



Informatica®
10.1

参照データガイド

本ソフトウェアおよびマニュアルには、Informatica LLC の所有権下にある情報が収められています。これらは使用および開示の制限等を定めた使用許諾契約のもとに提供され、著作権法により保護されています。当該ソフトウェアのリバースエンジニアリングは禁じられています。本マニュアルのいかなる部分も、いかなる手段（電子的複製、写真複製、録音など）によっても、Informatica LLC の事前の承諾なしに複製または転載することは禁じられています。このソフトウェアは、米国および/または国際的な特許、およびその他の出願中の特許によって保護されています。

合衆国政府によるソフトウェアの使用、複製または開示は、DFARS 227.7202-1 (a) および 227.7702-3 (a) (1995 年)、DFARS 252.227-7013(C) (1) (ii) (1988 年 10 月)、FAR 12.212 (a) (1995 年)、FAR 52.227-19、または FAR 52.227-14 (ALT III) に記載されているとおり、当該ソフトウェア使用許諾契約に定められた制限によって規制されます。

本製品または本書の情報は、予告なしに変更されることがあります。お客様が本製品または本書内に問題を発見された場合は、書面に当社までお知らせください。

Informatica、Informatica Platform、Informatica Data Services、PowerCenter、PowerCenterRT、PowerCenter Connect、PowerCenter Data Analyzer、PowerExchange、PowerMart、Metadata Manager、Informatica Data Quality、Informatica Data Explorer、Informatica B2B Data Transformation、Informatica B2B Data Exchange、Informatica On Demand、Informatica Identity Resolution、Informatica Application Information Lifecycle Management、Informatica Complex Event Processing、Ultra Messaging、Informatica Master Data Management、および Live Data Map は、Informatica LLC の米国および世界中の管轄地での商標または登録商標です。その他のすべての企業名および製品名は、それぞれの企業の商標または登録商標です。

本ソフトウェアまたはドキュメントの一部は、次のサードパーティが有する著作権に従います（ただし、これらに限定されません）。Copyright DataDirect Technologies. All rights reserved. Copyright (C) Sun Microsystems. All rights reserved. Copyright (C) RSA Security Inc. All rights reserved. Copyright (C) Ordinal Technology Corp. All rights reserved. Copyright (C) Aandacht c.v. All rights reserved. Copyright Genivia, Inc. All rights reserved. Copyright Isomorphic Software. All rights reserved. Copyright (C) Meta Integration Technology, Inc. All rights reserved. Copyright (C) Intalio. All rights reserved. Copyright (C) Oracle. All rights reserved. Copyright (C) Adobe Systems Incorporated. All rights reserved. Copyright (C) DataArt, Inc. All rights reserved. Copyright (C) ComponentSource. All rights reserved. Copyright (C) Microsoft Corporation. All rights reserved. Copyright (C) Rogue Wave Software, Inc. All rights reserved. Copyright (C) Teradata Corporation. All rights reserved. Copyright (C) Yahoo! Inc. All rights reserved. Copyright (C) Glyph & Cog, LLC. All rights reserved. Copyright (C) Thinkmap, Inc. All rights reserved. Copyright (C) Clearpace Software Limited. All rights reserved. Copyright (C) Information Builders, Inc. All rights reserved. Copyright (C) OSS Nokalva, Inc. All rights reserved. Copyright Edifecs, Inc. All rights reserved. Copyright Cleo Communications, Inc. All rights reserved. Copyright (C) International Organization for Standardization 1986. All rights reserved. Copyright (C) ej-technologies GmbH. All rights reserved. Copyright (C) Jaspersoft Corporation. All rights reserved. Copyright (C) International Business Machines Corporation. All rights reserved. Copyright (C) yWorks GmbH. All rights reserved. Copyright (C) Lucent Technologies. All rights reserved. Copyright (C) University of Toronto. All rights reserved. Copyright (C) Daniel Veillard. All rights reserved. Copyright (C) Unicode, Inc. Copyright IBM Corp. All rights reserved. Copyright (C) MicroQuill Software Publishing, Inc. All rights reserved. Copyright (C) PassMark Software Pty Ltd. All rights reserved. Copyright (C) LogiXML, Inc. All rights reserved. Copyright (C) 2003-2010 Lorenzi Davide, All rights reserved. Copyright (C) Red Hat, Inc. All rights reserved. Copyright (C) The Board of Trustees of the Leland Stanford Junior University. All rights reserved. Copyright (C) EMC Corporation. All rights reserved. Copyright (C) Flexera Software. All rights reserved. Copyright (C) Jinfonet Software. All rights reserved. Copyright (C) Apple Inc. All rights reserved. Copyright (C) Telerik Inc. All rights reserved. Copyright (C) BEA Systems. All rights reserved. Copyright (C) PDFlib GmbH. All rights reserved. Copyright (C) Orientation in Objects GmbH. All rights reserved. Copyright (C) Tanuki Software, Ltd. All rights reserved. Copyright (C) Ricebridge. All rights reserved. Copyright (C) Sencha, Inc. All rights reserved. Copyright (C) Scalable Systems, Inc. All rights reserved. Copyright (C) jQWidgets. All rights reserved. Copyright (C) Tableau Software, Inc. All rights reserved. Copyright (C) MaxMind, Inc. All rights reserved. Copyright (C) TMate Software s.r.o. All rights reserved. Copyright (C) MapR Technologies Inc. All rights reserved. Copyright (C) Amazon Corporate LLC. All rights reserved. Copyright (C) Highsoft. All rights reserved. Copyright (C) Python Software Foundation. All rights reserved. Copyright (C) BeOpen.com. All rights reserved. Copyright (C) CNRI. All rights reserved.

本製品には、Apache Software Foundation (<http://www.apache.org/>) によって開発されたソフトウェア、およびさまざまなバージョンの Apache License（まとめて「License」と呼んでいます）の下に許諾された他のソフトウェアが含まれます。これらのライセンスのコピーは、<http://www.apache.org/licenses/> で入手できます。適用法にて要求されないか書面に合意されない限り、ライセンスの下に配布されるソフトウェアは「現状のまま」で配布され、明示的あるいは黙示的かを問わず、いかなる種類の保証や条件も付帯することはありません。ライセンス下での許諾および制限を定める具体的文言については、ライセンスを参照してください。

本製品には、Mozilla (<http://www.mozilla.org/>) によって開発されたソフトウェア、ソフトウェア copyright The JBoss Group, LLC、コンテンツの無断複製・転載を禁じます、ソフトウェア copyright, Red Hat Middleware, LLC、コンテンツの無断複製・転載を禁じます、Copyright (C) 1999-2006 by Bruno Lowagie and Paulo Soares および GNU Lesser General Public License Agreement (<http://www.gnu.org/licenses/lgpl.html> を参照) に基づいて許諾されたその他のソフトウェアが含まれています。資料は、Informatica が無料で提供しており、一切の保証を伴わない「現状渡し」で提供されるものとし、Informatica LLC は市場性および特定の目的の適合性の黙示の保証などを含めて、一切の明示的及び黙示的保証の責任を負いません。

製品には、ワシントン大学、カリフォルニア大学アーバイン校、およびバンダービルト大学の Douglas C. Schmidt および同氏のリサーチグループが著作権を持つ ACE (TM) および TAO (TM) ソフトウェアが含まれています。Copyright (C) 1993-2006, All rights reserved.

本製品には、OpenSSL Toolkit を使用するために OpenSSL Project が開発したソフトウェア（copyright The OpenSSL Project. コンテンツの無断複製・転載を禁じます）が含まれています。また、このソフトウェアの再配布は、<http://www.openssl.org> および <http://www.openssl.org/source/license.html> にある使用条件に従います。

本製品には、Curl ソフトウェア Copyright 1996-2013, Daniel Stenberg, <daniel@haxx.se>が含まれます。All Rights Reserved. 本ソフトウェアに関する許諾および制限は、<http://curl.haxx.se/docs/copyright.html> にある使用条件に従います。すべてのコピーに上記の著作権情報とこの許諾情報が記載されている場合、目的に応じて、本ソフトウェアの使用、コピー、変更、ならびに配布が有償または無償で許可されます。

本製品には、MetaStuff, Ltd. のソフトウェアが含まれます。Copyright 2001-2005 (C) MetaStuff, Ltd. All Rights Reserved. 本ソフトウェアに関する許諾および制限は、<http://www.dom4j.org/license.html> にある使用条件に従います。

製品には、The Dojo Foundation のソフトウェアが含まれます。Copyright (C) 2004-2007. All Rights Reserved. 本ソフトウェアに関する許諾および制限は、<http://dojotoolkit.org/license> にある使用条件に従います。

本製品には、ICU ソフトウェアおよび他のソフトウェアが含まれます。Copyright International Business Machines Corporation. All rights reserved. 本ソフトウェアに関する許諾および制限は、<http://source.icu-project.org/repos/icu/icu/trunk/license.html> にある使用条件に従います。

本製品には、Per Bothner のソフトウェアが含まれます。Copyright (C) 1996-2006. All rights reserved. お客様がこのようなソフトウェアを使用するための権利は、ライセンスで規定されています。<http://www.gnu.org/software/kawa/Software-License.html> を参照してください。

本製品には、OSSP UUID ソフトウェアが含まれます。Copyright (C) 2002 Ralf S. Engelschall, Copyright (C) 2002 The OSSP Project Copyright (C) 2002 Cable & Wireless Deutschland. 本ソフトウェアに関する許諾および制限は、<http://www.opensource.org/licenses/mit-license.php> にある使用条件に従います。

本製品には、Boost (<http://www.boost.org/>) によって開発されたソフトウェア、または Boost ソフトウェアライセンスの下で開発されたソフトウェアが含まれます。本ソフトウェアに関する許諾および制限は、http://www.boost.org/LICENSE_1_0.txt にある使用条件に従います。

本製品には、University of Cambridge のが含まれます。Copyright (C) 1997-2007. 本ソフトウェアに関する許諾および制限は、<http://www.pcre.org/license.txt> にある使用条件に従います。

本製品には、The Eclipse Foundation のソフトウェアが含まれます。Copyright (C) 2007. All Rights Reserved. 本ソフトウェアに関する許諾および制限は、<http://www.eclipse.org/org/documents/epl-v10.php> および <http://www.eclipse.org/org/documents/edl-v10.php> にある使用条件に従います。

本製品には、<http://www.tcl.tk/software/tcltk/license.html>、<http://www.bosrup.com/web/overlib/?License>、<http://www.stlport.org/doc/license.html>、<http://www.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/license.html>、<http://www.jgraph.com/jgraphdownload.html>、<http://www.jcraft.com/jsch/LICENSE.txt>、http://jotm.objectweb.org/bsd_license.html に基づいて許諾されたソフトウェアが含まれています。<http://www.w3.org/Consortium/Legal/2002/copyright-software-20021231>、<http://www.slf4j.org/license.html>、<http://nanoxml.sourceforge.net/orig/copyright.html>、<http://www.json.org/license.html>、<http://forge.ow2.org/projects/javaservice/>、<http://www.postgresql.org/about/licence.html>、<http://www.sqlite.org/copyright.html>、<http://www.tcl.tk/software/tcltk/license.html>、<http://www.jaxen.org/faq.html>、<http://www.jdom.org/docs/faq.html>、<http://www.slf4j.org/license.html>、<http://www.iodbc.org/dataspace/iodbc/wiki/iODBC/License>、<http://www.keplerproject.org/md5/license.html>、<http://www.toedter.com/en/jcalendar/license.html>、<http://www.edankert.com/bounce/index.html>、<http://www.net-snmp.org/about/license.html>、<http://www.openmdx.org/#FAQ>、http://www.php.net/license/3_01.txt、<http://srp.stanford.edu/license.txt>、<http://www.schneider.com/blowfish.html>、<http://www.jmock.org/license.html>、<http://xsom.java.net>、<http://benalman.com/about/license/>、<https://github.com/CreateJS/EaselJS/blob/master/src/easeljs/display/Bitmap.js>、<http://www.h2database.com/html/license.html#summary>、<http://jsoncpp.sourceforge.net/LICENSE>、<http://jdbc.postgresql.org/license.html>、<http://protobuf.googlecode.com/svn/trunk/src/google/protobuf/descriptor.proto>、<https://github.com/rantav/hector/blob/master/LICENSE>、<http://web.mit.edu/Kerberos/krb5-current/doc/mitK5license.html>、<http://jibx.sourceforge.net/jibx-license.html>、<https://github.com/lyokato/libgeohash/blob/master/LICENSE>、<https://github.com/hjiang/jsonxx/blob/master/LICENSE>、<https://code.google.com/p/lz4/>、<https://github.com/jedisct1/libsodium/blob/master/LICENSE>、<http://one-jar.sourceforge.net/index.php?page=documents&file=license>、<https://github.com/EsotericSoftware/kryo/blob/master/license.txt>、<http://www.scala-lang.org/license.html>、<https://github.com/tinkerpop/blueprints/blob/master/LICENSE.txt>、<http://gee.cs.oswego.edu/dl/classes/EDU/oswego/cs/dl/util/concurrent/intro.html>、<https://aws.amazon.com/asl/>、<https://github.com/twbs/bootstrap/blob/master/LICENSE>、および <https://sourceforge.net/p/xmlunit/code/HEAD/tree/trunk/LICENSE.txt>。

本製品には、Academic Free License (<http://www.opensource.org/licenses/afl-3.0.php>)、Common Development and Distribution License (<http://www.opensource.org/licenses/cddl1.php>)、Common Public License (<http://www.opensource.org/licenses/cpl1.0.php>)、Sun Binary Code License Agreement Supplemental License Terms、BSD License (<http://www.opensource.org/licenses/bsd-license.php>)、BSD License (<http://opensource.org/licenses/BSD-3-Clause>)、MIT License (<http://www.opensource.org/licenses/mit-license.php>)、Artistic License (<http://www.opensource.org/licenses/artistic-license-1.0>)、Initial Developer's Public License Version 1.0 (<http://www.firebirdsql.org/en/initial-developer-s-public-license-version-1-0/>) に基づいて許諾されたソフトウェアが含まれています。

本製品には、ソフトウェア copyright (C) 2003-2006 Joe Walnes, 2006-2007 XStream Committers が含まれています。All rights reserved.本ソフトウェアに関する許諾および制限は、<http://j.org/license.html> にある使用条件に従います。本製品には、Indiana University Extreme! Lab によって開発されたソフトウェアが含まれています。詳細については、<http://www.extreme.indiana.edu/> を参照してください。

本製品には、ソフトウェア Copyright (C) 2013 Frank Balluffi and Markus Moeller が含まれています。All rights reserved.本ソフトウェアに関する許諾および制限は、MIT ライセンスの使用条件に従います。

特許については、<https://www.informatica.com/legal/patents.html> を参照してください。

免責: 本文書は、一切の保証を伴わない「現状渡し」で提供されるものとし、Informatica LLC は他社の権利の非侵害、市場性および特定の目的への適合性の黙示の保証などを含めて、一切の明示的および黙示的保証の責任を負いません。Informatica LLC では、本ソフトウェアまたはドキュメントに誤りのないことを保証していません。本ソフトウェアまたはドキュメントに記載されている情報には、技術的に不正確な記述や誤植が含まれる場合があります。本ソフトウェアまたはドキュメントの情報は、予告なしに変更されることがあります。

NOTICES

この Informatica 製品（以下「ソフトウェア」）には、Progress Software Corporation（以下「DataDirect」）の事業子会社である DataDirect Technologies からの特定のドライバ（以下「DataDirect ドライバ」）が含まれています。DataDirect ドライバには、次の用語および条件が適用されます。

1. DataDirect ドライバは、特定物として現存するままの状態提供され、商品性の保証、特定目的適合性の保証および法律上の瑕疵担保責任を含むすべての明示もしくは黙示の保証責任を負わないものとします。国または地域によっては、法律の強行規定により、保証責任の制限が禁じられる場合、強行規定の制限を受けるものとします。
2. DataDirect または第三者は、予見の有無を問わず発生した ODBC ドライバの使用に関するいかなる直接的、間接的、偶発的、特別、あるいは結果的損害に対して責任を負わないものとします。本制限事項は、すべての訴訟原因に適用されます。訴訟原因には、契約違反、保証違反、過失、厳格責任、詐称、その他の不法行為を含みますが、これらに限るものではありません。

発行日: 2018-06-08

目次

序文	8
Informatica のリソース	8
Informatica Network	8
Informatica ナレッジベース	8
Informatica マニュアル	8
Informatica 製品可用性マトリックス	9
Informatica Velocity	9
Informatica Marketplace	9
Informatica グローバルカスタマサポート	9
第 1 章 : 参照データの概要	10
参照データの概要	10
Informatica 参照データ	11
ユーザー定義の参照データ	11
参照テーブル	12
参照テーブルの構造	12
参照データウェアハウスの特権	13
パラメータと参照テーブル	13
参照データオブジェクトとバージョン管理	14
第 2 章 : Analyst ツールの参照テーブル	15
Analyst ツールの参照テーブルの概要	15
参照テーブルのプロパティ	15
参照テーブルの全般プロパティ	16
参照テーブルのカラムのプロパティ	16
参照テーブルエディタでの参照テーブルの作成	17
プロファイルデータからの参照テーブルの作成	17
プロファイルカラムデータからの参照テーブルの作成	18
値パターンからの参照テーブルの作成	19
フラットファイルからの参照テーブルの作成	20
Analyst ツールのフラットファイルプロパティ	20
フラットファイルからの参照テーブルの作成	21
データベーステーブルからの参照テーブルの作成	21
データベーステーブルからの参照テーブルの作成	22
バージョン管理されているモデルリポジトリでの参照テーブルの操作	22
参照テーブルの更新	23
カラムの管理	23
行の管理	24
値の検索と置換	25
参照テーブルのデータのエクスポート	25

管理されていない参照テーブルでの編集の有効化と無効化.	26
参照テーブル値のリフレッシュ.	26
監査証跡イベント.	27
監査証跡イベントの表示.	27
参照テーブルのルールおよびガイドライン.	28
第 3 章 : Developer ツールの参照データ.	29
Developer ツール参照データの概要.	29
参照データとトランスフォーメーション.	29
バージョン管理されているモデルリポジトリの参照データオブジェクトの操作.	30
参照データオブジェクトのチェックアウト.	30
参照データオブジェクトのチェックイン.	30
参照テーブル.	31
参照テーブルデータのプロパティ.	31
参照テーブルオブジェクトの作成.	32
フラットファイルからの参照テーブルの作成.	33
リレーショナルソースからの参照テーブルの作成.	34
コンテンツセット.	35
文字セット.	36
分類子モデル.	36
パターンセット.	37
確率モデル.	37
正規表現.	37
トークンセット.	38
確率的なモデルと分類子モデルのルールとガイドライン.	40
コンテンツセットの作成.	41
コンテンツセットでの参照データオブジェクトの作成.	41
第 4 章 : 分類子モデル.	42
分類子モデルの概要.	42
分類子モデルの構造.	43
分類子のスコア.	43
分類子トランスフォーメーションの例.	43
分類子モデルのオプション.	44
分類子モデルの参照データ.	45
分類子モデルのラベルデータ.	46
分類子モデルのラベルの管理.	46
分類子モデルの設定.	47
分類子モデルの作成.	47
分類子モデルへのデータソースからのデータの追加.	48
分類子モデルへの参照データ行の追加.	49
分類子モデルへのラベルの追加.	49
参照データ行へのラベルの割り当て.	49

未使用のラベル値の特定.	50
分類子モデルからの行の削除.	50
分類子モデルからのラベルの削除.	50
分類子モデルのコンパイル.	51
フィルタ操作と検索操作.	51
データ値を使用した参照データ行のフィルタリング.	51
ラベル値を使用した参照データ行のフィルタリング.	51
参照データ行内の値の検索.	51
コピーと貼り付けの操作.	52
別のコンテンツセットへの分類子モデルのコピー.	52
別のコンテンツセットからの分類子モデルのインポート.	52

第 5 章 : 確率モデル. 53

確率モデルの概要.	53
確率モデルの構造.	54
ラベラトランスフォーメーションの例.	54
パーサートランスフォーメーションの例.	55
確率モデルのオプション.	55
確率モデルのデータビュー.	56
確率モデルのラベルビュー.	57
確率モデルの参照データ.	58
確率モデルのラベルデータ.	58
オーバーフローラベル.	59
確率モデルのプロパティ.	60
確率モデルの設定.	60
空の確率モデルの作成.	61
データオブジェクトからの確率モデルの作成.	61
確率モデルへのデータソースからのデータの追加.	62
確率モデルへの参照データ行の追加.	63
確率モデルへのラベルの追加.	63
参照データ値へのラベルの割り当て.	64
複数のデータ値へのラベルの割り当て.	64
確率モデルからの行の削除.	65
確率モデルからのラベルの削除.	65
確率モデルのコンパイル.	65
確率モデル内のデータ行の検索.	65
ラベル割り当てによる参照データ値のフィルタリング.	66
未使用のラベル値の検索.	66
コピーと貼り付けの操作.	66
別のコンテンツセットへの確率モデルのコピー.	67
別のコンテンツセットからの確率モデルのインポート.	67
参照データ行のクリップボードへのコピー.	67

索引.....	68
---------	----

序文

『Informatica 参照データガイド』には Informatica Developer および Informatica Analyst で使用可能な参照データオブジェクトおよびファイルに関する情報が記載されています。本書は、データアナリストやデータスチュワードをはじめ、参照データを使用して組織のデータの精度と有用性を確認し向上させるユーザーを対象とします。

Informatica のリソース

Informatica Network

Informatica Network は、Informatica グローバルカスタマサポート、Informatica ナレッジベースなどの製品リソースをホストします。Informatica Network には、<https://network.informatica.com> からアクセスしてください。

メンバーは以下の操作を行うことができます。

- 1つの場所からすべての Informatica のリソースにアクセスできます。
- ドキュメント、FAQ、ベストプラクティスなどの製品リソースをナレッジベースで検索できます。
- 製品の提供情報を表示できます。
- 自分のサポート事例を確認できます。
- 最寄りの Informatica ユーザーグループネットワークを検索して、他のユーザーと共同作業を行えます。

Informatica ナレッジベース

ドキュメント、ハウツー記事、ベストプラクティス、PAM などの製品リソースを Informatica Network で検索するには、Informatica ナレッジベースを使用します。

ナレッジベースには、<https://kb.informatica.com> からアクセスしてください。ナレッジベースに関する質問、コメント、ご意見の連絡先は、Informatica ナレッジベースチーム (KB_Feedback@informatica.com) です。

Informatica マニュアル

使用している製品の最新のドキュメントを取得するには、https://kb.informatica.com/_layouts/ProductDocumentation/Page/ProductDocumentSearch.aspx にある Informatica ナレッジベースを参照してください。

このマニュアルに関する質問、コメント、ご意見の電子メールの送付先は、Informatica マニュアルチーム (infa_documentation@informatica.com) です。

Informatica 製品可用性マトリックス

製品可用性マトリックス（PAM）には、製品リリースでサポートされるオペレーティングシステム、データベースなどのデータソースおよびターゲットが示されています。Informatica Network メンバである場合は、PAM (<https://network.informatica.com/community/informatica-network/product-availability-matrices>) にアクセスできます。

Informatica Velocity

Informatica Velocity は、Informatica プロフェッショナルサービスによって開発されたヒントおよびベストプラクティスのコレクションです。数多くのデータ管理プロジェクトの経験から開発された Informatica Velocity には、世界中の組織と協力して優れたデータ管理ソリューションの計画、開発、展開、および維持を行ってきた弊社コンサルタントの知識が集約されています。

Informatica Network メンバである場合は、Informatica Velocity リソース (<http://velocity.informatica.com>) にアクセスできます。

Informatica Velocity についての質問、コメント、またはアイデアがある場合は、ips@informatica.com から Informatica プロフェッショナルサービスにお問い合わせください。

Informatica Marketplace

Informatica Marketplace は、お使いの Informatica 製品を強化したり拡張したりするソリューションを検索できるフォーラムです。Informatica の開発者およびパートナーの何百ものソリューションを利用して、プロジェクトで実装にかかる時間を短縮したり、生産性を向上させたりできます。Informatica Marketplace には、<https://marketplace.informatica.com> からアクセスできます。

Informatica グローバルカスタマサポート

Informatica Network の電話またはオンラインサポートからグローバルカスタマサポートに連絡できます。

各地域の Informatica グローバルカスタマサポートの電話番号は、Informatica Web サイト (<http://www.informatica.com/us/services-and-training/support-services/global-support-centers>) を参照してください。

Informatica Network メンバである場合は、オンラインサポート (<http://network.informatica.com>) を使用できます。

第 1 章

参照データの概要

この章では、以下の項目について説明します。

- [参照データの概要, 10 ページ](#)
- [Informatica 参照データ, 11 ページ](#)
- [ユーザー定義の参照データ, 11 ページ](#)
- [参照テーブル, 12 ページ](#)
- [参照データオブジェクトとバージョン管理, 14 ページ](#)

参照データの概要

Informatica トランスフォーメーションは、参照データを使用してデータを分析および更新できます。参照データオブジェクトは、Developer tool と Analyst ツールで作成できます。参照データオブジェクトとファイルをモデルリポジトリとファイルシステムにインポートすることもできます。Data Quality Content Installer を使用して、参照データオブジェクトをインポートし、参照データファイルをインストールできます。

次のタイプの参照データを作成して編集できます。

参照テーブル

参照テーブルには、データ値のセットの標準バージョンと代替バージョンが含まれています。Developer tool で参照テーブルをトランスフォーメーションに追加し、ソースのデータ値が正確で、正しい形式であることを検証します。

ほとんどの参照テーブルには、少なくとも 2 つのカラムが含まれています。1 つのカラムには値の標準または優先バージョンが含まれ、もう 1 つのカラムには代替バージョンが含まれます。参照テーブルをトランスフォーメーションに追加すると、テーブルに存在する値が入力ポートデータで検索されます。作業するデータプロジェクトに役立つデータを含むテーブルを作成できます。

コンテンツセット

コンテンツセットは、リポジトリまたはファイル内の参照データの値を指定するモデルリポジトリオブジェクトです。コンテンツセットをトランスフォーメーションに追加すると、トランスフォーメーションでコンテンツセット内のデータパターンと一致する値を求めて入力データが検索されます。

Data Quality Content Installer で、以下のタイプの参照データをインストールできます。

Informatica 参照テーブル

Informatica で開発するリポジトリオブジェクトおよびデータファイル。アクセラレータオブジェクトをモデルリポジトリにインポートする際は、Informatica 参照テーブルをインポートします。参照情報には、

市外局番、郵便番号の形式、名前、職業、頭字語などのタイプがあります。Informatica 参照テーブルは編集できます。

Informatica コンテンツセット

Informatica で開発するリポジトリオブジェクトおよびデータファイル。アクセラレータオブジェクトをモデルリポジトリにインポートする際は、コンテンツセットをインポートします。コンテンツセットには、データ品質トランスフォーメーションで検索操作の実行に使用できる、さまざまなタイプの参照データが含まれます。

アドレス参照データファイル

国内に配達可能なアドレスのデータを含む参照データファイル。アドレスバリデータトランスフォーメーションで参照データを読み取ります。アドレス参照データファイルを作成したり編集したりすることはできません。

アドレス参照データは、定義されている期間において最新のものであり、定期的（四半期ごとなど）にデータを更新する必要があります。

ID ポピュレーションファイル

個人、世帯、企業の ID に関する情報を含む参照データファイル。一致トランスフォーメーションと比較トランスフォーメーションでポピュレーションファイルを使用して、入力データに存在する可能性のある ID を検索します。ID ポピュレーションファイルを作成または編集することはできません。

Informatica 参照データ

アドレス参照データと ID ポピュレーションデータは、Informatica から購入してダウンロードできます。

各国のアドレスデータは年間契約で購入できます。サブスクリプションの期間内であれば、最新のアドレスデータを Informatica からいつでもダウンロードすることができます。

参照データは、Content Installer を実行するユーザーが、アプリケーションとは別にダウンロードしてインストールします。システムにインストールされている参照データについては、管理者に問い合わせてください。

ユーザー定義の参照データ

データオブジェクトの値を使用して、参照データオブジェクトを作成できます。

例えば、プロジェクトまたは組織に固有の値が含まれるデータオブジェクトまたはプロファイルカラムを選択できます。カラム値からカスタム参照データオブジェクトを作成します。

以下のことを検証する手段として、データカラムから参照データオブジェクトを作成できます。

- カラムのデータ行に同じタイプの情報が含まれている。
- ソース値が有効である。参照オブジェクトには有効な値のリストが含まれていることも、有効でない値のリストが含まれていることもあります。

以下の表に、参照データを含むことができるプロジェクトデータカラムの代表的な例を示します。

情報	参照データの例
最小在庫管理単位 (SKU) コード	SKU カラムを使用して、ある企業にとって有効な SKU コードの参照テーブルを作成します。参照テーブルを使用して、データセットで正しいまたは正しくない SKU コードを検索します。
従業員コード	従業員コードまたは従業員 ID カラムを使用して、有効な従業員コードの参照テーブルを作成します。参照テーブルを使用して、従業員データのエラーを検索します。
顧客口座番号	顧客口座カラムに対してプロファイルを実行し、口座番号パターンを識別します。プロファイルを使用して、正しくないデータパターンのトークンセットを作成します。トークンセットを使用して、正しい口座番号構造に従っていない口座番号を検索します。
顧客名	顧客名カラムに名、ミドルネーム、および姓が含まれている場合、カラムで想定される文字列構造を定義する確率モデルを作成できます。確率モデルを使用して、カラムに属さないデータ文字列を検索します。

参照テーブル

Analyst ツールと Developer tool で、参照テーブルを作成および更新します。

参照テーブルのメタデータはモデルリポジトリに格納されます。参照テーブルのカラムデータは、参照データウェアハウスまたは別のデータベースに格納できます。参照データウェアハウスにカラムデータを格納すると、Informatica サービスはテーブルを管理されている参照テーブルとして識別します。別のデータベースにカラムデータを格納すると、Informatica サービスはテーブルを管理されていない参照テーブルとして識別します。

コンテンツ管理サービスは、参照データウェアハウスのデータベース接続を格納します。参照データウェアハウスとしては、IBM DB2 データベース、Microsoft SQL Server データベース、または Oracle データベースを指定できます。

別のデータベースから参照データウェアハウスにデータをインポートする場合は、ネイティブ接続または ODBC 接続を使用してデータをインポートします。管理されていないデータベースを参照テーブルのデータソースとして指定する場合は、ネイティブ接続を使用してデータベースに接続します。

参照テーブルの構造

ほとんどの参照テーブルには、少なくとも 2 つのカラムが含まれています。1 つのカラムには、正しいまたは必要なバージョンのデータ値が含まれます。他のカラムには、ソースデータに含まれる代替バージョンなど、異なるバージョンの値が含まれます。

正しい値または必要な値が含まれたカラムは、有効カラムと呼ばれます。トランスフォーメーションがマッピング内の参照テーブルを読み取るときに、トランスフォーメーションは無効なカラムの値を探します。無効な値を見つけると、トランスフォーメーションは有効カラムからそれに相当する値を返します。有効な値を返すのではなく、単一の共通の値を返すようにトランスフォーメーションを設定することもできます。

有効カラムには、ZIP コードなど、公式に正しいデータを含むことができます。例えば、組織固有の最小在庫管理単位 (SKU) の値など、プロジェクトに関連のあるデータを含めることができます。また、検索で使用するために、既知のデータエラーを含む値など、不良データの有効カラムを作成することもできます。

例えば、小売企業の有効な SKU 番号のリストが含まれる参照テーブルを作成するとします。ラベラトランスフォーメーションに参照テーブルを追加し、トランスフォーメーションとのマッピングを作成します。製品デー

データベースに対してこのマッピングを実行します。マッピングが実行されると、ラベラーは有効な SKU の値を含まない製品レコードを識別するカラムを作成します。

参照テーブルおよびパーサートランスフォーメーション

パターンベースの解析処理でテーブルデータを使用するには、単一カラムの参照テーブルを作成します。パターンベースの解析を実行するようにパーサートランスフォーメーションを設定し、トランスフォーメーションの設定に参照データをインポートします。

参照データウェアハウスの特権

コンテンツ管理サービスは、特権を使用して参照テーブルへのユーザーアクションを制限します。Administrator ツールのセキュリティオプションでサービスの特権を確認または更新します。

参照テーブルを使用するには、コンテンツ管理サービスにおける以下の特権が必要です。

- 参照テーブルの作成
- 参照テーブルデータの編集
- 参照テーブルメタデータの編集

管理されていない参照テーブルのデータを編集するには、編集を許可するように参照テーブルオブジェクトが設定されていることも確認します。

注: 管理されていない参照テーブルのメタデータをデータベースアプリケーションで編集する場合は、Analyst ツールを使用してモデルリポジトリをテーブルと同期します。Developer tool で管理されていない参照テーブルを使用する前に、モデルリポジトリとテーブルを同期する必要があります。

パラメータと参照テーブル

パラメータを使用して、モデルリポジトリの参照テーブルを特定できます。Developer tool で参照テーブルを特定するパラメータを作成します。または、参照テーブルの場所をパラメータファイルに追加します。

Developer tool でパラメータを作成する場合は、マッピングでそのパラメータをトランスフォーメーションに追加します。参照テーブルの場所をパラメータファイルに追加する場合は、コマンドプロンプトでマッピングを実行するときにファイルを指定します。どちらの場合も、マッピングの実行時にデータ統合サービスがパラメータによって特定された参照テーブルを読み込みます。

参照テーブルを特定するパラメータは次のトランスフォーメーションに追加できます。

- 大文字小文字変換プログラムトランスフォーメーション
- ラベラトランスフォーメーション
- トークン解析モードでのパーサートランスフォーメーション
- 標準化トランスフォーメーション

注: コマンドプロンプトで、`infacmd ms runMapping` コマンドを使用してマッピングを実行します。

参照データオブジェクトとバージョン管理

参照データオブジェクトを保存するモデルリポジトリをバージョン管理アプリケーションと統合する場合は、オブジェクトにバージョン管理を適用することができます。バージョン管理は、参照テーブルとコンテンツセットに適用できます。

参照データオブジェクトは、バージョン管理をサポートするモデルリポジトリからチェックインおよびチェックアウトすることができます。チェックアウトを元に戻し、前バージョンのオブジェクトを取得し、オブジェクトを前バージョンに保存することができます。参照データオブジェクトでバージョン管理が行われていない場合は、モデルリポジトリで、編集する参照データオブジェクトがロックされます。その他ユーザーは、作業中のロックされたオブジェクトを編集できません。オブジェクトを閉じると、モデルリポジトリでロックが解除され、その他ユーザーはそのオブジェクトを編集できます。

注: バージョン管理は、モデルリポジトリに保存される管理されていない参照テーブルオブジェクトのメタデータに適用されます。バージョン管理は、管理されていない参照テーブルのデータには適用されません。前バージョンの管理されていない参照テーブルから、参照データを表示または保存することはできません。

第 2 章

Analyst ツールの参照テーブル

この章では、以下の項目について説明します。

- [Analyst ツールの参照テーブルの概要, 15 ページ](#)
- [参照テーブルのプロパティ, 15 ページ](#)
- [参照テーブルエディタでの参照テーブルの作成, 17 ページ](#)
- [プロファイルデータからの参照テーブルの作成, 17 ページ](#)
- [フラットファイルからの参照テーブルの作成, 20 ページ](#)
- [データベーステーブルからの参照テーブルの作成, 21 ページ](#)
- [バージョン管理されているモデルリポジトリでの参照テーブルの操作, 22 ページ](#)
- [参照テーブルの更新, 23 ページ](#)
- [監査証跡イベント, 27 ページ](#)
- [参照テーブルのルールおよびガイドライン, 28 ページ](#)

Analyst ツールの参照テーブルの概要

Analyst ツールの [デザイン] ワークスペースで参照テーブルを作成します。

参照テーブルは、フラットファイル、モデルリポジトリのデータソース、および別のデータベースのテーブルから作成できます。

参照テーブルは、プロファイルカラムまたはプロファイルカラム内のデータのサブセットから作成できます。また、プロファイルから選択するカラムパターンから参照テーブルを作成することも可能です。

参照テーブルを作成または更新する際は、テーブルおよびテーブルに含まれるデータ列でプロパティを設定します。

参照テーブルのプロパティ

Analyst ツールでは、参照テーブルのプロパティを表示および更新できます。参照テーブルには全般プロパティおよびカラムのプロパティが表示されます。全般プロパティには、参照テーブル名、作成日付、データベースの接続名、および有効なカラム名などが含まれます。カラムのプロパティには、カラム名、精度値、およびスケール値などが含まれます。

プロパティは読み取り専用モードで表示することができます。プロパティを更新するには、参照テーブルを編集またはチェックアウトします。

参照テーブルの全般プロパティ

全般プロパティには、参照テーブルオブジェクトに関する情報が含まれます。

次の表は、全般プロパティの説明です。

プロパティ	説明
名前	参照テーブル名。
説明	参照テーブルについてユーザーが入力したすべての説明。
場所	モデルリポジトリでの参照テーブルオブジェクトの場所。
有効なカラム	参照テーブルでの有効カラムの名前。
作成日時	参照テーブル名の作成日時。
作成者	参照テーブルを作成したユーザーのログイン名。
最終変更日	参照テーブルに対する最新の更新の日時。
最終変更者	最新の更新を行ったユーザーのログイン名。
接続名	参照データの値を保存するデータベースの接続名。
タイプ	参照テーブルタイプ。参照テーブルには管理されているものと、管理されていないものがあります。

参照テーブルのカラムのプロパティ

カラムのプロパティには、カラムのメタデータに関する情報が含まれます。

以下の表に、カラムのプロパティを示します。

プロパティ	説明
名前	列の名前です。
データタイプ	各カラム内のデータのデータ型。以下のいずれかのデータ型を選択できます。 <ul style="list-style-type: none">- bigint- date/time- 10 進数- ダブル- 整数- 文字列 空の参照テーブルを作成する場合や、フラットファイルから参照テーブルを作成する場合は、倍精度浮動小数点数型データ型を選択できません。
精度	各カラムの精度。精度は、カラムに格納できる最大桁数または最大文字数です。精度の値は、データ型に応じて設定します。
スケール	各カラムのスケール。スケールは、カラムに格納できる小数点以下の最大桁数です。小数のカラムに適用されます。スケールの値は、データ型に応じて設定します。

プロパティ	説明
説明	任意での各カラムの説明。
NULL 可能	カラムが NULL 値を持つことが可能かどうかを示します。
キー	キーカラム特定します。キーカラムを指定するテーブルから参照データをインポートした場合、Analyst ツールはキーカラムを特定できます。

参照テーブルエディタでの参照テーブルの作成

参照テーブルエディタでテーブル構造を定義し、参照テーブルにデータを追加します。

1. **【新規】** > **【参照テーブル】** をクリックします。
新しい参照テーブルウィザードが開きます。
2. **【参照テーブルエディタを使用します】** オプションを選択し、**【次へ】** をクリックします。
3. テーブルにカラムを追加するには、**【新しいカラムの追加】** オプションを使用します。
4. カラムごとにプロパティを設定します。
プロパティには、カラム名、データ型、精度、およびスケールなどがあります。
トランスフォーメーションで参照データの検索によって返すことができるデータがカラムに含まれている場合は、**【有効】** オプションを選択します。
5. 必要に応じてカラムを追加し、詳細な説明を参照テーブルのメタデータとして含めます。
6. 必要に応じて、テーブルの監査のメモを入力します。
監査のメモは、監査証跡ログに表示されます。
7. **【次へ】** をクリックします。
8. 参照テーブルの名前を入力し、モデルリポジトリでの参照テーブルオブジェクトの場所を選択します。
9. **【完了】** をクリックします。

プロファイルデータからの参照テーブルの作成

プロファイル内のソースデータに関連付けた参照テーブルを、プロファイルデータから作成できます。参照テーブルを使用して、ソースデータ内のさまざまなタイプの情報を検索します。

プロファイルを使用して参照テーブルを作成または更新するには、以下の方法があります。

- プロファイルのカラムを選択して参照テーブルに追加する。
- プロファイルカラムを参照し、カラムデータのサブセットを参照テーブルに追加する。
- プロファイルのカラムを選択し、そのカラムのパターン値を参照テーブルに追加する。

プロファイルカラムデータからの参照テーブルの作成


参照テーブルは、プロファイルデータカラム内の 1 つ以上の値から作成できます。プロファイルでカラムを選択し、参照テーブルに追加するカラム値を選択します。

1. Analyst ツールで、**【ライブラリ】** ワークスペースを開きます。
2. **【プロファイル】** アセットカテゴリを選択します。
ライブラリには、モデルリポジトリ内のプロファイルのリストが表示されます。
3. 参照テーブルに追加するカラムが含まれるプロファイルを開きます。
プロファイルの概要には、プロファイルのカラム名がリスト表示されます。
4. カラムデータを確認します。
カラムデータを表示するには、カラム名をクリックします。
5. 詳細なプロファイルビューで、参照テーブルに追加するデータ値を選択します。値は、1 つずつ選択するか、すべてまとめて選択できます。
6. カラム名を右クリックし、**【参照テーブルへの追加】** を選択します。
次の図に、詳細なプロファイルビューでのデータカラムを示します。

COLUMN2

▼ Values

20 distinct values (0 Non-unique values | 20 Unique values)



<input checked="" type="checkbox"/>	Value	Frequency	Length	Percentage
<input checked="" type="checkbox"/>	ThePickwickPapers	1	17	5.00%
<input checked="" type="checkbox"/>	TheOldCuriosityShop	1	19	5.00%
<input checked="" type="checkbox"/>	TheMysteryofEdwinDrood	1	22	5.00%
<input checked="" type="checkbox"/>	TheLifeandAdventuresofNi...	1	38	5.00%
<input checked="" type="checkbox"/>	TheLifeandAdventuresofMa...	1	38	5.00%

Drilldown

Add to Reference Table

Create Value Frequency Rule

Create Data domain

Length (min → max) 9 → 38
Value (min → max) AChristmasCarol → ThePickwickPapers

図の番号 1 は、**【参照テーブルへの追加】** オプションを示しています。

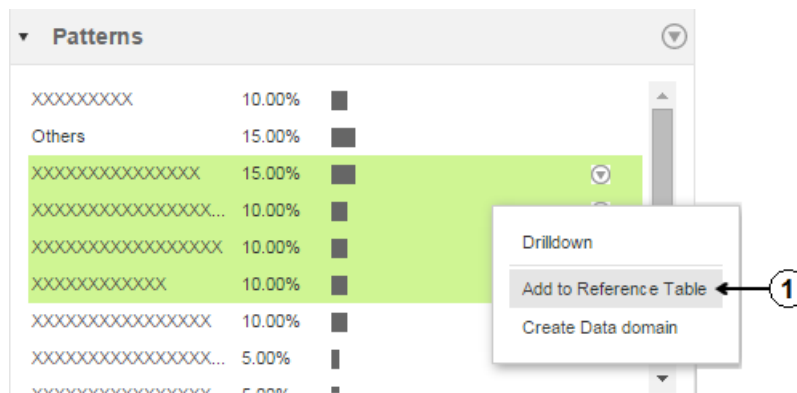
7. **参照テーブルへの追加**ウィザードが開きます。
【参照テーブルの作成】 オプションを選択します。
注: データを現在の参照テーブルに追加するオプションを選択することもできます。
8. **【次へ】** をクリックします。
カラム名は、デフォルトで参照テーブル名として表示されます。必要に応じて、名前を更新します。
9. 必要に応じて、説明とデフォルト値を入力します。
値が含まれていないテーブルレコードにはこのデフォルト値が使用されます。
10. **【次へ】** をクリックします。
11. カラムのプロパティを確認します。
必要に応じて、詳細な説明メタデータ用のカラムを作成するオプションを選択します。
12. **【次へ】** をクリックします。

13. 参照テーブル名と説明を確認します。
必要に応じて、監査のメモを入力します。
14. 参照テーブルオブジェクトのモデルリポジトリの場所を選択します。
15. **【完了】** をクリックします。

値パターンからの参照テーブルの作成

参照テーブルは、プロファイルカラムのカラムパターンからも作成できます。パターンは、1 つ以上のカラムフィールドにあるデータ値の構成を表します。プロファイルでカラムを選択し、作成する参照テーブルに追加するパターンを選択します。

1. Analyst ツールで、**【ライブラリ】** ワークスペースを開きます。
2. **【プロファイル】** アセットカテゴリを選択します。
ライブラリには、モデルリポジトリ内のプロファイルのリストが表示されます。
3. 参照テーブルに追加する値パターンが含まれるプロファイルを開きます。
プロファイルの概要には、プロファイルのカラム名がリスト表示されます。
4. 参照テーブルに追加するパターンデータを定義するカラムを選択します。
5. カラムデータパターンを確認します。
カラムデータを表示するには、カラム名をクリックします。
6. 詳細なプロファイルビューで、追加するカラムパターンを選択します。
7. 選択したパターンを右クリックし、**【参照テーブルへの追加】** を選択します。
次の図に、詳細なプロファイルビューでのカラムのデータパターンを示します。



図の番号 1 は、**【参照テーブルへの追加】** オプションを示しています。

8. **【参照テーブルへの追加】** ウィザードが開きます。
【参照テーブルの作成】 オプションを選択します。
注: データを現在の参照テーブルに追加するオプションを選択することもできます。
9. **【次へ】** をクリックします。
カラム名は、デフォルトで参照テーブル名として表示されます。必要に応じて、名前を更新します。
10. 必要に応じて、説明とデフォルト値を入力します。
値が含まれていないテーブルレコードにはこのデフォルト値が使用されます。
11. **【次へ】** をクリックします。

12. カラムのプロパティを確認します。
必要に応じて、詳細な説明メタデータ用のカラムを作成するオプションを選択します。
13. **【次へ】** をクリックします。
14. 参照テーブル名と説明を確認します。
必要に応じて、監査のメモを入力します。
15. 参照テーブルオブジェクトのモデルリポジトリの場所を選択します。
16. **【完了】** をクリックします。

フラットファイルからの参照テーブルの作成

CSV ファイルから参照データをインポートできます。ファイルデータをインポートするには、**新しい参照テーブルウィザード**を使用します。

参照テーブルの作成に使用するフラットファイルごとにプロパティを設定する必要があります。

Analyst ツールのフラットファイルプロパティ

フラットファイルを参照テーブルとしてインポートする場合、ファイル内の各カラムのプロパティを設定する必要があります。設定したオプションによって、Analyst ツールによるファイルのデータの読み取り方法が決まります。

次の表に、参照テーブルのファイルデータをインポートする際に設定可能なプロパティを記載します。

プロパティ	説明
区切り文字	データのカラムをパーティション化するのに使用される文字。[その他] フィールドを使用すれば、別の区切り文字を入力できます。 区切り文字は印字可能な文字でなければなりません。また、エスケープ文字や引用符を選択した場合には、それと別な文字でなければなりません。 印字されないマルチバイト文字を区切り文字として選択することはできません。
引用符	引用符はテキスト文字列の境界を定義します。 [なし]、[シングルクォーテーション]、または [ダブルクォーテーション] を選択してください。 引用符を選択すると、引用符で囲まれた文字列内の区切り文字が無視されます。
カラム名	1 行目からカラム名をインポートします。カラム名が最初の行に含まれている場合に、このオプションを選択します。 プレビューの先頭の行のデータがカラム名に使用されます。 デフォルトでは、無効になっています。
値	ある行から値のインポートを開始するオプション。ウィザードでのファイルのインポート時に読み取りを開始するプレビューの行番号を示します。

フラットファイルからの参照テーブルの作成

フラットファイルから参照テーブルデータを作成するときは、ファイルのカラム構造を使用したテーブルにファイルデータをインポートします。

1. **【新規】** > **【参照テーブル】** をクリックします。
【新しい参照テーブル】 ウィザードが表示されます。
2. **【フラットファイルのインポート】** オプションを選択します。
3. **【次へ】** をクリックします。
4. **【ファイルを選択】** をクリックして、フラットファイルを選択します。
5. フラットファイルのデータに一致するコードページを選択します。
6. **【アップロード】** をクリックしてファイルデータをアップロードします。
7. **【次へ】** をクリックします。
8. フラットファイルのプロパティを設定します。
プロパティにより、ファイルで使用される区切り文字と、ファイルの第 1 行目にカラム名を含めるかどうかを指定します。
9. 設定したプロパティをプレビューするには、**【プレビュー】** ペインを更新します。
10. **【次へ】** をクリックします。
11. カラムごとにプロパティを設定します。
プロパティには、カラム名、データ型、精度、およびスケールなどがあります。
トランスフォーメーションで参照データの検索によって返すことができるデータがカラムに含まれている場合は、**【有効】** オプションを選択します。
12. 必要に応じてカラムを追加し、詳細な説明を参照テーブルのメタデータとして含めます。
13. 必要に応じて、テーブルの監査のメモを入力します。
監査のメモは、監査証跡ログに表示されます。
14. **【次へ】** をクリックします。
15. 参照テーブルの名前を入力し、モデルリポジトリでの参照テーブルオブジェクトの場所を選択します。
16. 必要に応じて、テーブルの説明を入力します。
17. **【完了】** をクリックします。

データベーステーブルからの参照テーブルの作成

データベーステーブルから参照テーブルを作成する場合は、モデルリポジトリにメタデータオブジェクトを作成します。参照データデータウェアハウスにテーブルデータをインポートすることもできます。

管理されている参照テーブルを作成する場合は、カラムデータを参照データウェアハウスにインポートします。管理されていない参照テーブルを作成する場合は、カラムデータを格納するデータベーステーブルを指定します。管理されている参照テーブルは、ODBC 接続またはネイティブ接続から作成できます。管理されていない参照テーブルは、ネイティブ接続から作成できます。

参照テーブルを作成する前に、参照データを含むデータベースへの接続が Informatica ドメインに含まれていることを確認します。ドメインにデータベースへの接続が含まれていない場合、Analyst ツールで定義できません。

データベース接続を定義するには、**【管理】** > **【接続】** をクリックします。

データベーステーブルからの参照テーブルの作成

参照テーブルを作成するには、データベースに接続して参照データを含むテーブルを選択します。

1. **【新規】** > **【参照テーブル】** を選択します。

新しい参照テーブルウィザードが表示されます。

2. **【リレーショナルテーブルに接続する】** オプションを選択します。

参照データウェアハウスにデータを格納しない参照テーブルを作成する場合は、**【未管理のテーブル】** を選択します。

ユーザーが管理されていない参照テーブルを編集できるようにするには、**【編集可能】** オプションを選択します。

【次へ】 をクリックします。

3. 接続リストからデータベース接続を選択します。

【次へ】 をクリックします。

4. **【テーブル】** パネルでテーブルを選択します。

5. **【プロパティ】** パネルで、テーブルのプロパティを確認します。

必要に応じて、**【データプレビュー】** をクリックしてテーブルデータを表示します。

【次へ】 をクリックします。

6. **【カラム属性】** パネルで有効なカラムを選択します。

管理された参照テーブルを作成する場合は、**【カラム属性】** パネルで以下のアクションを実行します。

- 参照テーブルのカラム名を編集する。
- 行レベルの説明のため、メタデータカラムを追加する。

7. 必要に応じてカラムを追加し、詳細な説明を参照テーブルのメタデータとして含めます。

8. 必要に応じて、テーブルの監査のメモを入力します。

監査のメモは、監査証跡ログに表示されます。

9. **【次へ】** をクリックします。

10. 参照テーブルの名前を入力し、モデルリポジトリでの参照テーブルオブジェクトの場所を選択します。

11. 必要に応じて、参照テーブルの説明を入力します。

12. **【完了】** をクリックします。

バージョン管理されているモデルリポジトリでの参照テーブルの操作

参照テーブルを読み取り専用モードで開きます。参照テーブルで作業する場合は、編集モードに切り替えるか、またはモデルリポジトリから参照テーブルをチェックアウトする必要があります。

1. Informatica ツールバーで、**【開く】** をクリックします。

アセットライブラリが開きます。

2. **【参照テーブル】** アセットカテゴリを選択し、参照テーブル名を選択します。

参照テーブルが読み取り専用モードで開きます。

3. 現在のバージョンの参照テーブルを編集するには、**【編集】** をクリックします。

バージョン管理されているモデルリポジトリの参照テーブルを編集するには、参照テーブルをチェックアウトします。

4. 参照テーブルでの作業が完了したら、**【完了】** をクリックします。Analyst ツールで、参照テーブルに変更内容が保存されます。

バージョン管理されているモデルリポジトリから参照テーブルをチェックアウトした場合は、オブジェクトをチェックインします。バージョン管理されているモデルリポジトリの参照テーブルのバージョンは、オブジェクトをチェックインするまで更新されません。

参照テーブルの更新

参照テーブルに含まれるビジネスデータは経時的に変化します。参照テーブル中のデータやメタデータを確認および更新してテーブルに正確な情報が含まれていることを確認します。参照テーブルの更新は Analyst ツールで行います。データやメタデータの更新は管理された参照テーブルと管理されていない参照テーブルのどちらでも行えます。

参照テーブルのデータやメタデータには以下の操作を実行できます。

カラムの管理

カラムの追加、カラムの削除、およびカラムプロパティの編集が可能です。

行の編集

参照テーブルにデータ行を追加できます。

参照データ値の更新

参照データ値を編集できます。

データ値の置換

【検索と置換】 オプションを使用して、正確ではなくなった、あるいは組織に関係なくなったデータ値を置き換えます。カラムの値を検索し、別の値に置き換えることができます。カラムのすべての値を単一の値に置き換えられます。

参照テーブルのエクスポート

参照テーブルを、カンマ区切り（CSV）ファイル、辞書ファイル、Excel ファイルにエクスポートします。

管理されていないテーブルの編集の有効化または無効化

管理されていない参照テーブルを更新して、テーブルのデータやメタデータへの編集を有効または無効にします。

参照テーブルデータのリフレッシュ

参照テーブルデータを Analyst ツールに再ロードし、データの最新の変更を確認します。

カラムの管理

参照テーブルにカラムを追加し、カラムのプロパティを更新することができます。また、管理されていない参照テーブルの編集可能なステータスを更新することもできます。

1. **【開く】** をクリックします。
アセットライブラリが開きます。
2. **【参照テーブル】** アセットカテゴリを選択し、参照テーブル名を選択します。

参照テーブルが読み取り専用モードで開きます。

3. 現在のバージョンの参照テーブルを編集するには、**【編集】** をクリックします。

バージョン管理されているモデルリポジトリの参照テーブルを編集するには、参照テーブルをチェックアウトします。

4. **【アクション】** メニューを開き、**【カラムプロパティの変更】** を選択します。

【カラムプロパティの変更】 ダイアログボックスが開きます。ダイアログボックスのオプションを使用して以下の操作を実行します。

- カラムの追加
- このテーブルの有効なカラムを変更します。
- カラム名を変更します。
- カラムの説明テキストを更新します。
- 管理されていない参照テーブルの編集可能なステータスを更新します。
- テーブルの監査のメモを更新します。

5. 操作が完了したら、**【OK】** をクリックします。

行の管理

参照テーブルの行を追加、編集、または削除することができます。

1. **【開く】** をクリックします。

アセットライブラリが開きます。

2. **【参照テーブル】** アセットカテゴリを選択し、参照テーブル名を選択します。

参照テーブルが読み取り専用モードで開きます。

3. 現在のバージョンの参照テーブルを編集するには、**【編集】** をクリックします。

バージョン管理されているモデルリポジトリの参照テーブルを編集するには、参照テーブルをチェックアウトします。

4. データの行を編集します。データの行は以下の方法で編集できます。

- 行を追加するには、**【アクション】** > **【行の追加】** を選択します。

【行の追加】 ダイアログボックスで、有効なカラムおよび少なくとも 1 つの他のカラムに値を入力します。必要に応じて、監査のメモを入力します。

【OK】 をクリックして行を追加します。

- 単一データ値を更新するには、値をクリックしてデータを更新します。

データを更新したら、行レベルのオプションを使用してデータを受け入れまたは拒否します。データ行にデータを直接入力するときに監査のメモを入力することはできません。

- 行のデータ値を更新するには、**【アクション】** > **【行の編集】** を選択します。

【行の編集】 ダイアログボックスで、1 つ以上のカラムに値を入力します。必要に応じて、監査のメモを入力します。

【適用】 をクリックして、選択したカラムのデータを更新します。

- 複数の行の値を更新するには、編集する行を選択して、**【アクション】** > **【行の編集】** を選択します。

【複数の行の編集】 ダイアログボックスで、1 つ以上のカラムに値を入力します。必要に応じて、監査のメモを入力します。

【OK】 をクリックして、選択したカラムのデータを更新します。

- 行を削除するには、削除する行を選択して **【アクション】** > **【削除】** をクリックします。

【行の削除】 ダイアログボックスで、必要に応じて監査のメモを入力します。

【OK】 をクリックして行を削除します。

注: Developer ツールを使用して大きな参照テーブルで行データを編集します。たとえば、参照テーブルに 500 を超える行が含まれる場合には、Developer ツールでテーブルを編集します。

値の検索と置換

参照テーブル中のデータ値を検索したり置き換えたりできます。テーブルに更新するデータ値のインスタンスが 1 つ以上含まれるとき、検索と置換オプションを使用します。

1. **【開く】** をクリックします。
アセットライブラリが開きます。
2. **【参照テーブル】** アセットカテゴリを選択し、参照テーブル名を選択します。
参照テーブルが読み取り専用モードで開きます。
3. 現在のバージョンの参照テーブルを編集するには、**【編集】** をクリックします。
バージョン管理されているモデルリポジトリの参照テーブルを編集するには、参照テーブルをチェックアウトします。
4. **【アクション】** > **【検索と置換】** をクリックします。
【検索と置換】 ツールバーが表示されます。
5. ツールバーで検索条件を入力します。
 - **【検索】** フィールドでデータ値を入力します。
 - 検索するカラムを選択します。デフォルトでは、すべてのカラムが検索されます。
 - **【置換後の文字列】** フィールドにデータ値を入力します。
6. 以下のオプションを使用して値を個別に、または一括で置き換えます。
 - **【次へ】** と **【前へ】** オプションを使用して値を個別に検索します。
 - 値を置き換えるには、**【置換】** を選択します。
 - 値のすべてのインスタンスを表示するには、**【すべてハイライト】** を選択します。
 - すべてのインスタンスの値を置き換えるには、**【すべて置換】** を選択します。

参照テーブルのデータのエクスポート

参照テーブルのデータを、カンマ区切りファイル、辞書ファイル、または Microsoft Excel ファイルにエクスポートします。データは読み取り専用モードでエクスポートすることができます。

1. **【開く】** をクリックします。
アセットライブラリが開きます。
2. **【参照テーブル】** アセットカテゴリを選択し、参照テーブル名を選択します。
参照テーブルが読み取り専用モードで開きます。
3. **【アクション】** > **【データのエクスポート】** をクリックします。
【ファイルへのデータのエクスポート】 ダイアログボックスが開きます。

次の表で、ダイアログボックスのオプションについて説明します。

オプション	説明
ファイル名	データを入力するファイルの名前。エクスポート操作によりファイルが作成されます。
ファイル形式	データを入力するファイルの形式。次のいずれかの形式を選択します。 <ul style="list-style-type: none">• csv: カンマ区切りファイルデフォルト形式。• xls: Microsoft Excel ファイル。• dic: Informatica 辞書ファイル
先頭行としてフィールド名をエクスポートする	カラム名オプション。オプションを選択してカラム名を含むファイルの先頭行を指定します。
コードページ	参照データのコードページ。デフォルトのコードページは UTF-8 です。

4. **【OK】** をクリックしてファイルをエクスポートします。

管理されていない参照テーブルでの編集の有効化と無効化

管理されていない参照テーブルのデータ値とカラムに対する更新を有効または無効にすることができます。

参照テーブルの編集可能なステータスを変更する前にテーブルを保存してください。

1. **【開く】** をクリックします。
アセットライブラリが開きます。
2. **【参照テーブル】** アセットカテゴリを選択し、参照テーブル名を選択します。
参照テーブルが読み取り専用モードで開きます。
3. 現在のバージョンの参照テーブルを編集するには、**【編集】** をクリックします。
バージョン管理されているモデルリポジトリの参照テーブルを編集するには、参照テーブルをチェックアウトします。
4. **【アクション】** メニューを開き、**【カラムプロパティの変更】** を選択します。
【カラムプロパティの変更】 ダイアログボックスが開きます。
5. **【編集可能】** オプションを選択または解除します。

参照テーブル値のリフレッシュ

Analyst ツールに表示される参照テーブルの値は、更新する必要がある場合があります。

参照テーブル値を再ロードするには、**【アクション】** > **【リフレッシュ】** をクリックします。Analyst ツールにより、現行バージョンのデータベースからデータが取得されます。

監査証跡イベント

ユーザーが参照テーブルに加えた変更の監査証跡を表示することができます。監査証跡イベントを表示するには、参照テーブルの監査証跡ビューを使用します。Analyst ツールに表示される監査証跡イベントは、フィルタすることができます。

以下の表に、指定可能なフィルタオプションの説明を示します。

オプション	説明
日付	表示するアクションの開始日と終了日。カレンダーオプションを使用して日付を設定します。
タイプ	監査証跡イベントのタイプ。以下のイベントタイプを表示することができます。 <ul style="list-style-type: none">- データ: 参照テーブルのデータ値に関連するイベント。イベントには、行の追加、行の削除、行の更新を行う操作が含まれます。- メタデータ: 参照テーブルメタデータに関連するイベント。イベントには、参照テーブルを作成する操作、カラムを追加または削除する操作、および参照テーブルにチェックインする操作などが含まれます。 注: データイベントとメタデータイベントを同時に表示することはできません。
ユーザー	参照テーブルを編集したユーザー。フィルタは、ユーザーのフルネームとログイン名を表示します。
ステータス	監査証跡ログイベントのステータス。ステータスは、参照テーブルエディタで実行したアクションに対応します。例えば、ユーザーが参照テーブルを作成したことや行を追加したことがステータスで示されます。

監査証跡ログイベントには、監査証跡のコメントや、挿入、更新、または削除したカラム値も含まれます。

監査証跡イベントの表示

監査証跡イベントを表示して、ユーザーが参照テーブルに対して行った更新を確認します。監査証跡イベントは読み取り専用モードで表示することができます。

1. **【開く】** をクリックします。
アセットライブラリが開きます。
2. **【参照テーブル】** アセットカテゴリを選択し、参照テーブル名を選択します。
参照テーブルが読み取り専用モードで開きます。
3. **【監査証跡】** をクリックします。
4. フィルタオプションを設定します。
更新日付、更新タイプ、更新ステータス、および更新を実行したユーザーの名前でフィルタできます。
5. **【表示】** をクリックします。
指定したフィルタオプションのログイベントが表示されます。

参照テーブルのルールおよびガイドライン

Analyst ツールを使用して参照テーブルを操作する場合は、以下のルールおよびガイドラインを使用します。

- Oracle、IBM DB2、または Microsoft SQL Server のデータベースから参照テーブルをインポートする際、テーブル、ビュー、スキーマ、シノニム、カラム名のいずれかが小文字のみまたは大文字と小文字の混合の場合、Analyst ツールではプレビューできません。

大文字と小文字が区別されるデータベースに格納されたテーブルのデータをプレビューするには、データベース接続の「大文字と小文字が混在する識別子をサポート」属性を true に設定します。

- 参照テーブルを、1 つの形式の推測されるカラムパターンから作成する場合、Analyst ツールは異なる形式のカラムパターンを使用して参照テーブルに入力します。

例えば、カラムパターン X(5)の参照テーブルを作成すると、Analyst ツールは以下の形式のカラムパターンで参照テーブルを表示します。XXXXX。

- Oracle データベーステーブルをインポートする場合は、テーブル内の VARCHAR2 カラムの長さを確認してください。Analyst ツールは、長さが 1000 を超える VARCHAR2 カラムが含まれた Oracle データベーステーブルをインポートできません。
- 参照テーブルを読み込むには、テーブルデータ値を格納するデータベースへの接続に対する実行権限が必要です。例えば、参照データウェアハウスにデータ値が格納されている場合、参照データウェアハウスへの接続に対する実行権限が必要です。読み取りモードまたは書き込みモードで参照テーブルにアクセスする実行権限が必要です。データベース接続権限は、データベースのすべての参照データに適用されます。
- 参照テーブルを指定するトランスフォーメーションでマッピングを実行すると、マッピングでは、モデルリポジトリにある現在のバージョンの参照テーブルが使用されます。トランスフォーメーションを設定する際に、履歴バージョンの参照テーブルを選択することはできません。

別のユーザーが、Developer tool の同時セッションで以前のバージョンの参照テーブルをリストアした場合、参照テーブルのバージョンはセッション間で同一ではなくなります。この参照テーブルを使用するマッピングを設定および実行すると、現在のセッションで現在の参照テーブルのバージョンが識別されなくなるため、マッピングに失敗する可能性があります。マッピングで現在の参照テーブルが確実に使用されるようにするために、マッピングを実行する前にモデルリポジトリを更新します。

第 3 章

Developer ツールの参照データ

この章では、以下の項目について説明します。

- [Developer ツール参照データの概要, 29 ページ](#)
- [参照データとトランスフォーメーション, 29 ページ](#)
- [バージョン管理されているモデルリポジトリの参照データオブジェクトの操作, 30 ページ](#)
- [参照テーブル, 31 ページ](#)
- [コンテンツセット, 35 ページ](#)

Developer ツール参照データの概要

Developer ツールでは、参照データオブジェクトの設定プロパティを作成、更新、表示できます。

Developer ツールを使用して、以下のタイプのオブジェクトを作成および更新します。

参照テーブル

参照テーブルには、データ値のセットの標準バージョンと代替バージョンが含まれています。Developer ツールで参照テーブルをトランスフォーメーションに追加し、ソースのデータ値が正確で、正しい形式であることを確認します。

コンテンツセット

コンテンツセットは、リポジトリまたはファイル内の参照データの値を指定するモデルリポジトリオブジェクトです。コンテンツセットには、データ品質トランスフォーメーションで検索操作の実行に使用できる、さまざまなタイプの参照データが含まれます。

Developer ツールでは、アドレス参照データファイルや ID ポピュレーションファイルを使用することもできます。アドレスバリデータトランスフォーメーションを設定する場合は、アドレス参照データファイルを選択します。ID 照合分析用に一致トランスフォーメーションを設定する場合は、ID ポピュレーションファイルを選択します。

参照データとトランスフォーメーション

複数のトランスフォーメーションでは、データ品質タスクを実行するために参照データを読み取ります。

参照データを読み取ることができるトランスフォーメーションは次のとおりです。

- アドレスバリデータ。アドレス参照データを読み取り、住所に誤りがないかどうかを確認します。

- 大文字小文字変換プログラム。参照データテーブルを読み取り、大文字と小文字を変更する必要がある文字列を識別します。
- 分類子。コンテンツセットデータを読み取り、文字列の情報のタイプを識別します。
- 比較。重複分析の際に ID ポピュレーションのデータを読み取ります。
- ラベラ。コンテンツセットデータを読み取り、ラベル文字列を識別します。
- 一致。重複分析の際に ID ポピュレーションのデータを読み取ります。
- パーサー。コンテンツセットデータを読み取り、それが含む情報に基づいて文字列をパースします。
- 標準化。参照データテーブルを読み取り、文字列を共通の形式に標準化します。

Data Quality Content Installer ファイルセットには、インポートできる Informatica 参照データオブジェクトが含まれます。

バージョン管理されているモデルリポジトリの参照データオブジェクトの操作

バージョン管理されているモデルリポジトリの参照テーブルまたはコンテンツセットを操作する場合、リポジトリでは、オブジェクトにバージョン管理が適用されます。ユーザーは、オブジェクトにバージョン管理を適用するため、モデルリポジトリに対してオブジェクトをチェックインします。

参照テーブルまたはコンテンツセットのバージョン管理が行われていない場合は、バージョン管理システムの外部でオブジェクトを開いて更新できます。オブジェクトを開くと、モデルリポジトリでオブジェクトがロックされるため、別のユーザーはそのオブジェクトを操作できなくなります。

参照テーブルまたはコンテンツセットのバージョン管理が行われている場合は、オブジェクトを読み取り専用モードで開きます。オブジェクトで作業するには、モデルリポジトリからオブジェクトをチェックアウトします。または、オブジェクトをチェックアウトしてから開きます。オブジェクトをチェックインして、最新の変更が含まれるオブジェクトのバージョンを作成します。

参照データオブジェクトのチェックアウト

ユーザーがモデルリポジトリにチェックインした参照テーブルまたはコンテンツセットで作業するには、リポジトリからオブジェクトをチェックアウトします。

1. オブジェクトエクスプローラで、参照テーブルまたはコンテンツセットを参照します。
2. オブジェクト名を右クリックし、**【開く】**をクリックします。
オブジェクトが読み取り専用モードで開きます。
3. オブジェクト名を右クリックし、**【チェックアウト】**をクリックします。
オブジェクトを編集できます。

参照データオブジェクトのチェックイン

モデルリポジトリからチェックアウトした参照テーブルまたはコンテンツセットでの作業が終了したら、オブジェクトをチェックインします。

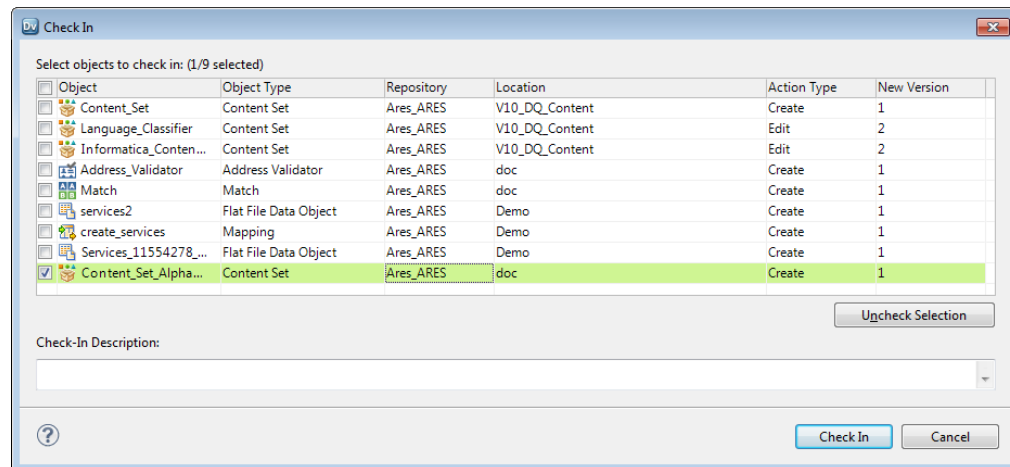
現在チェックアウトされているオブジェクトのリストを表示するには、参照テーブルエディタの下にある **【チェックアウトされたオブジェクト】** タブを開きます。

1. 参照テーブルまたはコンテンツセットに対して行われた変更をすべて保存します。

- オブジェクトエクスプローラで、参照テーブルまたはコンテンツセットを参照します。
- オブジェクト名を右クリックし、**[チェックイン]** をクリックします。

[チェックイン] ダイアログボックスが開きます。

次の図に **[チェックイン]** ダイアログボックスを示します。



- リポジトリでチェックインするオブジェクトを 1 つ以上選択します。

注: 現在のセッションで開かれていないオブジェクトをチェックインできます。チェックアウトされた状態になっている任意のオブジェクトをチェックインすることができます。

- 必要に応じて、操作の説明を入力します。
- [チェックイン]** をクリックします。

チェックイン操作では、オブジェクトのバージョン番号が更新されます。初めてオブジェクトをチェックインする場合には、モデルリポジトリで、オブジェクトのバージョン 1 が作成されます。

参照テーブル

Developer tool で参照テーブルをトランスフォーメーションに追加します。トランスフォーメーションを設定して、入力データで参照テーブル値を検索し、該当する有効な値を参照テーブルから出力として書き込むようにします。

Developer tool で参照テーブルを作成するには、以下の方法のいずれかを使用します。

- 空の参照テーブルを作成し、データ値を入力します。
- フラットファイルのデータから参照テーブルを作成する。
- データベーステーブル、シノニム、またはビューのデータから参照テーブルを作成します。

参照テーブルデータのプロパティ

Developer ツールで参照テーブルデータおよびメタデータのプロパティを表示できます。モデルリポジトリから参照テーブルを開くときに、Developer ツールには、プロパティが表示されます。

参照テーブルには全般プロパティおよびカラムのプロパティが表示されます。Developer ツールで参照テーブルのプロパティを表示できます。Analyst ツールで参照テーブルプロパティを表示および編集できます。

以下の表は、参照テーブルの全般プロパティの説明です。

プロパティ	説明
名前	参照テーブルの名前。
説明	参照テーブルのオプションの説明。

以下のテーブルは、参照テーブルのカラムのプロパティの説明です。

プロパティ	説明
有効	有効な参照データを含むカラムを特定します。
名前	各カラムの名前。
データタイプ	各カラムのデータのデータタイプ。
精度	各カラムの精度。
スケール	各カラムのスケール。
説明	カラムのコンテンツの説明。参照テーブルを作成したときは、オプションで説明を追加できます。
低レベルの説明のカラムを含めます	参照テーブルにカラムデータの説明のカラムが含まれることを示しています。
デフォルト値	カラムのフィールドのデフォルト値。参照テーブルを作成する場合は、オプションでデフォルト値を追加できます。
接続名	参照テーブルデータ値を含むデータベースへの接続の名前。

参照テーブルオブジェクトの作成

空の参照テーブルを作成して値を手動で追加する場合は、この方法を選択します。

1. Developer ツールのメニューから **【ファイル】** > **【新規】** > **【参照テーブル】** を選択します。
2. 新しいテーブルウィザードで、**【空の参照テーブル】** を選択します。
3. テーブルの名前を入力します。
4. テーブルのメタデータを格納するプロジェクトを選択します。

[場所] フィールドで、**【参照】** をクリックします。 **【場所の選択】** ダイアログボックスが開き、リポジトリ内のプロジェクトが表示されます。必要なプロジェクトを選択します。

【次へ】 をクリックします。

5. テーブルにカラムを 2 つ以上追加します。 **【新規】** オプションをクリックしてカラムを作成します。

以下の表に、各カラムのプロパティを示します。

プロパティ	デフォルト値
名前	column

プロパティ	デフォルト値
データ型	string
精度	10
位取り	0
説明	空:オプションのプロパティ。

- 有効な値を含むカラムを選択します。作成するカラムの順序は変更が可能です。
- 次のテーブルは、オプションのプロパティの説明です。

プロパティ	デフォルト値
行レベルの説明のカラムを含めます	オフ
監査のメモ	空
デフォルト値	空

【完了】をクリックします。

Developer ツールのワークスペースで参照テーブルが開きます。

フラットファイルからの参照テーブルの作成

フラットファイルに保存されたデータから参照テーブルを作成できます。

- Developer ツールのメニューから【ファイル】 > 【新規】 > 【参照テーブル】を選択します。
- 新しいテーブルウィザードで、【フラットファイルからの参照テーブル】を選択します。
- テーブルのデータソースとして使用するファイルを参照します。
- テーブルの名前を入力します。
- テーブルのメタデータを格納するプロジェクトを選択します。

【場所】フィールドで、【参照】をクリックします。【場所の選択】ダイアログボックスが開き、リポジトリ内のプロジェクトが表示されます。必要なプロジェクトを選択します。

【次へ】をクリックします。

- コードページとして UTF-8 を設定します。
- フラットファイルで使用する区切り文字を指定します。
- フラットファイルにカラム名が含まれている場合は、ファイルの 1 行目からカラム名をインポートするオプションを選択します。
- 次のテーブルは、オプションのテーブルプロパティの説明です。

プロパティ	デフォルト値
テキスト修飾子	引用符なし
次の行からインポートを開始	行 1

プロパティ	デフォルト値
行の区切り文字	\012 LF (\n)
連続した区切り文字を 1 文字として扱う	オフ
エスケープ文字	空
データ内のエスケープ文字を維持	オフ
プレビューする最大行数	500

【次へ】をクリックします。

- 有効な値を含むカラムを選択します。
- 次のテーブルは、オプションのプロパティの説明です。

プロパティ	デフォルト値
行レベルの説明のカラムを含めます	オフ
監査のメモ	空
デフォルト値	空
プレビューする最大行数	500

【完了】をクリックします。

Developer ツールのワークスペースで参照テーブルが開きます。

リレーショナルソースからの参照テーブルの作成

リレーショナルテーブル、シノニム、またはビューから参照テーブルを作成できます。

管理されている参照テーブルを作成する場合は、カラムデータを参照データウェアハウスにインポートします。管理されていない参照テーブルを作成する場合は、カラムデータを格納するデータベーステーブルを指定します。管理されている参照テーブルは、ODBC 接続またはネイティブ接続から作成できます。管理されていない参照テーブルは、ネイティブ接続から作成できます。

参照テーブルを作成する前に、参照データを含むデータベースへの接続が Informatica ドメインに含まれていることを確認します。

データベース接続は接続エクスプローラで設定できます。Developer tool に接続エクスプローラが表示されていない場合は、Developer tool のメニューから **【ウィンドウ】 > 【ビューの表示】 > 【接続エクスプローラ】** を選択します。

リレーショナルソースからの参照テーブルの作成

参照テーブルを作成するには、データベースに接続して参照データを含むテーブルを選択します。

- Developer tool のメニューから **【ファイル】 > 【新規】 > 【参照テーブル】** を選択します。
- テーブル作成ウィザードで、**【リレーショナルソースからの参照テーブル】** を選択します。

【次へ】をクリックします。

3. データベース接続を選択します。

【接続】フィールドで、【参照】をクリックします。【接続の選択】ダイアログボックスが開き、使用可能なデータベース接続が表示されます。

接続を選択する場合は【OK】をクリックします。

4. データベースリソースを選択します。

【リソース】フィールドで、【参照】をクリックします。【リソースの選択】ダイアログボックスが開き、データベース接続上のリソースが表示されます。データベースを参照し、データベーステーブル、シノニム、またはビューを選択します。

必要に応じて、リソースのエンティティ情報をプレビューすることもできます。

5. テーブルの名前を入力します。

6. 参照テーブルオブジェクトの場所を選択します。

【場所】フィールドで、【参照】をクリックします。【場所の選択】ダイアログボックスが開き、リポジトリ内のプロジェクトが表示されます。

場所を選択し、【次へ】をクリックします。

7. 参照データウェアハウスにデータを格納しない参照テーブルを作成する場合は、【未管理のテーブル】を選択します。

ユーザーが管理されていない参照テーブルを編集できるようにするには、【編集可能】オプションを選択します。

【次へ】をクリックします。

8. 有効な値を含むカラムを選択します。

9. 次の表に、指定できるオプションのプロパティを示します。

プロパティ	デフォルト値
行レベルの説明のカラムを含めます	オフ
説明	オフ
デフォルト値	空
監査のメモ	空
プレビューする最大行数	500

10. 【完了】をクリックします。

コンテンツセット

コンテンツセットは、他の参照データオブジェクトのデータまたはメタデータを格納するモデルリポジトリオブジェクトです。コンテンツセットには、文字セット、パターンセット、トークンセット、正規表現、確率的なモデル、および分類子モデルを含めることができます。コンテンツセットを使用して、単一プロジェクト、情報タイプ、ビジネス目的に関連する参照データオブジェクトを定義し、整理します。

Developer ツールには、モデルリポジトリに表示されないシステム定義の文字セットやトークンセットが含まれます。システム定義オブジェクトを表示および使用するには、ラベラトランスフォーメーション、パーサートランスフォーメーション、または標準化トランスフォーメーションでストラテジを設定します。

文字セット

文字セットには、特定の文字および文字範囲を識別する式が含まれます。文字セットは、文字ラベル適用モードを使用するラベラトランスフォーメーションで使用できます。

文字範囲は、連続する文字コードの範囲を指定します。例えば、文字範囲"[A-C]"は大文字の"A"、"B"、および"C"に一致します。この文字範囲は、小文字の"a"、"b"、または"c"には一致しません。

文字セットを使用すると、ラベル適用操作の一部として特定の文字または文字範囲を識別できます。例えば、電話番号が格納されたカラムのすべての数字にラベルを適用することができます。数字にラベルを適用した後、パーサトランスフォーメーションでパターンを識別し、問題のあるパターンを別の出力ポートに書き込むことができます。

文字セットのプロパティ

文字セットの文字ラベル適用操作に関するプロパティを設定します。

次の表に、ユーザー定義の文字セットのプロパティを示します。

プロパティ	説明
ラベル	文字セットに一致するデータに対してラベラトランスフォーメーションで適用するラベルを定義します。
標準モード	開始範囲と終了範囲のフィールドを含む簡単な編集ビューを使用できます。
開始範囲	文字範囲の最初の文字を指定します。
終了範囲	文字範囲の最後の文字を指定します。1 文字の範囲を指定する場合は、このフィールドを空白のままにします。
詳細モード	範囲文字と区切り文字を使用して文字範囲を手動で入力できる詳細な編集ビューを使用できます。
範囲文字	文字範囲を示す記号を一時的に変更します。範囲文字は、文字セットを閉じるとデフォルトの文字に戻ります。
区切り文字	文字範囲を区切る記号を一時的に変更します。区切り文字は、文字セットを閉じるとデフォルトの文字に戻ります。

分類子モデル

分類子モデルは、入力文字列を分析し、文字列が含む可能性の高い情報のタイプを決定します。分類子モデルは分類子トランスフォーメーションで使用します。

分類子モデルには、参照データ行とラベル値が含まれます。行は、分類子トランスフォーメーションに接続する可能性のあるポートの入力データを表します。ラベル値は、データ行に含まれる情報のタイプを説明します。分類子モデルを設定するときに、モデルの参照データの各行にラベルを割り当てます。

参照データ行を分類子モデルのラベルにリンクするには、モデルをコンパイルします。コンパイル処理で、データ行とラベル値間の一連の論理的な関連付けが生成されます。モデルを読み取るマッピングを実行すると、データ統合サービスがモデルのロジックを分類子トランスフォーメーションの入力データに適用します。データ統合サービスは、入力データの各フィールドの情報を最も正確に説明するラベルを返します。

Developer ツールで分類子モデルを作成します。モデルリポジトリは、分類子モデルオブジェクトを格納します。Developer ツールは、データ行、ラベル、およびコンパイルデータを、Informatica ディレクトリ構造内のファイルに書き込みます。

パターンセット

パターンセットには、トークンラベル適用操作の出力のデータパターンを識別する式が含まれます。パターンセットを使用すると、トークン化されたデータの出力ポートを分析し、一致する文字列を1つ以上の出力ポートに書き込むことができます。パターンセットは、パターン解析モードを使用するパーサートランスフォーメーションで使用します。

例えば、名前と頭文字を識別するパターンセットを使用するようにパーサートランスフォーメーションを設定することができます。このトランスフォーメーションでは、パターンセットを使用して、トークンラベル適用モードのラベラトランスフォーメーションの出力を分析します。出力に含まれる名前と頭文字を別のポートに書き込むようにパーサートランスフォーメーションを設定できます。

パターンセットのプロパティ

パターンセットのパターンに関するプロパティを設定します。

次の表に、ユーザー定義のパターンセットのプロパティを示します。

プロパティ	説明
パターン	パターンパーサーで検索するパターンを定義します。パターンは、1つのパターンセットに複数入力することができます。ワイルドカード、文字、および文字列の組み合わせで構成されるパターンを入力できます。

確率モデル

確率モデルは、入力データ値を分析し、値が含む可能性の高い情報のタイプを決定します。確率モデルは、ラベラトランスフォーメーションおよびパーサートランスフォーメーションで使用します。

確率モデルには、参照データ値とラベル値が含まれます。参照データ値は、トランスフォーメーションに接続する入力ポートのデータを表します。ラベル値は、参照データ値に含まれる情報のタイプを説明します。モデルの各参照データ値にラベルを割り当てます。

確率モデルのラベルに参照データ値をリンクするには、モデルをコンパイルします。コンパイル処理では、データ値とラベル間の、一連の論理的な関連付けが生成されます。モデルを読み取るマッピングを実行すると、データ統合サービスはモデルのロジックをトランスフォーメーション入力データに適用します。データ統合サービスは、入力データ値を最も正確に描写しているラベルを返します。

Developer tool で確率モデルを作成します。モデルリポジトリは、確率モデルオブジェクトを格納します。Developer ツールは、データ値、ラベル、およびコンパイルデータを、Informatica ディレクトリ構造内のファイルに書き込みます。

正規表現

コンテンツセットのコンテキストにおいては、正規表現とは、解析操作やラベル適用操作で利用できる式のことを指します。正規表現を使用して入力データの1つ以上の文字列を識別できます。正規表現は、トークン解析モードを使用するパーサートランスフォーメーションで使用できます。また、トークンラベル適用モードを使用するラベラトランスフォーメーションでも使用できます。

パーサートランスフォーメーションでは、正規表現を使用して入力データのパターンを一致させ、一致するすべての文字列を1つ以上の出力に解析します。例えば、正規表現を使用して入力データに含まれるすべての電子メールアドレスを識別し、電子メールアドレスの構成要素ごとに異なる出力に解析することができます。

ラベラトランスフォーメーションでは、正規表現を使用して入力パターンを一致させ、1つのラベルを作成します。正規表現の出力が複数になる場合、複数のラベルは生成されません。

正規表現のプロパティ

プロパティは、正規表現が出力文字列を識別して書き込む方法を決定します。

次の表に、ユーザー定義の正規表現のプロパティを示します。

プロパティ	説明
出力の数	正規表現が書き込む出力ポートの数を定義します。
正規表現	パーサートランスフォーメーションが文字列の一致に使用するパターンを定義します。
テスト表現	正規表現をテストするために入力するデータです。このフィールドにデータを入力すると、正規表現に一致する文字列が強調表示されます。
次の式	正規表現に一致する次の文字列に移動し、その文字列のフォントを太字にします。
前の式	正規表現に一致する前の文字列に移動し、その文字列のフォントを太字にします。

トークンセット

トークンセットには、特定のトークンを識別する式が含まれます。トークンセットは、トークンラベル適用モードを使用するラベラートランスフォーメーションで使用できます。また、トークン解析モードを使用するパーサートランスフォーメーションでも使用できます。

トークンセットを使用すると、ラベル適用操作や解析操作の一部として特定のトークンを識別できます。例えば、トークンセットを使用して、"AccountName@DomainName"という形式のすべての電子メールアドレスにラベルを適用できます。トークンにラベルを適用したら、パーサートランスフォーメーションを使用して、指定した出力ポートに電子メールアドレスを書き込むことができます。

トークンセットのプロパティ

トークンセットのラベル適用操作に関するプロパティを設定します。

次の表に、ユーザー定義の文字セットのプロパティを示します。

プロパティ	トークンセットのモード	説明
名前	該当なし	トークンセットの名前を定義します。
説明	該当なし	トークンセットの説明を記述します。
トークンセットのオプション	該当なし	トークンセットが正規表現モードと文字モードのどちらを使用するかを定義します。
ラベル	正規表現	トークンセットに一致するデータに対してラベラートランスフォーメーションで適用するラベルを定義します。

プロパティ	トークンセットのモード	説明
正規表現	正規表現	ラベラトランスフォーメーションが文字列の一致に使用するパターンを定義します。
テスト表現	正規表現	正規表現をテストするために入力するデータです。このフィールドにデータを入力すると、正規表現に一致する文字列が強調表示されます。
次の式	正規表現	正規表現に一致する次の文字列に移動し、その文字列のフォントを太字にします。
前の式	正規表現	正規表現に一致する前の文字列に移動し、その文字列のフォントを太字にします。
ラベル	文字	文字セットに一致するデータに対してラベラトランスフォーメーションで適用するラベルを定義します。
標準モード	文字	開始範囲と終了範囲のフィールドを含む簡単な編集ビューを使用できます。
開始範囲	文字	文字範囲の最初の文字を指定します。
終了範囲	文字	文字範囲の最後の文字を指定します。1 文字の範囲を指定する場合は、このフィールドを空白のままにします。
詳細モード	文字	範囲文字と区切り文字を使用して文字範囲を手動で入力できる詳細な編集ビューを使用できます。
範囲文字	文字	文字範囲を示す記号を一時的に変更します。範囲文字は、文字セットを閉じるとデフォルトの文字に戻ります。
区切り文字	文字	文字範囲を区切る記号を一時的に変更します。区切り文字は、文字セットを閉じるとデフォルトの文字に戻ります。

確率的なモデルと分類子モデルのルールとガイドライン

モデルリポジトリの各確率的なモデルと分類子モデルは、Informatica ディレクトリ構造内のファイル特定します。ファイルには、Developer ツールでモデルに追加するデータ値とラベルが含まれます。ファイルには、データ値とラベル間の関連付けを定義するコンパイルロジックも含まれます。

確率的なモデルまたは分類子モデルで作業する場合は、次のルールとガイドラインに従います。

- モデルを含むマッピングを実行する場合、データ統合サービスは、コンパイルされたモデルのロジックをトランスフォーメーション入力データに適用します。データ統合サービスは、マッピングを実行するときに、モデル内のデータ値やラベルを読み取りません。
- オプションで、データ値やラベルを確率的なモデルや分類子モデルから削除できます。例えば、モデルから機密データまたは独自データを削除することができます。個々のデータ値やラベルは、Developer ツールで削除できます。すべてのデータ値やラベルは、モデルリポジトリからモデルをエクスポートするときに削除できます。

注: モデルからすべてのデータ値とラベルを削除すると、モデルをコンパイルできません。

- 1 つ以上のデータ値またはラベルをモデルから削除すると、コンパイルされたモデルのロジックは、モデルファイル内の現在のデータを表さなくなります。モデルのロジック、およびデータ値とラベルを同期するには、モデルをもう一度コンパイルします。現在のモデルのロジックを維持する場合は、モデルをコンパイルしないでください。
- 分類子モデルや確率的なモデルのデータを保護するには、Informatica ディレクトリ構造内のモデルファイルをバックアップします。モデルからすべてのデータ値とラベルを削除する前に、ファイルをバックアップします。
- コンテンツ管理サービスホストマシンでモデルファイルを検索します。

確率的なモデルファイルの、デフォルトの場所とファイル名拡張子は次のとおりです。

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

分類子モデルファイルの、デフォルトの場所とファイル名拡張子は次のとおりです。

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

- Informatica インストールをアップグレードする場合、マッピングでモデルを使用する前に、確率的なモデルと分類子モデルのコンパイルが必要になる場合があります。モデルにデータが何も含まれない場合は、Informatica ディレクトリ構造内の現在のファイルを、データを含むバックアップファイルに置き換えます。

分類子モデルおよび確率モデル内のラベルの管理

確率モデルまたは分類子モデル内のラベルを確認および更新するには、**【ラベルの管理】** ダイアログボックスを使用します。

- 分類子モデルの入っているコンテンツセットを開きます。
- モデル名を選択し、**【編集】** をクリックします。
- 【ラベルの管理】** ダイアログボックスを開きます。
ダイアログボックスにモデルのラベルがリスト表示されます。

関連項目：

- [「分類子モデルのラベルの管理」 \(ページ 46\)](#)
- [「確率モデルの設定」 \(ページ 60\)](#)

コンテンツセットの作成

コンテンツセットを作成して、単一プロジェクト、情報タイプ、ビジネス目的を参照する参照データオブジェクトを管理します。

1. **Object Explorer** ビューで、コンテンツセットを格納するプロジェクトまたはフォルダーを選択します。
2. **【ファイル】 > 【新規】 > 【コンテンツセット】** をクリックします。
3. コンテンツセットの名前を入力します。
4. 必要に応じて、**【参照】** を選択して、コンテンツセットのモデルリポジトリの場所を変更します。
5. **【完了】** をクリックします。

コンテンツセットでの参照データオブジェクトの作成

文字セット、パターンセット、トークンセット、正規表現、確率的なモデル、および分類子モデルをコンテンツセットで作成できます。

1. エディタでコンテンツセットを開き、**【コンテンツ】** ビューを選択します。
2. 参照データオブジェクトタイプを選択します。
3. **【追加】** をクリックします。
4. 参照データオブジェクトの名前を入力します。
必要に応じて、オブジェクトの説明を入力します。
5. 参照データオブジェクトのプロパティを設定します。
6. **【完了】** をクリックします。

ヒント: 参照データオブジェクトを 1 つのコンテンツセットから別のコンテンツセットにコピーできます。**【コピー先】** および **【貼り付け元】** オプションを使用して、コンテンツセットでオブジェクトのコピーを作成します。**Ctrl** キーを使用して、複数のコンテンツセットオブジェクトを選択できます。

第 4 章

分類子モデル

この章では、以下の項目について説明します。

- [分類子モデルの概要, 42 ページ](#)
- [分類子モデルの構造, 43 ページ](#)
- [分類子のスコア, 43 ページ](#)
- [分類子トランスフォーメーションの例, 43 ページ](#)
- [分類子モデルのオプション, 44 ページ](#)
- [分類子モデルの参照データ, 45 ページ](#)
- [分類子モデルのラベルデータ, 46 ページ](#)
- [分類子モデルの設定, 47 ページ](#)
- [フィルタ操作と検索操作, 51 ページ](#)
- [コピーと貼り付けの操作, 52 ページ](#)

分類子モデルの概要

分類子モデルは、コンテンツセット内の参照データオブジェクトです。複数の値を含む長い文字列を分析するには、分類子モデルを使用します。分類子モデルは、各文字列で最も共通性のあるタイプの情報を特定します。

分類子モデルを分類子トランスフォーメーションに追加します。このトランスフォーメーションは、分類子モデルのデータと各入力行のデータの間に共通する値を探します。このトランスフォーメーションは共通する値を使用して、各行が表す情報のタイプを分類します。

入力データに次のような特徴がある場合は、分類子モデルを使用します。

- 入力データにテキストがある。分類子モデルは、自然言語のプロセスをテキストデータに適用して、テキスト内の情報の種類を特定します。自然言語のプロセスでは、入力文字列の中の関係性のある単語が検出されます。自然言語のプロセスでは、関連性の無い単語は無視されます。
- 入力データの文字列に複数の値がある。たとえば、各フィールドに電子メールメッセージの内容が含まれているようなデータカラムを作成することができます。

分類子トランスフォーメーションが文字列のデータ型を読み取ります。このトランスフォーメーションでは、入力文字列の長さに制限を設けていません。

分類モデルを Developer ツールでコンパイルします。モデルをコンパイルするときは、モデル内の似たデータ値同士の関連付けを作成します。分類子トランスフォーメーションは、このコンパイルされたデータを使用して入力データの情報を検索します。

分類子モデルの構造

分類子モデルには、参照データ値とラベル値が含まれます。参照データ値は、分類するデータを表します。ラベル値は、分類子トランスフォーメーションがデータ内で識別可能な情報のタイプを指定します。

分類子モデルにもコンパイルデータが入っています。分類子トランスフォーメーションはコンパイルデータを使用して、モデルの参照データとトランスフォーメーションの入力データ間の類似度を測定します。分類子モデルをコンパイルするときに、コンパイルデータを作成または更新します。分類子トランスフォーメーションが入力データをモデルデータと比較するときに、入力データの各行を説明するラベル値を返します。

Developer tool は、参照データ値、ラベル値、およびコンパイルデータを Informatica ディレクトリ構造内のファイルに書き込みます。モデルリポジトリの分類子モデルオブジェクトがファイル名を格納します。分類子モデルを保存するときに、現在の参照データ値とラベル値をファイルに書き込みます。モデルをコンパイルするときに、ファイルのコンパイルデータを更新します。Developer tool のモデルのプロパティからファイル名を読み取ることができます。

分類子のスコア

分類子トランスフォーメーションは、入力データの各行を分類子モデル中の参照データのすべての行と比較します。このトランスフォーメーションで各比較のスコアを計算します。このスコアは、入力行と参照データ行間での類似の度合いを表しています。

分類子トランスフォーメーションを含んだマッピングを実行すると、マッピングが参照データの中のスコアの最も高い行を特定するラベルを返します。スコアは 0 から 1 までの範囲です。スコアが高ければ、入力データとモデルデータの間に強い一致があることを示しています。

分類子のスコアを確認して、ラベル出力が入力データの各行を正確に説明しているか確かめます。また、分類子モデルが入力データに対して適切であるかについても、スコアを見て確認することができます。トランスフォーメーション出力において低いスコアの割合が高い場合、この分類子モデルは不適切であると言えます。比較を改善するために、モデルを再度コンパイルしてください。コンパイルしたモデルでもスコアが良くならない場合は、トランスフォーメーションのモデルを置き換えてください。

分類子トランスフォーメーションの例

分類子モデルと分類子トランスフォーメーションを使用することで、電子メールのメッセージをその中のテキストに基づいて分類することができます。

例えば、ソフトウェアメーカーのカスタマサポートセンターのデータスチュワードで、顧客からサポートセンターに送られてきた電子メールのメッセージを確認しているとします。同社にはさまざまな国の顧客がいるため、サポートセンターにも各国の言語で記載された電子メールが届きます。各電子メールを顧客に適切に対応可能な部門に転送するために、電子メールを言語別に分類することにします。

電子メールを並び替えるには、以下の手順を実行します。

1. 電子メールのメッセージを単一のファイルまたはデータベーステーブルに書き込みます。
2. モデルリポジトリにファイルまたはデータベーステーブルを読み取るデータオブジェクトを作成します。
3. メッセージが使用する各言語のデータオブジェクトをモデルリポジトリに作成します。
4. 言語ごとにサンプルテキストが入った分類子モデルを作成します。

注: 電子メールメッセージのデータから採ったサンプルデータを、モデルのソースデータとして使用することができます。

5. 分類子モデルを再利用可能な分類子トランスフォーメーションに追加します。
6. 分類子トランスフォーメーションをメッセージデータに適用するマッピングを設定します。

マッピングを設定するには、以下の手順を実行します。

- 分類子トランスフォーメーションとデータオブジェクトをマッピングに追加します。
- 分類子トランスフォーメーションの入力ポートをソースのデータオブジェクトに接続します。
- 分類子トランスフォーメーションの出力ポートをターゲットのデータオブジェクトに接続します。

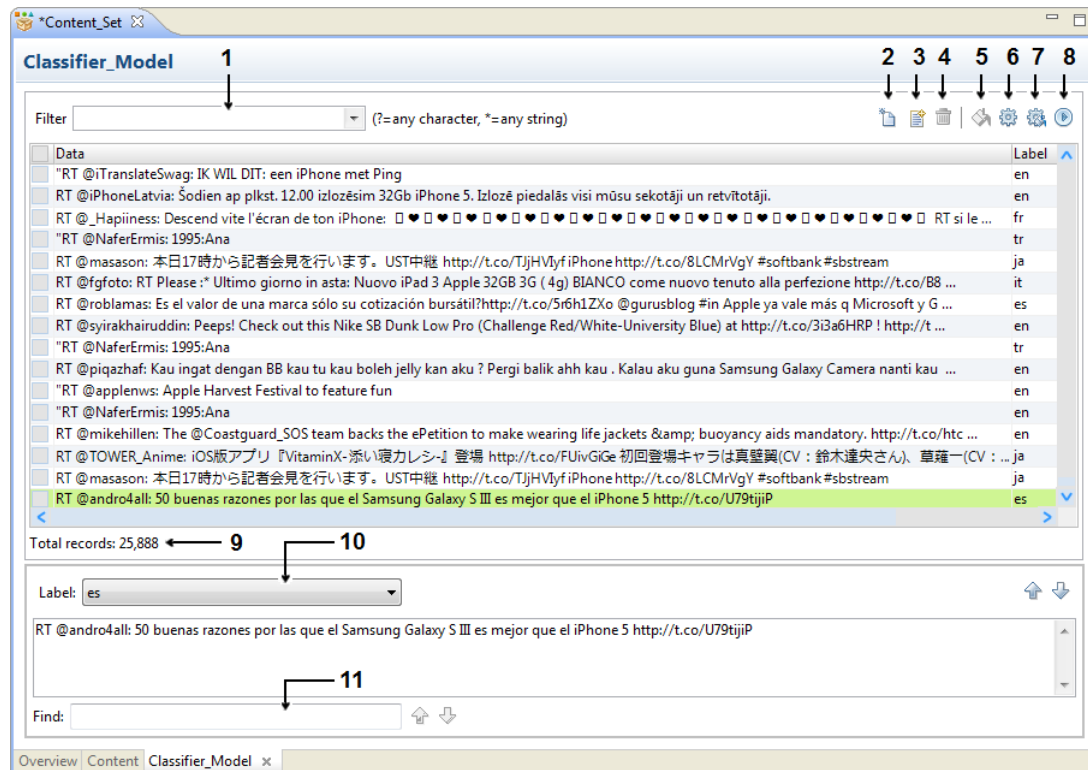
このマッピングを実行すると、分類子トランスフォーメーションが電子メールメッセージを分析して、電子メールテキストを正しいデータターゲットに書き込みます。データターゲットを各部門のチームメンバーと共有できます。

分類子モデルのオプション

Developer tool では、エディタの上部ペインと下部ペインに分類子モデルのデータが表示されます。上部ペインは各行に参照データと、ユーザーがデータに割り当てたラベルが表示されます。下部ペインには選択した行のコンテンツが表示されます。

上部ペインでは、参照データ行を確認したり、ラベルを使用していない行を特定したりします。下部ペインでは、行のコンテンツを確認したり、行にラベルを割り当てたりします。上部ペインの各行にはデータの約 100 文字が表示されます。下部ペインには選択した行の全データが表示されます。

次の図に、分類子モデルのエディタを示します。



エディタには次のオプションがあります。

1. [フィルタ] フィールド
データ値またはユーザーが指定したラベルに基づいて参照データ行のリストをフィルタリングします。
2. 行の追加
空白の参照データ行を挿入します。
3. データの追加
モデルリポジトリのデータオブジェクトからデータをインポートします。
4. 削除
ユーザーが選択した参照データ行を削除します。チェックボックスを使用して行を選択します。
5. ラベルの割り当て
ユーザーが選択した 1 つ以上の参照データ行にラベルを割り当てます。チェックボックスを使用して行を選択します。
6. プロパティの編集
分類子モデルのプロパティを表示します。
7. ラベルの管理
[ラベルの管理] ダイアログボックスが開きます。このダイアログボックスを使用して、分類子モデルのラベル値を追加または削除します。
8. コンパイル
分類子モデルをコンパイルします。
9. 総レコード数
分類子モデルの参照データの行数を示します。
10. [ラベル] フィールド
参照データの現在の行に適用可能なラベル値を表示します。
11. [検索] フィールド
参照データの現在の行内でユーザーが指定したデータ値を検索します。

分類子モデルの参照データ

分類子モデルには、文、段落、テキストページなどの参照データカラムが含まれています。参照データは、分類子トランスフォーメーションがマッピングで読み取ることのできるテキスト入力のさまざまな種類を表しています。モデルを作成する際は、マッピング実行時に検索されるはずのテキストの種類が参照データに含まれているか確認してください。

マッピングソースデータを使用して分類子モデルを作成することができます。ソースデータのサンプルを選択し、データサンプルをモデルにコピーします。

分類子モデルの参照データを使用するときは、次に挙げるルールとガイドラインを考慮してください。

- 参照データフィールドは任意の長さです。テキストページを各データフィールドに入力することができます。
- データオブジェクトから参照データをインポートします。
- 参照データの値は編集できません。ただし、データ行を削除することはできます。

- 分類子モデルをコンパイルする場合、コンパイルプロセスでは、参照データに含まれるすべての数値が無視されます。

分類子モデルのラベルデータ

分類子モデルには、参照データ行の情報のタイプを要約する説明ラベルが 1 つ以上含まれます。参照データの各行にラベルを割り当てます。

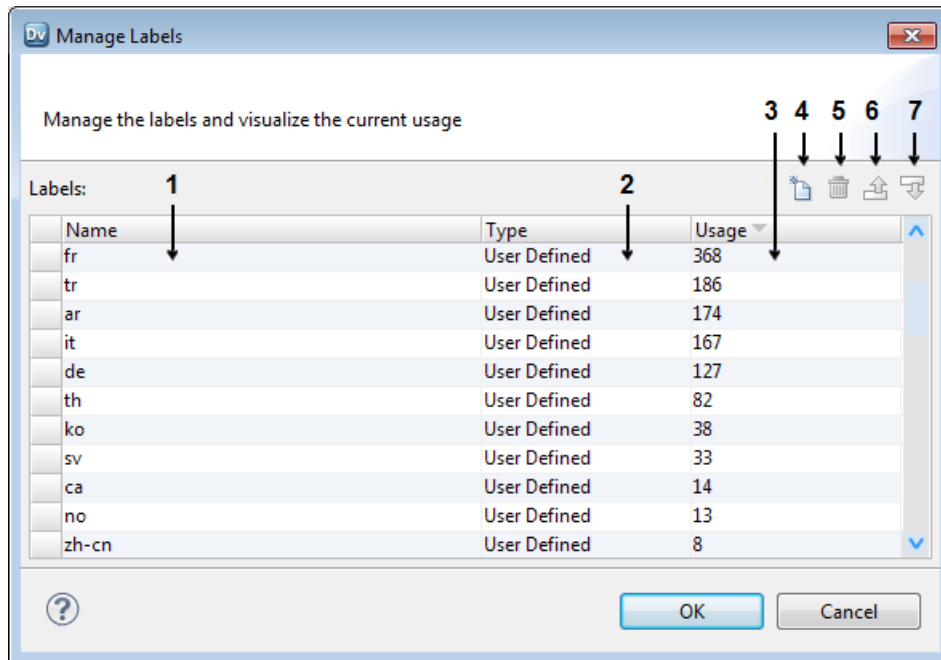
データソースからのデータを分類子モデルに追加するときに、カラムをラベルデータカラムとして指定できます。また、モデルでラベルを作成することもできます。

ラベルは、それが説明する参照データの値から独立しています。ラベルを使用する参照データ行を削除しても、そのラベルはモデルから削除されません。ラベルを削除しても、そのラベルに関連付けられていた参照データの値は削除されません。

分類子モデルのラベルの管理

[ラベルの管理] ダイアログボックスを使用して、分類子モデルのラベル値を確認および更新します。ラベル値を並べ替えて更新することもできます。

次の図に、**[ラベルの管理]** ダイアログボックスを示します。



[ラベルの管理] ダイアログボックスの要素は次のとおりです。

1. [名前] カラム。

分類子トランスフォーメーションが入力データ行に適用可能なラベル値が含まれます。ラベルを名前順に並べ替えることができます。

2. [タイプ] カラム。

ラベル値のソースを識別します。分類子モデルはすべてのラベルをユーザー定義の値として識別します。

3. [使用状況] カラム。
各ラベルを使用する参照データの行数を示します。ラベルを行数順に並べ替えることができます。
4. [追加] ボタン。
分類子モデルにラベルを追加します。行の [名前] カラムにラベル値を入力します。
注: ラベル値を更新するには、値をダブルクリックして必要な値を入力します。
5. [削除] ボタン。
分類子モデルからラベルを削除します。
6. 上矢印。
ラベルをダイアログボックスの 1 つ上の行に移動します。
7. 下矢印。
ラベルをダイアログボックスの 1 つ下の行に移動します。

分類子モデルの設定

分類子モデルを設定する手順では、最初に分類するデータを処理します。モデルに追加する参照データのコンテンツは、分類子トランスフォーメーションに接続するデータを示している必要があります。このトランスフォーメーションは、入力データのデータ値およびパターンを分類子モデルのデータ値およびパターンと比較します。

分類子トランスフォーメーションで使用可能な分類子モデルを作成するには、以下のタスクを実行します。

1. モデルに追加する参照データ値とラベル値を特定します。
分類するデータのフラグメントを使用できます。モデルリポジトリにデータフラグメントを読み取るデータオブジェクトを作成します。
2. コンテンツセットを作成し、分類子モデルをそのセットに追加します。
3. 参照データ値をモデルに追加します。
4. ラベル値をモデルに追加します。
モデルリポジトリのデータオブジェクトからデータをインポートできます。参照データの 1 つの行または 1 つのラベルを入力することもできます。
5. 参照データの各行にラベルを割り当てます。
単一の操作で、複数の行にラベルを割り当てることができます。
6. モデルをコンパイルします。
分類子モデルをコンパイルしたら、分類子トランスフォーメーションでそのモデルを使用できます。

分類子モデルの作成

データオブジェクトを分類子モデルデータのソースに使用します。

分類子モデルのパフォーマンスは、分類子トランスフォーメーションへの入力データをモデルの参照データのソースとして使用した場合に最適になります。

1. Object Explorer で、コンテンツセットを開くか作成します。
2. [コンテンツ] ビューを選択します。
3. [分類子モデル] を選択し、[追加] をクリックします。

[分類子モデル] ウィザードが開きます。

4. 分類子モデルの名前を入力します。

必要に応じて、このモデルの説明文を入力します。

5. モデルリポジトリを参照して、インポートするデータを含むデータオブジェクトを選択します。

ソーシャルメディアのデータオブジェクトは選択しないでください。

[次へ] をクリックします。

6. データオブジェクトのカラムを確認して、モデルに追加するカラムを 1 つ以上選択します。同じ操作で参照データカラムとラベルカラムを追加できます。

- データのカラムを参照データとしてインポートするには、カラム名を選択して [データ] をクリックします。

データカラムを複数選択してもかまいません。Developer tool が、ユーザーが選択したカラムのコンテンツを 1 つのカラムに統合します。

- データのカラムをラベル値としてインポートするには、カラム名を選択して [ラベル] をクリックします。

参照データ値とラベル値をインポートするときに、Developer tool が各行のラベルを同じ行の参照データ文字列に割り当てます。カラムを選択する前に、データをプレビューできます。モデルを作成した後にラベルの割り当てを変更できます。

[次へ] をクリックします。

7. データソースからインポートする行数を選択します。

デフォルトでは、Developer tool がデータソースからすべての行をインポートします。数を入力すると、モデルがデータセットの先頭から行を数えます。

8. [完了] をクリックしてモデルを保存します。

モデルを作成したら、ラベルの割り当てを確認してモデルをコンパイルします。

分類子モデルへのデータソースからのデータの追加

単一の操作で、参照データ値またはラベル値の複数の行を分類子モデルにインポートできます。

1. 分類子モデルの入っているコンテンツセットを開きます。

2. モデル名を選択し、[編集] をクリックします。

3. [データの付加] をクリックします。

[分類子モデル] ウィザードが開きます。

4. モデルリポジトリを参照して、インポートするデータを含むデータオブジェクトを選択します。

ソーシャルメディアのデータオブジェクトは選択しないでください。

[次へ] をクリックします。

5. データオブジェクトのカラムを確認して、モデルに追加するカラムを 1 つ以上選択します。同じ操作で参照データカラムとラベルカラムを追加できます。

- データのカラムを参照データとしてインポートするには、カラム名を選択して [データ] をクリックします。

データカラムを複数選択してもかまいません。Developer tool が、ユーザーが選択したカラムのコンテンツを 1 つのカラムに統合します。

- データのカラムをラベル値としてインポートするには、カラム名を選択して [ラベル] をクリックします。

参照データ値とラベル値をインポートするときに、Developer tool が各行のラベルを同じ行の参照データ文字列に割り当てます。カラムを選択する前に、データをプレビューできます。モデルを作成した後にラベルの割り当てを変更できます。

【次へ】をクリックします。

6. データソースからインポートする行数を選択します。

デフォルトでは、Developer tool がデータソースからすべての行をインポートします。数を入力すると、モデルがデータセットの先頭から行を数えます。

7. 【完了】をクリックしてモデルを保存します。

分類子モデルへの参照データ行の追加

分類子モデルに参照データの行を 1 つ追加できます。

1. 分類子モデルの入っているコンテンツセットを開きます。
2. モデル名を選択し、【編集】をクリックします。
3. 【行の追加】をクリックします。

Developer tool が参照データの現在の行の下に行を追加します。

4. 行に参照データ値を入力します。

Windows のショートカットを使用して行にデータを貼り付けることができます。

分類子モデルへのラベルの追加

分類子モデルに 1 つのラベルを追加できます。

1. 分類子モデルの入っているコンテンツセットを開きます。
2. モデル名を選択し、【編集】をクリックします。
3. 【ラベルの管理】ダイアログボックスを開きます。

ダイアログボックスにモデルのラベルがリスト表示されます。

4. 【新規】をクリックします。

Developer tool がラベルのリストの一番下に行を追加します。

5. [名前] カラムのデフォルト値をダブルクリックして、ラベル名を入力します。
6. 【OK】をクリックします。

ラベルを作成したら、参照データの 1 つ以上の行にラベルを割り当てることができます。【ラベルの管理】ダイアログボックスの [使用状況] カラムに、そのラベルを使用している行数が表示されます。

参照データ行へのラベルの割り当て

単一の操作で、1 つ以上の参照データ行にラベルを割り当てることができます。

1. モデルの入っているコンテンツセットを開きます。
2. モデル名を選択し、【編集】をクリックします。
3. 1 つ以上の参照データ行を選択します。チェックボックスオプションを使用して行を選択します。

注: フィルタオプションを使用すると、ユーザーが指定したデータ値を含むすべての行を表示できます。値を含む行をすべて選択するには、[すべて選択] チェックボックスオプションを使用します。

4. 【ラベルの割り当て】をクリックします。

Developer tool に分類子モデルのラベルがリスト表示されます。

5. ラベル値を選択して、**【割り当て】** をクリックします。

Developer tool がラベル値に基づいてユーザーが選択した参照データ行を更新します。

必要に応じてモデルをコンパイルし、ラベル名を分類子モデルのロジックに追加します。

未使用のラベル値の特定

【ラベルの管理】 ダイアログボックスを使用して、分類子モデルに未使用のままのラベル値がないか検索します。**【ラベルの管理】** ダイアログボックスに、分類子モデルのラベル値の使用状況データが表示されます。使用状況データを使用して、ラベル値を使用する参照データの行数を確認したり、未使用のラベル値を見つけたりします。

1. 分類子モデルの入っているコンテンツセットを開きます。
2. モデル名を選択し、**【編集】** をクリックします。
3. **【ラベルの管理】** ダイアログボックスが開きます。ダイアログボックスに分類子モデルのラベルがリスト表示されます。
4. 各ラベルの **【使用状況】** カラムのデータを確認します。

【使用状況】 カラムに、そのラベルを使用する参照データの行数が示されます。ラベル値が未使用の場合は、**【使用状況】** カラムの値がゼロになります。

分類子モデルからの行の削除

単一のアクションで、分類子モデルから参照データ行を 1 つ以上削除できます。

1. モデルの入っているコンテンツセットを開きます。
2. モデル名を選択し、**【編集】** をクリックします。
3. 1 つ以上の参照データ行を選択します。チェックボックスオプションを使用して行を選択します。
4. **【削除】** をクリックします。

Developer tool が分類子モデルからユーザーが選択した行を削除します。

操作を元に戻すには、キーボードの Ctrl+Z キーを押します。

分類子モデルからのラベルの削除

【ラベルの管理】 ダイアログボックスを使用して、分類子モデルからラベルを削除します。

1. モデルの入っているコンテンツセットを開きます。
2. モデル名を選択し、**【編集】** をクリックします。
3. **【ラベルの管理】** ダイアログボックスを開きます。
4. **【削除】** をクリックします。
5. **【はい】** をクリックして操作を確認します。

Developer tool がモデルからラベルを削除します。Developer tool が、ラベルを使用する参照データ行を削除することはありません。

6. **【OK】** をクリックしてダイアログボックスを閉じます。

操作を元に戻すには、キーボードの Ctrl+Z キーを押します。

分類子モデルのコンパイル

分類子モデルでラベルの値または参照データの値を編集するたびに、モデルをコンパイルする必要があります。モデルをコンパイルするときは、モデル内のコンパイルデータを更新します。

▶ コンパイルデータを更新するには、モデルを Developer ツールで開いて **【コンパイル】** をクリックします。

フィルタ操作と検索操作

フィルタオプションを使用して、ユーザーが指定した条件を満たす参照データ行を表示または非表示にします。フィルタを適用すると、分類子モデルが表示するデータ行に対して追加のアクションを実行できます。例えば、すべてのデータ行にラベル値を適用できます。

フィルタオプションを使用して、以下のタスクを実行します。

- ユーザーが入力した値を含む参照データ行を検索する。
- ユーザーが選択したラベルを使用する参照データ行を検索する。
- ラベルを使用しない参照データ行を検索する。

参照データの行内のデータ値を検索することもできます。

データ値を使用した参照データ行のフィルタリング

フィルタを使用して、1つ以上の参照データ行に予想どおりのデータ値が含まれていることを確認します。

1. 分類子モデルの入っているコンテンツセットを開きます。
2. モデル名を選択し、**【編集】** をクリックします。
3. **【フィルタ】** フィールドに値を入力します。

入力する値にワイルドカード文字を含めることもできます。

Developer tool に、フィルタテキストを含む参照データ行が表示されます。

ラベル値を使用した参照データ行のフィルタリング

フィルタを使用して、ユーザーが選択したラベルを使用する参照データ行を表示または非表示にします。

1. モデルの入っているコンテンツセットを開きます。
2. モデル名を選択し、**【編集】** をクリックします。
3. **【フィルタ】** メニューからラベル値を選択します。

Developer tool に、そのラベル値を使用する参照データ行が表示されます。

注: ラベルを使用していない参照データ行を検索するには、**【フィルタ】** メニューから **【ラベルなし】** オプションを選択します。

参照データ行内の値の検索

【検索】 フィールドを使用して、ユーザーが選択したデータ値を行内で検索します。

1. モデルの入っているコンテンツセットを開きます。
2. モデル名を選択し、**【編集】** をクリックします。
3. 参照データの行を選択します。

4. [検索] フィールドに値を入力します。
モデルに参照データ行内の値の出現箇所が強調表示されます。
5. 上矢印または下矢印を使用して、行内の値の他の出現箇所を確認します。

コピーと貼り付けの操作

分類子モデルを、モデルリポジトリの中のあるコンテンツセットから別のコンテンツセットにコピーすることができます。分類子モデルをコピーして、リソースを他の Developer ツールユーザーと共有します。

モデルを別のコンテンツセットにコピーすることも、モデルを現在のコンテンツセットにインポートすることもできます。リポジトリの中の複数のコンテンツセットから複数のモデルを 1 回の操作でインポートすることができます。

モデルをコピーする際に、コンテンツ管理サービスがモデルデータファイルのコピーをサービスマシンに作成します。各モデルが異なるデータファイルを使用します。

別のコンテンツセットへの分類子モデルのコピー

分類子モデルを、モデルリポジトリの中のあるコンテンツセットから別のコンテンツセットにコピーすることができます。分類子モデルをコピーする際は、モデルオブジェクトおよびコピー元とコピー先のコンテンツセットを指定します。

1. 分類子モデルの入っているコンテンツセットを開きます。
2. 分類子モデルを選択し、**[コピー先]** をクリックします。
3. モデルリポジトリの中を参照して、コンテンツセットを選択します。
現在のプロジェクトあるいは別のプロジェクトのコンテンツセットに分類子モデルをコピーすることができます。
4. **[OK]** をクリックします。
Developer ツールが、選択されたコンテンツセットに分類子モデルをコピーします。

別のコンテンツセットからの分類子モデルのインポート

分類子モデルを、モデルリポジトリであるコンテンツセットから別のコンテンツセットにインポートすることができます。分類子モデルをインポートする際は、1 つまたは複数のモデルオブジェクトおよびコピー元とコピー先のコンテンツセットを指定します。

1. 分類子モデルを入れるコンテンツセットを開きます。
2. 分類子モデルを選択し、**[貼り付け元]** をクリックします。
3. モデルリポジトリの中を参照して、分類子モデルを選択します。
現在のプロジェクトあるいは別のプロジェクトのコンテンツセットから分類子モデルを貼り付けることができます。
4. **[OK]** をクリックします。
Developer ツールが、現在のコンテンツセットに分類子モデルを貼り付けます。

第 5 章

確率モデル

この章では、以下の項目について説明します。

- [確率モデルの概要, 53 ページ](#)
- [確率モデルの構造, 54 ページ](#)
- [ラベラトランスフォーメーションの例, 54 ページ](#)
- [パーサートランスフォーメーションの例, 55 ページ](#)
- [確率モデルのオブション, 55 ページ](#)
- [確率モデルの参照データ, 58 ページ](#)
- [確率モデルのラベルデータ, 58 ページ](#)
- [確率モデルのプロパティ, 60 ページ](#)
- [確率モデルの設定, 60 ページ](#)
- [コピーと貼り付けの操作, 66 ページ](#)

確率モデルの概要

確率的なモデルは、コンテンツセットで作成する参照データオブジェクトです。確率的なモデルを使用して、複数のデータ値を含むデータ文字列を分析します。確率的なモデルは、文字列の中の各値の情報の種類を特定します。確率的なモデルをラベラトランスフォーメーションおよびパーサートランスフォーメーションに追加することができます。

ラベラトランスフォーメーションで確率モデルを使用して、説明的なラベルを入力文字列の中の各値に適用します。ラベラトランスフォーメーションでは、ラベルを単一の出力ポートに書き込みます。パーサートランスフォーメーションの確率的なモデルを使用して、入力文字列の各値を値の情報を表すポートに書き込みます。パーサートランスフォーメーションでは、情報タイプごとに出力ポートを作成します。

Developer tool で確率モデルを設計してコンパイルします。確率的なモデルを定義する際、一連のデータ行をモデルに追加し、各行内のそれぞれの値にラベルを割り当てます。確率モデルをコンパイルすると、Developer tool により、データ値と追加したラベルとの関連付けが作成されます。ラベラトランスフォーメーションとパーサートランスフォーメーションでは、自然言語のプロセスを使用して、確率モデルのデータと入力ポートのデータを比較します。

自然言語のプロセスでは、以下の方法を使用して、データ値内の情報のタイプを識別します。

- 自然言語のプロセスでは、似ているデータ値を認識して、同じラベルを値に適用できます。
- 自然言語のプロセスでは、データ値と文字列内の隣接する値を比較できます。自然言語のプロセスでは、値のシーケンスを分析して、各文字列の用法を把握し、文字列が表す情報タイプを確認できます。

確率モデルの構造

確率モデルには、参照データ値の行とラベル値の行があります。参照データ値は、トランスフォーメーションの入力データに表示される可能性があるさまざまな値を表します。ラベル値は、入力値に含まれる情報のタイプを識別します。

確率モデルは、また編集データを含みます。ラベラトランスフォーメーションとパーサートランスフォーメーションは、コンパイルデータを使用して、モデルの参照データとトランスフォーメーションの入力データ間の類似度を測定します。確率モデルをコンパイルするときに、コンパイルデータを作成または更新します。

1つのデータ行には、1つの値または複数の値を含めることができます。各データ行は異なる構造を持つことができます。同じラベルを、データ行の異なる値に割り当てられます。また、行内の複数の位置に表示される同一の値に異なるラベルを割り当てることもできます。マッピングの実行時、データ統合サービスは入力文字列内の値の相対位置を考慮します。確率モデルをコンパイルする前に、各ラベルを1つ以上のデータ値に割り当てます。

Developer tool は、参照データ値、ラベル値、およびコンパイルデータを Informatica ディレクトリ構造内のファイルに書き込みます。モデルリポジトリの確率モデルオブジェクトがファイル名を格納します。確率モデルを保存するときに、現在の参照データ値とラベル値をファイルに書き込みます。モデルをコンパイルするときに、ファイルのコンパイルデータを更新します。Developer tool のモデルのプロパティからファイル名を読み取ることができます。

注: 確率モデルの機能を最適化するには、各データ行に複数の参照データ値が含まれていることを確認します。各行の値の順序は、トランスフォーメーションの入力データに値が出現する順序とできるだけ一致させる必要があります。データ行に単一の参照データ値がある場合、ラベラトランスフォーメーションまたはパーサートランスフォーメーションが確率分析時に自然言語のプロセスを適用できません。

ラベラトランスフォーメーションの例

ある保険会社の顧客データベースには複数のデータエントリエラーがあります。あなたは保険会社でデータ管理を任されています。そこで、ラベラトランスフォーメーションでマッピングを設定して、各カラムに入っているさまざまな種類のデータを判定することにします。

次のテーブルは、顧客データベースからのサンプルデータの説明です。

行 ID	フィールド 1	フィールド 2	フィールド 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 LEVINGER	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

このマッピングを実行すると、ラベラトランスフォーメーションが入力データを確率モデルの参照データと比較します。ラベラトランスフォーメーションが、各入力ポートのデータのラベルを選択します。このトランスフォーメーションがラベルを出力ポートに書き込みます。各出力行には一連のラベルがあり、対応する入力行でのデータ構造を定義しています。

次のテーブルは、ラベラトランスフォーメーションが出力ポートに追加するラベルの説明です。

行 ID	出力ラベル
1	電話番号 社名 連絡先
2	電話番号 連絡先 社名
3	電話番号 社名 連絡先
4	電話番号 連絡先 社名
5	電話番号 社名 連絡先
6	organization number contact
7	organization number contact
8	連絡先 社名 電話番号

パーサートランスフォーメーションの例

あるスーパーでは、データベーステーブル内の単一のコラムに製品の説明を格納しています。製品の説明には、さまざまな種類の情報を表している複数のデータ値が入っています。あなたはスーパーでデータ管理を任されています。そこで、製品の説明の中のさまざまな情報の種類ごとにコラムを作成することにします。

パーサートランスフォーメーションでマッピングを設定して、データ値を正しいフィールドに整理します。

次のデータフラグメントにはオレンジジュースの商品説明が含まれています。

Sunnydream オレンジジュース無糖 340 グラム

次のテーブルは、パーサートランスフォーメーションが入力データから作成する出力データの説明です。

製品名	製品タイプ	製品の詳細	製品のサイズ
Sunnydream	オレンジジュース	無糖	340 グラム

確率モデルのオプション

確率モデルを編集する際、データビューまたはラベルビューで作業できます。

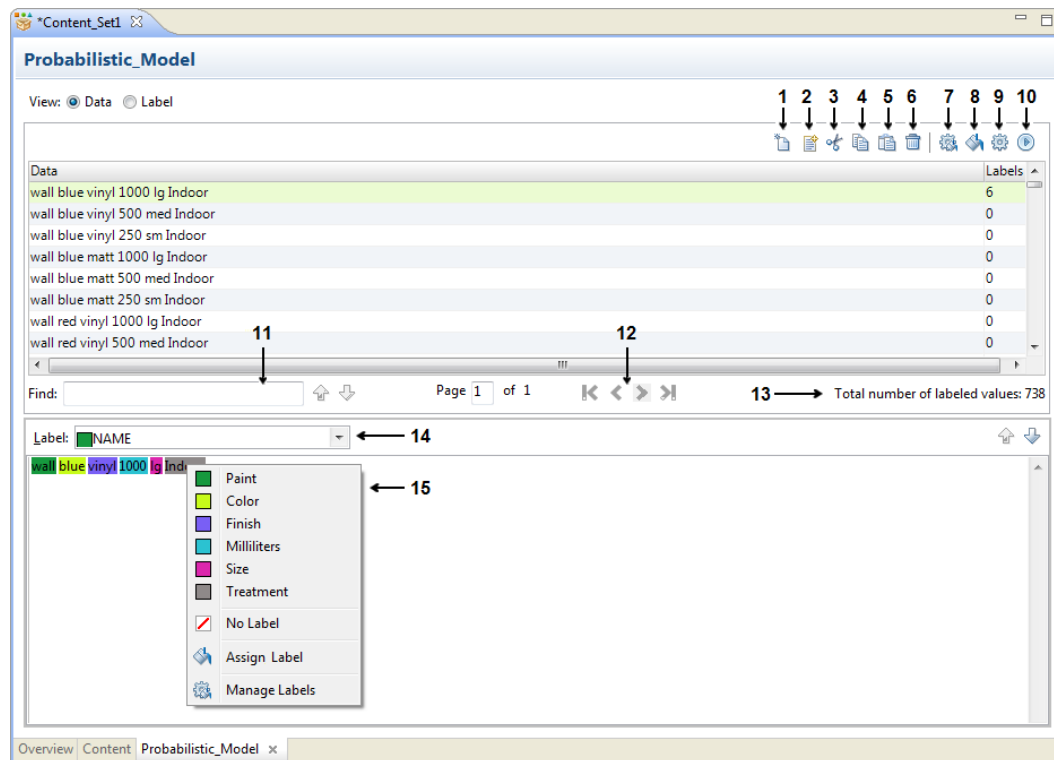
データビューでは、モデルに参照データ行を追加したり、各行のデータ値にラベルを割り当てたりします。ラベルビューでは、モデル内のラベル値の使用に関する詳細な情報を確認できます。データビューでもラベルビューでも確率モデルにラベルを追加できます。

確率モデルのデータビュー

データビューには、確率モデルの参照データ行と、ユーザーが各行に割り当てたラベル値の数が表示されます。また、現在のモデルで値に割り当てたラベルの合計数も示されます。

参照データ行を選択すると、[検索] フィールドの下にあるエディタに行の値が表示されます。行の参照データ値にラベルを割り当てるには、エディタの値を右クリックしてラベル値を選択します。

次の図に、データビューを選択したときに使用可能な確率モデルのオプションを示します。



データビューには次のオプションがあります。

1. 行の追加
空白のデータ行を挿入します。
2. データの追加
モデルリポジトリのデータオブジェクトからデータをインポートします。
3. 切り取り
データ行を確率モデルから削除してクリップボードに追加します。
4. コピー
データ行をクリップボードにコピーします。
5. 貼り付け
データ行をクリップボードから確率モデルに貼り付けます。
6. 削除
確率モデルからデータ行を削除します。
7. ラベルの管理
[ラベルの管理] ダイアログボックスが開きます。このダイアログボックスを使用して、確率モデルのラベル値を追加または削除します。

8. ラベルの割り当て

ユーザーが選択した 1 つ以上の参照データ値にラベルを割り当てます。モデル内の参照データ値のすべての出現箇所にラベルを割り当てるオプションを使用できます。

9. プロパティの編集

確率モデルのプロパティを表示します。

10. コンパイル

確率モデルをコンパイルします。

11. [検索] フィールド

ユーザーが入力した参照データ値を含むモデル内の行を検索します。上矢印または下矢印を使用して、値を含む行に移動します。

12. 進む矢印と戻る矢印

モデル内のデータ値を含む前後の行に移動します。

13. ラベル付けされた値の合計数

ラベルを使用する参照データ値の数を示します。

14. [ラベル] フィールド

ユーザーが選択した参照データ値に適用できるラベル値を表示します。

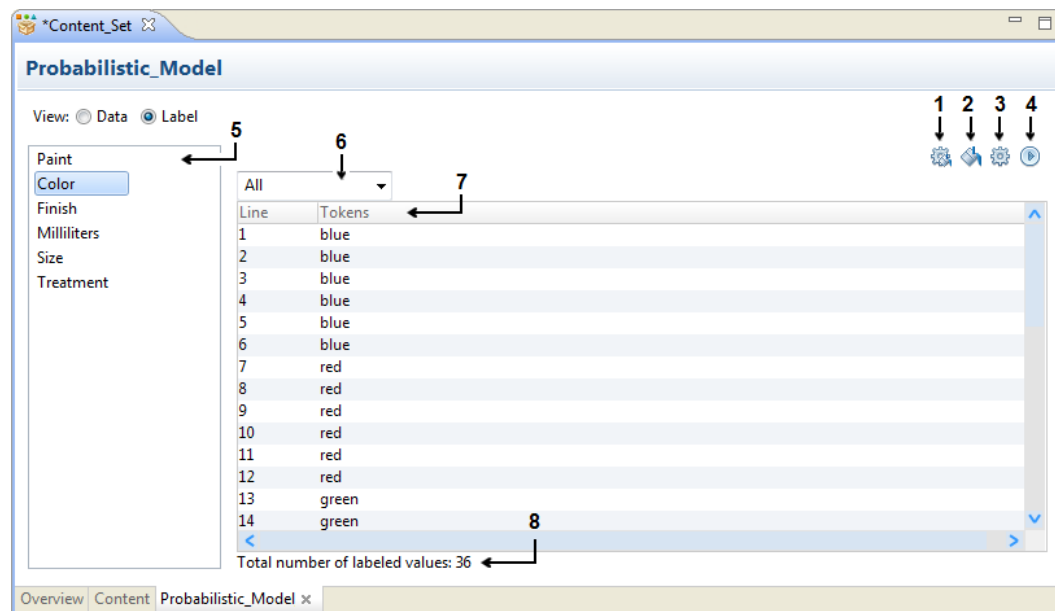
15. [ラベル] メニュー

1 つ以上の参照データ値にラベルを割り当てる場合に使用できるオプションをリスト表示します。メニューを開くには、参照データエディタの参照データ値を右クリックします。

確率モデルのラベルビュー

ラベルビューには、確率モデルで定義したラベルがリスト表示されます。ラベルを選択すると、ユーザーが各行のラベルに割り当てたデータ値がラベルビューに表示されます。

次の図に、データビューを選択したときに使用可能な確率モデルのオプションを示します。



ラベルビューには次のオプションがあります。

1. ラベルの管理

【ラベルの管理】 ダイアログボックスが開きます。このダイアログボックスを使用して、確率モデルのラベル値を追加または削除します。

2. ラベルの割り当て

ユーザーが選択した 1 つ以上の参照データ値にラベルを割り当てます。

単一の操作で、ラベルを 1 つのデータ値に割り当てても、複数の値に割り当ててもできます。

3. プロパティの編集

確率モデルのプロパティを表示します。

4. コンパイル

確率モデルをコンパイルします。

5. ラベル値のリスト

モデル内の参照データ値に割り当て可能なラベルをリスト表示します。

6. 割り当てフィルタ

選択するラベルを使用する参照データ値のリストをフィルタします。フィルタオプションは、データ値へのラベルの割り当て方法に基づいて参照データ値を表示または非表示にします。

フィルタを適用すると、ラベルビューに表示されるラベル付けされた値の合計数に、フィルタ条件を満足する値の数が反映されます。

7. 参照データ値カラム

現在のラベルを使用する参照データ値をリスト表示します。

8. ラベル付けされた値の合計数

現在のラベルを使用する参照データ値の数を示します。

確率モデルの参照データ

確率モデルの参照データ値は、マッピングでトランスフォーメーションに接続する可能性がある入力データのタイプを表します。

Developer ツールで参照データ行を追加、編集、削除することができます。データはクリップボードから貼り付けることも、データソースからインポートすることもできます。参照データ値を追加したら、各行の各データ値にラベルを割り当てます。

確率モデルのラベルデータ

確率モデルのラベル値は、参照データ値に含まれる可能性のある情報タイプを表します。参照データ行をモデルに追加するときに、それぞれの行の各値にラベルを割り当てます。モデルに追加したラベルは、ラベルビュー、およびデータビューのメニューオプションに表示されます。

モデル内のどのラベルでも、任意の参照データの値に割り当てることができます。参照データの複数の行に同じ値があるが、その意味が異なる場合は、それぞれの行の各値に異なるラベルを割り当てられます。

ラベル値の範囲は、ラベラ変換（ラベラ変換）またはパーサ変換（パーサ変換）が確率分析時に読み込む入力ポートの範囲と対応させることができます。確率モデルには、変換（変換）が各入力ポートのデータ値に適用できるラベル値が 1 つ以上必要です。

例えば倉庫で、8 つのカラムを定義するカンマ区切りのファイルにインベントリデータが保存されているとします。このインベントリデータをデータベーステーブルに解析するマッピングを設計します。各データカラムにラベル値がある確率モデルを作成します。マッピングの実行時に、パーサ変換（パーサ変換）が入力データの各値をターゲットテーブルの正しいカラムに書き込みます。

次のテーブルに、確率モデルで作成するインベントリデータのカラムとラベル値のカラムの例を示します。

インベントリカラム名	ラベル名
Product_Name	Product_Name
Quantity	Quantity
場所	場所
Barcode	Barcode
SKU	Stock_Keeping_Unit
Arrival_Date	Arrival_Date
Cost_Price	Cost_Price

注: 入力カラム名を使用することも、他の名前を使用することも可能です。名前が一致している必要はありません。

オーバーフローラベル

変換（変換）が入力データ値にラベルを適用できない場合、変換（変換）はそのデータ値をオーバーフローデータとして処理します。ラベラ変換（ラベラ変換）が特定不能なデータ値にオーバーフローラベルを適用します。パーサ変換（パーサ変換）が特定不能なデータ値をオーバーフローポートに書き込みます。

次のテーブルに、パーサ変換（パーサ変換）が確率モデルで認識されないアドレスデータ要素の解析にオーバーフローポートを使用する場合とその方法を示します。

入力データ	Street_Name ポート	Street_Descriptor ポート	Overflow ポート
Park Place	Park	Place	オーバーフローデータなし
Park Avenue	Park	Avenue	オーバーフローデータなし
Madison Avenue	Madison	Avenue	オーバーフローデータなし
Central Park	Central	Park	オーバーフローデータなし

入力データ	Street_Name ポート	Street_Descriptor ポート	Overflow ポート
Washington Square Park	Washington	Square	Park
Madison Square Garden	Madison	Square	Garden

入力値の数がモデルのラベル数より多い場合も、パーサートランスフォーメーションが値をオーバーフローポートに書き込みます。トランスフォーメーションで確率モデルを使用する前に、入力データを検証し、モデルに正しい数のラベル値があることを確認します。

確率モデルのプロパティ

確率モデルの全般プロパティと詳細プロパティを確認できます。

プロパティエディタを開くには、データビューまたはラベルビューの **【プロパティの編集】** オプションを選択します。

全般プロパティには、確率モデルの名前、モデルの説明、モデルデータファイルの名前が表示されます。詳細プロパティには、Developer ツールが確率モデルのコンパイルに使用する計算プロパティが表示されます。

確率モデルのコンパイルにおける基本要素は *n-gram* です。n-gram とは、それが他の文字の前または後ろに付くことで単語が形成されるような一連の文字です。マッピングが実行されているときに、確率モデルの参照データカラムにある各値に対して複数の n-gram が、ラベラトランスフォーメーションまたはパーサートランスフォーメーションによって作成されます。このトランスフォーメーションが入力データの値を、参照データの値および n-gram と比較します。確率モデルの詳細プロパティにより、確率モデルが n-gram およびその他のモデル機能を処理する方法が決定されます。

注: 詳細プロパティのデフォルト値は、確率的な分析および確率モデルのコンパイルの優先設定を表します。詳細プロパティを編集すると、確率的な分析の精密度に悪影響を及ぼす恐れがあります。行う変更の影響を理解していない限り、詳細プロパティは編集しないでください。

関連項目：

- [「確率的なモデルと分類子モデルのルールとガイドライン」 \(ページ 40\)](#)

確率モデルの設定

確率モデルを設定する手順では、最初に行う分析のタイプを処理します。ラベラトランスフォーメーションでは、確率モデルを使用して、入力文字列内の各値の情報のタイプを識別します。パーサートランスフォーメーションでは、確率モデルを使用して、入力文字列内のデータ値を異なる出力ポートに解析します。

データのラベル付けとデータの解析に同じ確率モデルを使用できます。ラベラトランスフォーメーションでモデルを使用するときは、ユーザーが選択した入力ポートごとにトランスフォーメーションが出力ポートを1つ作成します。パーサートランスフォーメーションでモデルを使用するときは、識別する入力データのタイプごとにトランスフォーメーションが出力ポートを作成します。

確率モデルを作成するには、以下のタスクを実行します。

1. モデルに追加する参照データ値とラベル値を特定します。
分析するデータのフラグメントを使用できます。モデルリポジトリにデータフラグメントを読み取るデータオブジェクトを作成します。
2. コンテンツセットを作成し、そのセットに確率モデルを追加します。
3. 参照データ値をモデルに追加します。
4. ラベル値をモデルに追加します。
モデルリポジトリのデータオブジェクトからデータをインポートできます。参照データの 1 つの行または 1 つのラベルを入力することもできます。
確率モデルを使用してデータを解析するには、モデルに、トランスフォーメーションが作成する必要がある各出力ポートのラベル値があることを確認します。
5. それぞれの行の各データ値を参照するラベルを割り当てます。
単一の操作で、複数の参照データ値にラベルを割り当てることができます。
6. モデルをコンパイルします。
確率モデルをコンパイルしたら、トランスフォーメーションでそのモデルを使用できます。

空の確率モデルの作成

参照データやラベルデータが含まれていない確率モデルオブジェクトを作成できます。空のモデルを作成して、そのモデルにデータを追加またはインポートします。

1. Object Explorer で、コンテンツセットを開くか作成します。
2. [コンテンツ] ビューを選択します。
3. **[確率モデル]** を選択し、**[追加]** をクリックします。
確率モデルウィザードが開きます。
4. **[確率モデル]** オプションを選択します。
[次へ] をクリックします。
5. 確率モデルの名前を入力します。
必要に応じて、このモデルの説明文を入力します。
6. **[完了]** をクリックします。

データオブジェクトからの確率モデルの作成

データオブジェクトを確率モデルのデータのソースとして使用することができます。

確率モデルのパフォーマンスは、ラベラトランスフォーメーションまたはパーサートランスフォーメーションへの入力データをモデルの参照データのソースとして使用した場合に最適になります。

1. Object Explorer で、コンテンツセットを開くか作成します。
2. [コンテンツ] ビューを選択します。
3. **[確率モデル]** を選択し、**[追加]** をクリックします。
確率モデルウィザードが開きます。
4. **[データオブジェクトからの確率モデル]** オプションを選択します。
[次へ] をクリックします。
5. 確率モデルの名前を入力します。

必要に応じて、このモデルの説明文を入力します。

6. モデルリポジトリを参照して、インポートするデータを含むデータオブジェクトを選択します。

ソーシャルメディアのデータオブジェクトは選択しないでください。

【次へ】 をクリックします。

7. データオブジェクトのカラムを確認して、モデルに追加するカラムを 1 つ以上選択します。同じ操作で参照データカラムとラベルカラムを追加できます。

- データのカラムを参照データとしてインポートするには、カラム名を選択して **【データ】** をクリックします。

データカラムを複数選択してもかまいません。Developer tool が、ユーザーが選択したカラムのコンテンツを 1 つのカラムに統合します。

- データのカラムをラベル値としてインポートするには、カラム名を選択して **【ラベル】** をクリックします。

参照データ値とラベル値をインポートするときに、Developer tool が各行のラベルを同じ行の参照データ文字列に割り当てます。カラムを選択する前に、データをプレビューできます。モデルを作成した後にラベルの割り当てを変更できます。

【次へ】 をクリックします。

8. データソースからインポートする行数を選択します。

デフォルトでは、Developer tool がデータソースからすべての行をインポートします。数を入力すると、モデルがデータセットの先頭から行を数えます。

9. インポートするデータ値の区切り文字を指定します。

参照データ値とラベル値に異なる区切り文字を指定できます。デフォルトの区切り文字は空白文字です。

10. **【完了】** をクリックしてモデルを保存します。

確率モデルを作成したら、ラベルの割り当てを確認してモデルをコンパイルします。

確率モデルへのデータソースからのデータの追加

単一の操作で、参照データ値またはラベル値の複数の行を確率モデルにインポートできます。

1. 確率モデルの入っているコンテンツセットを開きます。

2. モデル名を選択し、**【編集】** をクリックします。

3. **【データの付加】** をクリックします。

確率モデルウィザードが開きます。

4. モデルリポジトリを参照して、インポートするデータを含むデータオブジェクトを選択します。

ソーシャルメディアのデータオブジェクトは選択しないでください。

【次へ】 をクリックします。

5. データオブジェクトのカラムを確認して、モデルに追加するカラムを 1 つ以上選択します。同じ操作で参照データカラムとラベルカラムを追加できます。

- データのカラムを参照データとしてインポートするには、カラム名を選択して **【データ】** をクリックします。

データカラムを複数選択してもかまいません。Developer tool が、ユーザーが選択したカラムのコンテンツを 1 つのカラムに統合します。

- データのカラムをラベル値としてインポートするには、カラム名を選択して **【ラベル】** をクリックします。

参照データ値とラベル値をインポートするときに、Developer tool が各行のラベルを同じ行の参照データ文字列に割り当てます。カラムを選択する前に、データをプレビューできます。モデルを作成した後にラベルの割り当てを変更できます。

【次へ】 をクリックします。

6. データソースからインポートする行数を選択します。

デフォルトでは、Developer tool がデータソースからすべての行をインポートします。数を入力すると、モデルがデータセットの先頭から行を数えます。

7. インポートするデータ値の区切り文字を指定します。

参照データ値とラベル値に異なる区切り文字を指定できます。デフォルトの区切り文字は空白文字です。

8. **【完了】** をクリックしてモデルを保存します。

確率モデルへの参照データ行の追加

データビューを使用して、確率モデルに空の行を追加します。

1. モデルの入っているコンテンツセットを開きます。
モデル名を選択し、**【編集】** をクリックします。
2. データビューを選択します。
3. モデルに空の行を追加するには、**【新規】** をクリックします。
4. 追加した行を選択し、行に 1 つ以上の参照データ値を入力します。
5. 確率モデルを保存します。

モデルを保存後、行の各値にラベルを割り当てます。必要に応じて、モデルをコンパイルします。

確率モデルへのラベルの追加

確率モデルにはラベルを 1 つ追加できます。モデルデータの値が表す情報タイプすべてにラベルを追加します。パーサートランスフォーメーションで確率モデルを使用する場合、トランスフォーメーションが作成するべき出力ポートごとに、ラベルを追加します。

1. モデルの入っているコンテンツセットを開きます。
2. モデル名を選択し、**【編集】** をクリックします。
3. データビューまたはラベルビューで、**【ラベルの管理】** をクリックします。
【ラベルの管理】 ダイアログボックスが表示されます。
4. **【ラベルの管理】** ダイアログボックスで、**【新規作成】** をクリックします。
ダイアログボックスの最初の空白行に、ラベルが表示されます。
5. ラベル名を編集します。必要に応じて、ラベルの色を更新します。
6. **【OK】** をクリックして、ラベルをモデルに追加します。
7. 確率モデルを保存します。

ラベルを追加した後、1 つ以上のデータ値にそのラベルを割り当てます。

参照データ値へのラベルの割り当て

参照データ行の 1 つのデータ値にラベルを割り当てることができます。

同じデータ値が行内の複数の場所または複数の行に表示される場合は、そのデータ値に異なるラベルを割り当てられます。

1. モデルの入っているコンテンツセットを開きます。
2. モデル名を選択し、**【編集】** をクリックします。
3. データビューを選択します。
4. ラベルのないデータ値や間違ったラベルの付いたデータ値を検索します。ラベルを使用するデータ値は色分けされています。
5. そのデータ値を含むデータ行を選択します。
エディタに行が表示されます。
6. エディタでデータ値を右クリックし、コンテキストメニューからラベルを選択します。
Developer ツールにより、データ値にラベルが割り当てられます。
7. 確率モデルを保存します。

確率モデルを保存後、必要に応じてモデルをコンパイルします。

複数のデータ値へのラベルの割り当て

単一の操作で、複数の参照データ値にラベルを割り当てることができます。

1. モデルの入っているコンテンツセットを開きます。
2. モデル名を選択し、**【編集】** をクリックします。
3. **【ラベルの割り当て】** をクリックします。
【複数値へのラベルの割り当て】 ダイアログボックスが開きます。
4. **【検索】** フィールドに 1 文字以上を入力します。
【検索】 フィールドにはワイルドカード文字も入力できます。
5. 必要に応じて、追加の検索条件を選択します。
次のオプションを選択またはクリアできます。
 - 大文字/小文字の一致
検索操作で大文字と小文字が区別されます。このオプションとワイルドカード文字を併用しないでください。
 - 文字列全体と一致検索操作で、参照データ値の文字とユーザーが入力した文字の完全な一致が検索されます。このオプションとワイルドカード文字を併用しないでください。
 - ラベル付けされた値の無視
検索操作で、ラベルを使用する参照データ値がスキップされます。
6. 検索条件に一致した参照データ値に割り当てるラベルを選択します。
【ラベルなし】 オプションも選択できます。このオプションを選択すると、ユーザーが入力した文字を含む参照データ値からラベルが削除されます。
7. **【開始】** をクリックします。

ユーザーが定義した検索条件に一致するすべての参照データ値に、Developer tool がラベルを割り当てます。

注: 単一の操作でラベルを割り当てた参照データを表示するには、ラベルビューの **【一括によって割り当て済み】** を使用します。

確率モデルからの行の削除

単一のアクションで、確率モデルから参照データ行を 1 つ以上削除できます。

1. モデルの入っているコンテンツセットを開きます。
2. モデル名を選択し、**【編集】** をクリックします。
3. データビューで、参照データ行を 1 つ以上選択します。
4. **【削除】** をクリックします。

Developer tool が分類子モデルからユーザーが選択した行を削除します。

操作を元に戻すには、キーボードの Ctrl+Z キーを押します。

確率モデルからのラベルの削除

モデルからラベル値を削除する場合、そのラベルを使用した参照データ値はすべてそのモデルに残ります。各参照データ値に別のラベル値を割り当てます。

1. モデルの入っているコンテンツセットを開きます。
2. モデル名を選択し、**【編集】** をクリックします。
3. データビューまたはラベルビューで、**【ラベルの管理】** をクリックします。
4. **【ラベルの管理】** ダイアログボックスで、ラベル値を選択します。
5. **【削除】** をクリックします。
6. **【OK】** をクリックしてラベルを削除します。
7. 確率モデルを保存します。

注: ラベルは確率的なモデル内の構造要素です。モデルをトランスフォーメーションに追加した後にラベルを追加または削除した場合、そのモデルを使用している操作が無効になります。更新したモデルを使用するには、トランスフォーメーション操作を削除して再作成します。

確率モデルのコンパイル

確率モデルのデータまたはラベルの割り当てを更新するときに、モデルをコンパイルできます。モデルをコンパイルして、現在の参照データ値と現在のラベル値間の関連付けに基づいてモデルのロジックを更新します。

確率モデルをコンパイルする前に、各ラベル値が少なくとも 1 つの参照データ値を識別することを確認します。

▶ モデルをコンパイルするには、モデルを Developer ツールで開いて **【コンパイル】** をクリックします。

確率モデル内のデータ行の検索

データビューを使用して、ユーザーが入力した値を含む参照データ行を検索します。

1. 確率モデルの入っているコンテンツセットを開きます。
2. モデル名を選択し、**【編集】** をクリックします。
3. データビューを選択します。
4. **【検索】** フィールドに 1 文字以上を入力します。
入力した値が含まれる、モデル内の最初の行がデータビューに表示されます。
5. 上矢印または下矢印を使用して、値を含む他の行に移動します。

ラベル割り当てによる参照データ値のフィルタリング

ラベルビューを使用して、指定するラベルを使用する参照データ値を検索します。ラベルの割り当てに使用した方法に基づいて結果をフィルタします。

1. 確率モデルの入っているコンテンツセットを開きます。
2. モデル名を選択し、**【編集】** をクリックします。
3. ラベルビューで、ラベル値を選択します。

確率モデルに、そのラベルを使用する参照データ値がリスト表示されます。モデルには、そのラベルを使用するデータ値の数も示されます。

4. ラベルを使用する参照データ値のリストにフィルタを適用します。

次のフィルタのいずれかを選択します。

- すべて。ラベルを使用する参照データ値を表示します。[すべて] はデフォルトのオプションです。
- ユーザーによって割り当て済み。ラベルを割り当てたときに個別に選択した参照データ値すべてを表示します。
- 一括によって割り当て済み。一括割り当て操作の一部としてラベルを割り当てた参照データ値を表示します。

確率モデルには、フィルタ条件を満たす参照データ値が示されます。

未使用のラベル値の検索

ラベルビューを使用して、参照データ値に割り当てられていないラベル値を検索します。各ラベルは最低 1 つ以上の参照データ値に割り当てする必要があります。

1. 確率モデルの入っているコンテンツセットを開きます。
2. モデル名を選択し、**【編集】** をクリックします。
3. ラベルビューで、ラベル値を選択します。

確率モデルに、そのラベルを使用する参照データ値がリスト表示されます。また、そのラベルを使用するデータ値の合計数も示されます。

データ値の合計数がゼロの場合、そのラベルは確率モデルのどの参照データ値にも割り当てられていません。

コピーと貼り付けの操作

確率モデルは、モデルリポジトリの中のあるコンテンツセットから別のコンテンツセットにコピーすることができます。確率モデルをコピーして、リソースを他の Developer ツールユーザーと共有します。

モデルを別のコンテンツセットにコピーすることも、モデルを現在のコンテンツセットにインポートすることもできます。リポジトリの中の複数のコンテンツセットから複数のモデルを 1 回の操作でインポートすることができます。

モデルをコピーする場合、コンテンツ管理サービスは、モデルデータファイルのコピーを Informatica サービスホストマシンに作成します。各モデルが異なるデータファイルを使用します。

別のコンテンツセットへの確率モデルのコピー

確率モデルは、モデルリポジトリの中のあるコンテンツセットから別のコンテンツセットにコピーすることができます。確率モデルをコピーする際は、モデルオブジェクトおよびコピー元とコピー先のコンテンツセットを指定します。

1. 確率モデルの入っているコンテンツセットを開きます。
2. 確率モデルを選択し、**【コピー先】** をクリックします。
3. モデルリポジトリの中を参照して、コンテンツセットを選択します。

この確率モデルは、現在のプロジェクトあるいは別のプロジェクトの中のコンテンツセットにコピーすることができます。

4. **【OK】** をクリックします。

Developer ツールが、選択されたコンテンツセットに確率モデルをコピーします。

別のコンテンツセットからの確率モデルのインポート

確率モデルを、モデルリポジトリの中のあるコンテンツセットから別のコンテンツセットにインポートすることができます。確率モデルをインポートする際は、1 つまたは複数のモデルオブジェクトおよびコピー元とコピー先のコンテンツセットを指定します。

1. 確率モデルを入れるコンテンツセットを開きます。
2. 確率モデルを選択し、**【貼り付け元】** をクリックします。
3. モデルリポジトリの中を参照して、確率モデルを選択します。

現在のプロジェクトあるいは別のプロジェクトのコンテンツセットから確率モデルを貼り付けることができます。

4. **【OK】** をクリックします。

Developer ツールが、現在のコンテンツセットに確率モデルを貼り付けます。

参照データ行のクリップボードへのコピー

参照データの 1 つ以上の行を確率モデルからクリップボードにコピーできます。行を別の確率モデルに貼り付けることも可能です。

1. 確率モデルの入っているコンテンツセットを開きます。
2. モデル名を選択し、**【編集】** をクリックします。
3. データビューで、参照データの行を 1 つ以上選択します。
4. Ctrl+C キーを使用して、行をクリップボードにコピーします。

この操作で、参照データと参照データに割り当てたラベル値がコピーされます。

Ctrl+V キーを使用して、その行をテキストエディタまたは別の確率モデルのデータビューに貼り付けます。

索引

A

Analyst ツール

参照データ値の検索と置換 [25](#)

か

確率的なモデル

ルールおよびガイドライン [40](#)

確率モデル

コンテンツセットの [37](#)

カラムの管理

参照テーブル [23](#)

カラムパターンからの参照テーブルの作成

参照テーブル [19](#)

監査テーブルイベントの表示

参照テーブル [27](#)

管理されていない参照テーブル

定義 [12](#)

モデルリポジトリとの同期 [13](#)

管理されている参照テーブル [12](#)

き

行の管理

参照テーブル [24](#)

こ

コンテンツ管理サービス

参照テーブル特権 [13](#)

コンテンツセット

確率モデル [37](#)

正規表現 [37](#)

トークンセット [38](#)

バージョン管理 [14](#), [22](#), [30](#)

パターンセット [37](#)

分類子モデル [36](#)

文字セット [36](#)

さ

参照テーブル

Analyst ツールでの値の検索と置換 [25](#)

Analyst ツールでのリフレッシュ [26](#)

Analyst ツールの概要 [15](#)

Analyst ツールのプロパティ [15](#)

Developer tool の概要 [31](#)

Developer ツールのプロパティ [31](#)

カラムの管理 [23](#)

カラムパターンからの参照テーブルの作成 [19](#)

監査証拠テーブルの表示 [27](#)

参照テーブル (続く)

管理されていない参照テーブル [12](#)

管理されている/管理されていない [12](#)

管理されている参照テーブル [12](#)

行の管理 [24](#)

コンテンツ管理サービス [12](#)

参照データウェアハウス [12](#)

参照テーブルのインポート [21](#)

参照テーブルのエクスポート [25](#)

参照テーブルの手動作成 [17](#)

特権 [13](#)

バージョン管理 [14](#), [22](#), [30](#)

パターンベースの解析 [12](#)

プロファイルカラムからの参照テーブルの作成 [18](#)

参照テーブルのインポート

参照テーブル [21](#)

参照テーブルのエクスポート

参照テーブル [25](#)

参照テーブルの手動作成

参照テーブル [17](#)

せ

正規表現 [37](#)

と

トークンセット [38](#)

特権

コンテンツ管理サービス [13](#)

は

バージョン管理

Analyst ツールの参照テーブル [22](#)

Developer tool の参照テーブル [30](#)

コンテンツセット [14](#), [30](#)

参照テーブル [14](#)

パターンセット [37](#)

ふ

プロファイルカラムデータからの参照テーブルの作成

参照テーブル [18](#)

分類子モデル

コンテンツセットの [36](#)

ルールおよびガイドライン [40](#)

も

文字セット [36](#)