Anzahl der Beschriftungen im Modell übersteigt. Überprüfen Sie vor der Verwendung eines probabilistischen Modells in einer Umwandlung die Eingabedaten und stellen Sie sicher, dass das Modell die korrekte Anzahl an Beschriftungswerten enthält.

Eigenschaften des probabilistischen Modells

Sie können die allgemeinen und die erweiterten Eigenschaften für ein probabilistisches Modell überprüfen.

Wählen Sie zum Öffnen des Eigenschafteneditors die Option **Eigenschaften bearbeiten** in der Daten- oder Beschriftungsansicht aus.

In den allgemeinen Eigenschaften werden der Name des probabilistischen Modells, eine Beschreibung des Modells und der Name der Modelldatendatei angezeigt. In den erweiterten Eigenschaften werden die Berechnungseigenschaften angezeigt, die vom Developer-Tool zum Kompilieren des probabilistischen Modells verwendet werden.

Das Basiselement bei der Kompilierung von probabilistischen Modellen ist das *n-gram*. Ein N-Gramm ist eine Sequenz von Buchstaben, die anderen Buchstaben folgen oder vorausgehen und zusammen ein vollständiges Wort ergeben. Wenn ein Mapping ausgeführt wird, erzeugt die Beschriftungs- oder Parser-Umwandlung mehrere N-Gramme für jeden Wert in der Referenzdatenspalte des probabilistischen Modells. Die Transformation vergleicht die Eingabedatenwerte mit den Referenzdatenwerten und den N-Grammen. Die erweiterten Eigenschaften für ein probabilistisches Modell geben vor, wie das probabilistische Modell n-grams und andere Modellfunktionen verarbeitet.

Hinweis: Bei den Standardwerten in den erweiterten Eigenschaften handelt es sich um die bevorzugten Eigenschaften für die probabilistische Analyse und die Kompilierung des probabilistischen Modells. Wenn Sie eine erweiterte Eigenschaft bearbeiten, hat dies unter Umständen negative Auswirkungen auf die Genauigkeit der probabilistischen Analyse. Bearbeiten Sie die erweiterten Eigenschaften nur dann, wenn Sie sich im Klaren über die Auswirkungen der Änderungen sind.

Verwandte Themen:

- “Regeln und Richtlinien für probabilistische Modelle und Klassifizierermodelle” auf Seite 41

Konfiguration des probabilistischen Modells

Die Schritte zum Konfigurieren eines probabilistischen Modells beginnen mit dem Typ der auszuführenden Analyse. Verwenden Sie ein probabilistisches Modell in einer Beschriftungsumwandlung, um die

Informationstypen in jedem Wert in einer Eingabezeichenfolge anzugeben. Verwenden Sie ein probabilistisches Modell in einer Parser-Umwandlung, um die Datenwerte in einer Eingabezeichenfolge auf verschiedene Ausgabeports zu parsen.

Sie können dasselbe probabilistische Modell verwenden, um Daten zu beschriften und zu parsen. Wenn Sie das Modell in einer Beschriftungsumwandlung verwenden, erstellt die Umwandlung einen einzelnen Ausgabeport für jeden von Ihnen ausgewählten Eingabeport. Wenn Sie das Modell in einer Parser-Umwandlung verwenden, erstellt die Umwandlung einen Ausgabeport für jeden Eingabedatentyp, den sie erkennt.

Führen Sie zum Erstellen eines probabilistischen Modells die folgenden Aufgaben durch:

1. Geben Sie die Referenzdaten- und Beschriftungswerte an, die dem Modell hinzugefügt werden sollen.

   Sie können ein Fragment der zu analysierenden Daten verwenden. Erstellen Sie ein Datenobjekt in dem Modellrepository, das das Datenfragment liest.
2. Erstellen Sie einen Inhaltssatz und fügen Sie ihm ein probabilistisches Modell hinzu.
3. Fügen Sie dem Modell die Referenzdatenwerte hinzu.
4. Fügen Sie dem Modell die Beschriftungswerte hinzu.

   Sie können die Daten aus dem Datenobjekt im Modellrepository importieren. Sie können auch eine einzelne Referenzdatenzeile oder eine einzelne Beschriftung eingeben.

   Stellen Sie bei der Verwendung des probabilistischen Modells zum Parsen von Daten sicher, dass das Modell einen Beschriftungswert für jeden Ausgabeport enthält, der von der Umwandlung erstellt werden muss.
5. Weisen Sie eine Beschriftung zu, um auf jeden Datenwert in jeder Zeile zu verweisen.

   Sie können mehreren Referenzdatenwerten in einem einzelnen Vorgang eine Beschriftung zuweisen.
6. Kompilieren Sie das Modell.

Nach dem Kompilieren des probabilistischen Modells können Sie das Modell in einer Umwandlung verwenden.

Erstellen eines leeren probabilistischen Modells

Sie können ein probabilistisches Modellobjekt erstellen, das weder Referenz- noch Beschriftungsdaten enthält. Erstellen Sie das leere Modell und fügen Sie dem Modell Daten hinzu oder importieren Sie Daten in das Modell.

1. Öffnen oder erstellen Sie im Objekt-Explorer einen Inhaltssatz.
2. Wählen Sie die Ansicht „Inhalt“ aus.
3. Wählen Sie **Probabilistische Modelle** aus und klicken Sie auf **Hinzufügen**.

   Der Assistent für die probabilistischen Modelle wird aufgerufen.
4. Wählen Sie die Option **Probabilistisches Modell** aus.

   Klicken Sie auf **Weiter**.
5. Geben Sie einen Namen für das probabilistische Modell ein.

   Geben Sie optional eine Textbeschreibung des Modells ein.
6. Klicken Sie auf **Fertig stellen**.

Erstellen eines probabilistischen Modells aus einem Datenobjekt

Sie können ein Datenobjekt als Quelle für probabilistische Modelldaten verwenden.

Ein probabilistisches Modell arbeitet optimal, wenn Sie die Eingabedaten aus der Beschriftungs- oder Parser-Umwandlung als Quelle für die Modellreferenzdaten verwenden.

1. Öffnen oder erstellen Sie im Objekt-Explorer einen Inhaltssatz.
2. Wählen Sie die Ansicht **Inhalt** aus.
3. Wählen Sie **Probabilistische Modelle** aus und klicken Sie auf **Hinzufügen**.

   Der Assistent für die probabilistischen Modelle wird aufgerufen.
4. Wählen Sie die Option **Probabilistisches Modell aus Datenobjekten** aus.

   Klicken Sie auf **Weiter.**
5. Geben Sie einen Namen für das probabilistische Modell ein.

   Geben Sie optional eine Textbeschreibung des Modells ein.
6. Durchsuchen Sie das Modellrepository und wählen Sie das Datenobjekt aus, das die zu importierenden Daten enthält.

   Wählen Sie kein Datenobjekt der sozialen Medien aus.

   Klicken Sie auf **Weiter**.
7. Überprüfen Sie die Spalten im Datenobjekt und wählen Sie mindestens eine Spalte zum Hinzufügen zum Modell aus. Sie können Referenzdatenspalten und eine Beschriftungsspalte im selben Vorgang hinzufügen.
   - Wählen Sie zum Importieren einer Datenspalte als Referenzdaten den Namen der Spalte aus und klicken Sie auf **Daten**.

     Sie können mehrere Datenspalten auswählen. Das Developer Tool führt die Inhalte der von Ihnen ausgewählten Spalten in einer Spalte zusammen.
   - Wählen Sie zum Importieren einer Datenspalte als Beschriftungswerte den Namen der Spalte aus und klicken Sie auf **Beschriftung**.

   Wenn Sie Referenzdaten- und Beschriftungswerte importieren, weist das Developer Tool die Beschriftung in jeder Zeile zur Referenzdatenzeichenfolge in derselben Zeile zu. Sie können die Daten in der Vorschau anzeigen, bevor Sie die Spalten auswählen. Sie können die Beschriftungszuweisungen nach dem Erstellen des Modells ändern.

   Klicken Sie auf **Weiter**.
8. Wählen Sie die Anzahl der Zeilen aus, die Sie aus der Datenquelle importieren möchten.

   Das Developer Tool importiert standardmäßig alle Zeilen aus der Datenquelle. Wenn Sie eine Anzahl eingeben, zählt das Modell die Zeilen vom Anfang des Datensatzes.
9. Geben Sie die Delimiter für die zu importierenden Datenwerte an.

   Sie können verschiedene Delimiter für Referenzdaten- und Beschriftungswerte angeben. Als Standarddelimiter wird ein Leerzeichen verwendet.
10. Klicken Sie auf **Fertig stellen** und speichern Sie das Modell.

Überprüfen Sie nach dem Erstellen des probabilistischen Modells die Beschriftungszuweisungen und kompilieren Sie das Modell.

Anhängen von Daten aus einer Datenquelle an ein probabilistisches Modell

Sie können mehrere Zeilen mit Referenzdatenwerten und Beschriftungswerten in einem einzelnen Vorgang in ein probabilistisches Modell importieren.

1. Öffnen Sie den Inhaltssatz, der das probabilistische Modell enthält.
2. Wählen Sie den Modellnamen aus und klicken Sie auf **Bearbeiten**.
3. Klicken Sie auf **Daten anfügen**.

   Der Assistent für die probabilistischen Modelle wird aufgerufen.
4. Durchsuchen Sie das Modellrepository und wählen Sie das Datenobjekt aus, das die zu importierenden Daten enthält.

   Wählen Sie kein Datenobjekt der sozialen Medien aus.

   Klicken Sie auf **Weiter**.
5. Überprüfen Sie die Spalten im Datenobjekt und wählen Sie mindestens eine Spalte zum Hinzufügen zum Modell aus. Sie können Referenzdatenspalten und eine Beschriftungsspalte im selben Vorgang hinzufügen.
   - Wählen Sie zum Importieren einer Datenspalte als Referenzdaten den Namen der Spalte aus und klicken Sie auf **Daten**.

     Sie können mehrere Datenspalten auswählen. Das Developer Tool führt die Inhalte der von Ihnen ausgewählten Spalten in einer Spalte zusammen.
   - Wählen Sie zum Importieren einer Datenspalte als Beschriftungswerte den Namen der Spalte aus und klicken Sie auf **Beschriftung**.

   Wenn Sie Referenzdaten- und Beschriftungswerte importieren, weist das Developer Tool die Beschriftung in jeder Zeile zur Referenzdatenzeichenfolge in derselben Zeile zu. Sie können die Daten in der Vorschau anzeigen, bevor Sie die Spalten auswählen. Sie können die Beschriftungszuweisungen nach dem Erstellen des Modells ändern.

   Klicken Sie auf **Weiter**.
6. Wählen Sie die Anzahl der Zeilen aus, die Sie aus der Datenquelle importieren möchten.

   Das Developer Tool importiert standardmäßig alle Zeilen aus der Datenquelle. Wenn Sie eine Anzahl eingeben, zählt das Modell die Zeilen vom Anfang des Datensatzes.
7. Geben Sie die Delimiter für die zu importierenden Datenwerte an.

   Sie können verschiedene Delimiter für Referenzdaten- und Beschriftungswerte angeben. Als Standarddelimiter wird ein Leerzeichen verwendet.
8. Klicken Sie auf **Fertig stellen** und speichern Sie das Modell.

Hinzufügen einer Referenzdatenzeile zu einem probabilistischen Modell

Verwenden Sie die Datenansicht, um einem probabilistischen Modell eine leere Zeile hinzuzufügen.

1. Öffnen Sie den Inhaltssatz, der das Modell enthält.

   Wählen Sie den Modellnamen aus und klicken Sie auf **Bearbeiten**.
2. Wählen Sie die Datenansicht aus.
3. Klicken Sie zum Hinzufügen einer leeren Zeile zum Modell auf **Neu**.

4. Wählen Sie die hinzugefügte Zeile aus und geben Sie mindestens einen Referenzdatenwert in die Zeile ein.
5. Speichern Sie das probabilistische Modell.

Weisen Sie nach dem Speichern des Modells jedem Wert in der Zeile eine Beschriftung zu. Kompilieren Sie das Modell optional.

Hinzufügen einer Beschriftung zu einem probabilistischen Modell

Sie können eine einzelne Beschriftung zu einem probabilistischen Modell hinzufügen. Fügen Sie jedem Informationstyp, den die Modelldatenwerte darstellen, eine Beschriftung hinzu. Wenn Sie das probabilistische Modell in einer Parser-Umwandlung nutzen, fügen Sie eine Beschriftung für jeden Ausgabeport hinzu, den die Umwandlung erwartungsgemäß erzeugen wird.

1. Öffnen Sie den Inhaltssatz, der das Modell enthält.
2. Wählen Sie den Modellnamen aus und klicken Sie auf **Bearbeiten**.
3. Klicken Sie in der Daten- oder Beschriftungsansicht auf **Beschriftungen verwalten**.

   Das Dialogfeld **Beschriftungen verwalten** wird angezeigt.
4. Klicken Sie im Dialogfeld **Beschriftungen verwalten** auf **Neu**.

   In der ersten leeren Zeile des Dialogfelds wird eine Beschriftung angezeigt.
5. Bearbeiten Sie die Beschriftung. Aktualisieren Sie optional die Farbe für die Beschriftung.
6. Klicken Sie auf **OK**, um dem Modell die Beschriftung hinzuzufügen.
7. Speichern Sie das probabilistische Modell.

Weisen Sie die Beschriftung nach dem Hinzufügen mindestens einem Datenwert zu.

Zuweisen einer Beschriftung zu einem Referenzdatenwert

Sie können einem Datenwert in einer Referenzdatenzeile eine Beschriftung zuweisen.

Sie können demselben Datenwert verschiedene Beschriftungen zuweisen, wenn der Datenwert an verschiedenen Positionen in der Zeile oder in verschiedenen Zeilen angezeigt wird.

1. Öffnen Sie den Inhaltssatz, der das Modell enthält.
2. Wählen Sie den Modellnamen aus und klicken Sie auf **Bearbeiten**.
3. Wählen Sie die Datenansicht aus.
4. Suchen Sie nach einem Datenwert, der keine oder eine falsche Beschriftung aufweist. Datenwerte, die eine Beschriftung verwenden, sind farbkodiert.
5. Wählen Sie die Datenzeile aus, die den Datenwert enthält.

   Die Zeile wird im Editor angezeigt.
6. Klicken Sie mit der rechten Maustaste auf einen Datenwert im Editor und wählen Sie eine Beschriftung aus dem Kontextmenü aus.

   Das Developer-Tool weist dem Datenwert die Beschriftung zu.
7. Speichern Sie das probabilistische Modell.

Kompilieren Sie das probabilistische Modell nach dem Speichern optional.

Zuweisen einer Beschriftung zu mehreren Datenwerten

Sie können mehreren Referenzdatenwerten in einem einzelnen Vorgang eine Beschriftung zuweisen.

1. Öffnen Sie den Inhaltssatz, der das Modell enthält.
2. Wählen Sie den Modellnamen aus und klicken Sie auf **Bearbeiten**.
3. Klicken Sie auf **Beschriftung zuweisen**.

   Das Dialogfeld **Mehreren Werten eine Beschriftung zuweisen** wird geöffnet.
4. Geben Sie mindestens ein Zeichen im Feld „Suchen“ ein.

   Sie können Platzhalterzeichen im Feld „Suchen“ eingeben.
5. Wählen Sie optional weitere Suchkriterien aus.

   Sie können die folgenden Optionen aktivieren oder deaktivieren:

   - Groß-/Kleinschreibung beachten.

     Gibt an, dass beim Suchvorgang die Groß-/Kleinschreibung berücksichtigt wird. Verwenden Sie keine Platzhalterzeichen mit der Option.
   - Mit ganzer Zeichenfolge vergleichen. Gibt an, dass nach einer vollständigen Übereinstimmung zwischen den Zeichen in den Referenzdatenwerten und den von Ihnen eingegebenen Zeichen gesucht wird. Verwenden Sie keine Platzhalterzeichen mit der Option.
   - Beschriftete Werte ignorieren.

     Gibt an, dass beim Suchvorgang alle Referenzdatenwerte übersprungen werden, die eine Beschriftung verwenden.
6. Wählen Sie eine Beschriftung zum Zuweisen zu den Referenzdatenwerten aus, die mit den Suchkriterien übereinstimmen.

   Sie können auch die Option **Keine Beschriftung** auswählen. Wählen Sie die Option aus, um die Beschriftung aus den Referenzdatenwerten zu löschen, die die von Ihnen eingegebenen Zeichen enthält.
7. Klicken Sie auf **Start**.

   Das Developer Tool weist die Beschriftung allen Referenzdatenwerten zu, die den von Ihnen definierten Suchkriterien entsprechen.

   Hinweis: Verwenden Sie in der Beschriftungsansicht den Filter **Von Bulk zugewiesen**, um die Referenzdatenwerte anzuzeigen, die Sie in einem einzelnen Vorgang beschriftet haben.

Löschen von Zeilen aus einem probabilistischen Modell

Sie können eine oder mehrere Referenzdatenzeilen in einer einzelnen Aktion aus einem probabilistischen Modell löschen.

1. Öffnen Sie den Inhaltssatz, der das Modell enthält.
2. Wählen Sie den Modellnamen aus und klicken Sie auf **Bearbeiten**.
3. Wählen Sie in der Datenansicht mindestens eine Referenzdatenzeile aus.
4. Klicken Sie auf **Löschen**.

   Das Developer Tool entfernt die von Ihnen ausgewählten Zeilen aus dem Klassifizierermodell.

Verwenden Sie die Tastenkombination Strg + Z, um den Vorgang rückgängig zu machen.

Löschen einer Beschriftung aus einem probabilistischen Modell

Beim Löschen eines Beschriftungswerts aus einem Modell verbleiben Referenzdatenwerte, die die Beschriftung verwendet hat, im Modell. Weisen Sie jedem Referenzdatenwert einen anderen Beschriftungswert zu.

1. Öffnen Sie den Inhaltssatz, der das Modell enthält.
2. Wählen Sie den Modellnamen aus und klicken Sie auf **Bearbeiten**.
3. Klicken Sie in der Daten- oder Beschriftungsansicht auf **Beschriftungen verwalten**.
4. Wählen Sie im Dialogfeld **Beschriftungen verwalten** einen Beschriftungswert aus.
5. Klicken Sie auf **Löschen**.
6. Klicken Sie auf **OK**, um die Beschriftung zu löschen.
7. Speichern Sie das probabilistische Modell.

Hinweis: Eine Beschriftung ist ein Strukturelement in einem probabilistischen Modell. Wenn Sie nach dem Hinzufügen des Modells zu einer Umwandlung eine Beschriftung hinzufügen oder entfernen, wird der Vorgang, der das Modell verwendet, ungültig gemacht. Löschen und erstellen Sie den Umwandlungsvorgang neu, um das aktualisierte Modell zu verwenden.

Kompilieren des probabilistischen Modells

Wenn Sie die Daten oder die Beschriftungszuweisungen in einem probabilistischen Modell aktualisieren, können Sie das Modell kompilieren. Kompilieren Sie das Modell, um die Modelllogik mit den Verbindungen zwischen den aktuellen Referenzdatenwerten und den aktuellen Beschriftungswerten zu aktualisieren.

Stellen Sie vor dem Kompilieren des probabilistischen Modells sicher, dass jeder Beschriftungswert mindestens einen Referenzdatenwert angibt.

- Um das Modell zu kompilieren, öffnen Sie das Modell im Developer-Tool und klicken Sie auf **Kompilieren**.

Suchen nach Datenzeilen in einem probabilistischen Modell

Verwenden Sie die Datenansicht, um nach den Referenzdatenzeilen zu suchen, die einen von Ihnen eingegebenen Wert enthalten.

1. Öffnen Sie den Inhaltssatz, der das probabilistische Modell enthält.
2. Wählen Sie den Modellnamen aus und klicken Sie auf **Bearbeiten**.
3. Wählen Sie die Datenansicht aus.
4. Geben Sie mindestens ein Zeichen im Feld **Suchen** ein.

   In der Datenansicht wird die erste Zeile in dem Modell angezeigt, das den eingegebenen Wert enthält.
5. Verwenden Sie den nach oben oder nach unten weisenden Pfeil, um in andere Zeilen zu wechseln, die den Wert enthalten.

Filtern von Referenzdatenwerten nach Beschriftungszuweisungen

Verwenden Sie die Beschriftungsansicht, um Referenzdatenwerte zu finden, die eine von Ihnen angegebene Beschriftung verwenden. Filtern Sie die Ergebnisse basierend auf der Methode, die Sie für die Zuweisung der Beschriftung verwendet haben.

1. Öffnen Sie den Inhaltssatz, der das probabilistische Modell enthält.
2. Wählen Sie den Modellnamen aus und klicken Sie auf **Bearbeiten**.

3. Wählen Sie in der Beschriftungsansicht einen Beschriftungswert aus.

   Im probabilistischen Modell wird eine Liste der Referenzdatenwerte angezeigt, die die Beschriftung verwenden. Im Modell wird darüber hinaus die Zahl der Datenwerte angezeigt, die die Beschriftung verwenden.

4. Wenden Sie einen Filter auf die Liste mit den Referenzdatenwerten an, die dieses Beschriftung verwenden.

   Wählen Sie einen der folgenden Filter aus:

   - Alle. Zeigt die Referenzdatenwerte an, die die Beschriftung verwenden. Alle ist die Standardoption.
   - Vom Benutzer zugewiesen. Zeigt alle Referenzdatenwerte an, die Sie bei der Zuweisung der Beschriftung einzeln ausgewählt haben.
   - Von Bulk zugewiesen. Zeigt die Referenzdatenwerte an, denen Sie die Beschriftung im Rahmen des Massenzuweisungsvorgangs zugewiesen haben.

   Im probabilistischen Modell werden die Referenzdatenwerte angezeigt, die den Filterbedingungen entsprechen.

Suchen nach ungenutzten Beschriftungswerten

Verwenden Sie die Beschriftungsansicht, um nach einem beliebigen Beschriftungswert zu suchen, der keinem Referenzdatenwert zugewiesen wurde. Sie müssen jede Beschriftung mindestens einem Referenzdatenwert zuweisen.

1. Öffnen Sie den Inhaltssatz, der das probabilistische Modell enthält.
2. Wählen Sie den Modellnamen aus und klicken Sie auf **Bearbeiten**.
3. Wählen Sie in der Beschriftungsansicht einen Beschriftungswert aus.

   Im probabilistischen Modell wird eine Liste der Referenzdatenwerte angezeigt, die die Beschriftung verwenden. Im Modell wird darüber hinaus die Gesamtzahl der Datenwerte angezeigt, die die Beschriftung verwenden.

   Wenn sich die Gesamtzahl der Datenwerte auf null beläuft, haben Sie die Beschriftung keinem Referenzdatenwert im probabilistischen Modell zugewiesen.

Kopieren und Einfügen

Sie können ein probabilistisches Modell aus einem Inhaltssatz in einen anderen in einem Modellrepository kopieren. Kopieren Sie ein probabilistisches Modell, um Ressourcen mit anderen Benutzern des Developer-Tools zu teilen.

Sie können ein Modell in einen anderen Inhaltssatz kopieren oder in den aktuellen Inhaltssatz importieren. Sie können mehrere Modelle aus mehreren Inhaltssätzen im Repository in einem einzigen Vorgang importieren.

Wenn Sie ein Modell kopieren, erstellt der Content-Management-Dienst eine Kopie der Modelldatendatei auf dem Computer, der die Informatica-Dienste hostet. Jedes Modell verwendet eine andere Datendatei.

Kopieren eines probabilistischen Modells in einen anderen Inhaltssatz

Sie können ein probabilistisches Modell aus einem Inhaltssatz in einen anderen in einem Modell-Repository kopieren. Beim Kopieren eines probabilistischen Modells geben Sie das Modellobjekt sowie die Quell- und Zielinhaltssätze an.

1. Öffnen Sie den Inhaltssatz, der das probabilistische Modell enthält.
2. Wählen Sie ein probabilistisches Modell und klicken Sie auf **Kopieren nach**.
3. Durchsuchen Sie das Modell-Repository und wählen Sie einen Inhaltssatz.

   Sie können das probabilistische Modell in einen Inhaltssatz des aktuellen oder eines anderen Projekts kopieren.
4. Klicken Sie auf **OK**.

   Das Developer-Tool kopiert das probabilistische Modell in den gewählten Inhaltssatz.

Importieren eines probabilistischen Modells aus einem anderen Inhaltssatz

Sie können ein probabilistisches Modell aus einem Inhaltssatz in einen anderen in einem Modell-Repository importieren. Beim Importieren eines probabilistischen Modells geben Sie ein oder mehrere Modellobjekte sowie die Quell- und Zielinhaltssätze an.

1. Öffnen Sie den Inhaltssatz, der das probabilistische Modell enthalten soll.
2. Wählen Sie ein probabilistisches Modell und klicken Sie auf **Einfügen aus**.
3. Durchsuchen Sie das Modell-Repository und wählen Sie ein probabilistisches Modell.

   Sie können das probabilistische Modell aus einem Inhaltssatz des aktuellen oder eines anderen Projekts einfügen.
4. Klicken Sie auf **OK**.

   Das Developer-Tool fügt das probabilistische Modell in den aktuellen Inhaltssatz ein.

Kopieren von Referenzdatenzeilen in die Zwischenablage

Sie können mindestens eine Referenzdatenzeile aus einem probabilistischen Modell in die Zwischenablage kopieren. Sie können die Zeilen in ein anderes probabilistisches Modell einfügen.

1. Öffnen Sie den Inhaltssatz, der das probabilistische Modell enthält.
2. Wählen Sie den Modellnamen aus und klicken Sie auf **Bearbeiten**.
3. Wählen Sie in der Datenansicht mindestens eine Referenzdatenzeile aus.
4. Verwenden Sie die Tastenkombination Strg + C, um die Zeilen in die Zwischenablage zu kopieren.

   Bei diesem Vorgang werden die Referenzdaten und die Beschriftungswerte kopiert, die den Referenzdaten zugewiesen wurden.

Sie können die Tastenkombination Strg + V verwenden, um die Zeilen in einen Texteditor oder in die Datenansicht eines anderen probabilistischen Modells zu kopieren.

ANHANG A

Referenzdaten und Informatica Big Data Management

- Referenzdaten und Informatica Big Data Management - Übersicht, 74

Referenzdaten und Informatica Big Data Management - Übersicht

Informatica Big Data Management® ist eine Big-Data-Lösung, die eine Informatica-Domäne und Clientanwendungen mit einem Hadoop-Cluster verbindet. Sie können Mappings vom Developer Tool bis zum Cluster pushen und die Mappings auf den Knoten im Cluster ausführen.

Wenn Sie ein Mapping pushen, die Umwandlungen enthält, welche Referenzdaten lesen, kann der Pushdown-Vorgang alle Referenzdaten kopieren, die von den Umwandlungen verwendet werden. Der Pushdown-Vorgang kopiert Referenztabellendaten, Inhaltssatzdaten und Identitätspopulationsdaten auf den Cluster. Nachdem das Mapping ausgeführt wurde, löscht der Cluster die Referenzdaten, die der Pushdown-Vorgang mit dem Mapping kopiert hat.

Hinweis: Der Pushdown-Vorgang kopiert keine Referenzdaten für die Adressvalidierung. Wenn Sie ein Mapping pushen, das die Adressvalidierung durchführt, müssen Sie die Referenzdaten-Dateien für die Adressvalidierung auf jedem DataNode installieren, der das Mapping ausführt. Der Cluster löscht die Referenzdaten-Dateien für die Adressvalidierung nach der Ausführung des Mapping der Adressvalidierung nicht.

Referenzdaten für die Adressvalidierung

Wenn Sie ein Adressvalidierungs-Mapping in einer Hadoop-Umgebung ausführen, müssen sich die Adressenreferenzdaten-Dateien auf jedem DataNode befinden, auf dem das Mapping ausgeführt wird. Informatica Big Data Management wird mit einem Shell-Skript installiert, mit dem Sie die Dateien auf den DataNodes installieren können.

Verwenden Sie das Shell-Skript, um die Adressenreferenzdaten-Dateien in einem Arbeitsgang auf den DataNodes zu installieren. Das Skript liest eine Datei mit den Namen oder IP-Adressen der Knoten. Das Skript kopiert die Adressenreferenzdaten-Dateien auf jeden Knoten, den die Datei identifiziert.

Der Name des Skripts lautet `copyRefDataToComputeNodes.sh`.

Suchen Sie das Skript in folgendem Verzeichnis in der Installation von Informatica Big Data Management:

`<Informatica-Installationsverzeichnis>/tools/dq/av`

In der folgenden Tabelle werden die Optionen beschrieben, die das Skript verwendet:

Option	Beschreibung
-n	Die Datei mit der Liste der Namen oder IP-Adressen der DataNodes im Hadoop-Cluster. Geben Sie den Namen oder die IP-Adresse für jeden Knoten in einer eigenen Zeile in der Datei ein. Standardmäßig liest das Skript die Datei aus dem Verzeichnis `$BASEDIR/HadoopDataNodes`, wobei `$BASEDIR` der Speicherort des Shell-Skripts ist.
-p	Eine Aufforderung zur Bestätigung, dass Sie die Adressenreferenzdaten-Dateien installieren möchten. Standardmäßig zeigt das Skript eine Aufforderung an, um zu bestätigen, dass Sie die Dateien aus dem Quellverzeichnis in die Zielverzeichnisse auf den DataNodes kopieren möchten. Wenn Sie das Shell-Skript nach einem Zeitplan ausführen, können Sie die Aufforderung deaktivieren. Der Standardwert für die Option ist J. Um die Aufforderung zu deaktivieren, legen Sie den Wert auf N fest.
-s	Das Quellverzeichnis für die Adressenreferenzdaten-Dateien, die das Skript auf die Knoten kopiert. Standardmäßig liest das Skript die Dateien aus dem Verzeichnis `/reference_data` auf dem lokalen Computer. **Hinweis:** Adressenreferenzdaten-Dateien verwenden die Dateierweiterung „.MD“. Das Quellverzeichnis muss ausschließlich die Adressenreferenzdaten-Dateien enthalten.
-t	Das Verzeichnis auf jedem Knoten, auf den das Skript die Adressenreferenzdaten-Dateien kopiert. Standardmäßig kopiert das Skript die Dateien in das Verzeichnis `/reference_data` auf jedem Knoten.
-u	Der Benutzername des Benutzers, der das Skript ausführt. Der Benutzer muss über einen sicheren Shell-Zugriff auf die Knoten ohne Passwort verfügen.

Installieren der Adressenreferenzdaten-Dateien

Führen Sie das Shell-Skript `copyRefDataToComputeNodes.sh` aus, um Adressenreferenzdaten-Dateien auf den DataNodes in einem Hadoop-Cluster zu installieren. Oder definieren Sie einen Auftrag, um das Shell-Skript in den von Ihnen angegebenen Zeitabständen in einer Auftragsplaner-Anwendung auszuführen.

Überprüfen Sie vor dem Ausführen des Skripts oder Definieren des Auftrags die zur Auswahl stehenden Werte, die Sie für das Skript angeben. Sie können die Standardwerte annehmen oder die Werte aktualisieren.

Installieren der Adressenreferenzdaten-Dateien an der Eingabeaufforderung

Führen Sie die folgenden Schritte aus, um die Dateien an der Eingabeaufforderung zu installieren:

1. Öffnen Sie an der Eingabeaufforderung das folgende Verzeichnis:
   `<Informatica-Installationsverzeichnis>/tools/dq/av`
2. Führen Sie `copyRefDataToComputeNodes.sh` aus.
   Geben Sie optional einen oder mehrere Werte für die Skriptoptionen ein. Wenn Sie keinen Wert für eine Option eingeben, wird das Skript mit dem Standardwert für die Option ausgeführt.

   Standardmäßig werden Sie vom Skript aufgefordert, die Installation der Dateien zu bestätigen. Um die Dateien zu installieren, geben Sie „J“ ein.

Installieren der Adressenreferenzdaten-Dateien mit einem geplanten Auftrag

Sie können einen Auftrag definieren, um das Shell-Skript in bestimmten Zeitabständen auszuführen, die Sie angeben. Fügen Sie den Auftrag zu einer Auftragsplaner-Anwendung hinzu. Wenn Sie einen Auftrag zum Installieren der Dateien definieren, müssen Sie die Aufforderung zum Bestätigen der Installation deaktivieren.

Um die Aufforderung zu deaktivieren, legen Sie die folgende Option für das Shell-Skript fest:

```
-p n
```

INDEX

A

Analyst Tool
  Referenzdatenwerte suchen und ersetzen 26
Audit-Trail-Ereignisse anzeigen
  Referenztabellen 28

B

Berechtigungen
  Content-Managementdienst 13
Big Data Management
  Installationsskript für Adressenreferenzdaten 74
  Installieren von Adressenreferenzdaten 75
  Referenzdatenanforderungen 74

C

Content-Managementdienst
  Referenztabellen-Berechtigungen 13

E

Erstellen einer Referenztabelle aus Profilspaltendaten
  Referenztabellen 18
Erstellen einer Referenztabelle aus Spaltenmustern
  Referenztabellen 19
Exportieren einer Referenztabelle
  Referenztabellen 26

H

Hadoop-Umgebung
  Installationsskript für Adressenreferenzdaten 74
  Installieren von Adressenreferenzdaten 75
  Referenzdatenanforderungen 74

I

Importieren einer Referenztabelle
  Referenztabellen 21
Inhaltssätze
  Klassifizierermodelle 38
  Mustersätze 38
  Probabilistische Modelle 38
  reguläre Ausdrücke 39
  Tokensätze 40
  Versionskontrolle 14, 23, 31
  Zeichensätze 37

K

Klassifizierermodelle
  in Inhaltssätzen 38
  Regeln und Richtlinien 41

M

Manuelles Erstellen einer Referenztabelle
  Referenztabellen 17
Mustersätze 38

N

nicht verwaltete Referenztabellen
  Aktivieren und Deaktivieren von Bearbeitungen 27
  Definition 12
  Synchronisierung mit dem Modellrepository 13

P

Probabilistische Modelle
  in Inhaltssätzen 38
  Regeln und Richtlinien 41

R

Referenztabellen
  Aktualisieren im Analyst Tool 27
  Audit-Trail-Tabellen anzeigen 28
  Berechtigungen 13
  Content-Managementdienst 12
  Das Analyst-Tool – Übersicht 15
  Developer tool – Übersicht 32
  Eigenschaften im Analyst Tool 16
  Eigenschaften im Developer-Tool 33
  Erstellen einer Referenztabelle aus Profilspalten 18
  Erstellen einer Referenztabelle aus Spaltenmustern 19
  Exportieren einer Referenztabelle 26
  im musterbasierten Parsing-Modus 12
  Importieren einer Referenztabelle 21
  Manuelles Erstellen einer Referenztabelle 17
  nicht verwaltete Referenztabellen 12
  Referenzdaten-Warehouse 12
  Suchen und Ersetzen von Werten im Analyst Tool 26
  Versionskontrolle 14, 23, 31
  Verwalten von Spalten 24
  Verwalten von Zeilen 25
  verwaltet und nicht verwaltet 12
  verwaltete Referenztabellen 12
reguläre Ausdrücke 39

T

Tokensätze 40

V

Versionskontrolle
  Inhaltssätze 14, 31
  Referenztabellen 14
  Referenztabellen im Analyst Tool 23
  Referenztabellen im Developer Tool 31
Verwalten von Spalten
  Referenztabellen 24
Verwalten von Zeilen
  Referenztabellen 25
verwaltete Referenztabellen 12

Z

Zeichensätze 37