



Informatica®

10.1.1 HotFix 2

Guia de Dados de Referência

Este software e a documentação são fornecidos somente sob um contrato de licença separado, contendo restrições sobre uso e divulgação. Não está permitida de forma alguma a reprodução ou a transmissão de qualquer parte deste documento (seja por meio eletrônico, fotocópia, gravação ou quaisquer outros meios) sem o consentimento prévio da Informatica LLC.

DIREITOS DO GOVERNO DOS ESTADOS UNIDOS Programas, softwares, bancos de dados, bem como a documentação e os dados técnicos relacionados, distribuídos a clientes do Governo dos EUA são "softwares de computador comerciais" ou "dados técnicos comerciais", de acordo com o Regulamento de Aquisição Federal aplicável e os regulamentos suplementares específicos da agência. Como tal, a utilização, duplicação, divulgação, modificação e adaptação estão sujeitas às restrições e aos termos de licença estabelecidos no contrato governamental aplicável e, na medida do que for aplicável pelos termos do contrato governamental, aos direitos adicionais estabelecidos no FAR 52.227-19, Licença de Software de Computador Comercial.

Informatica e o logotipo Informatica são marcas comerciais ou marcas registradas da Informatica LLC nos Estados Unidos e em muitas jurisdições por todo o mundo. Uma lista atual das marcas comerciais da Informatica está disponível na Internet em <https://www.informatica.com/trademarks.html>. Os nomes de outras companhias e produtos podem ser nomes ou marcas comerciais de seus respectivos proprietários.

Partes deste software e/ou documentação estão sujeitas a copyright detido por terceiros. Os avisos de terceiros necessários são incluídos no produto.

Consulte as patentes em <https://www.informatica.com/legal/patents.html>.

ISENÇÃO DE RESPONSABILIDADE: a Informatica LLC fornece esta documentação no estado em que se encontra, sem garantia de qualquer tipo, expressa ou implícita, incluindo, mas não limitando-se, as garantias implícitas de não infração, comercialização ou uso para um determinado propósito. A Informatica LLC não garante que este software ou documentação não contenha erros. As informações fornecidas neste software ou documentação podem incluir imprecisões técnicas ou erros tipográficos. As informações deste software e documentação estão sujeitas a alterações a qualquer momento sem aviso prévio.

AVISOS

Este produto da Informatica (o "Software") traz determinados drivers (os "drivers da DataDirect") da DataDirect Technologies, uma empresa em funcionamento da Progress Software Corporation ("DataDirect"), que estão sujeitos aos seguintes termos e condições:

1. OS DRIVERS DA DATADIRECT SÃO FORNECIDOS NO ESTADO EM QUE SE ENCONTRAM, SEM GARANTIA DE QUALQUER TIPO, EXPRESSA OU IMPLÍCITA, INCLUINDO, MAS NÃO LIMITANDO-SE, AS GARANTIAS IMPLÍCITAS DE COMERCIALIZAÇÃO, ADEQUAÇÃO A UMA FINALIDADE ESPECÍFICA E NÃO INFRAÇÃO.
2. EM NENHUM CASO, A DATADIRECT OU SEUS FORNECEDORES TERCEIRIZADOS SERÃO RESPONSÁVEIS, EM RELAÇÃO AO CLIENTE FINAL, POR QUAISQUER DANOS DIRETOS, INDIRETOS, INCIDENTAIS, ESPECIAIS, CONSEQUENCIAIS OU DE MAIS QUE POSSAM ADVIR DO USO DE DRIVERS ODBC, SENDO OU NÃO ANTERIORMENTE INFORMADOS DAS POSSIBILIDADES DE TAIS DANOS. ESTAS LIMITAÇÕES SE APLICAM A TODAS AS CAUSAS DE AÇÃO, INCLUINDO, SEM LIMITAÇÕES, QUEBRA DE CONTRATO, QUEBRA DE GARANTIA, NEGLIGÊNCIA, RESPONSABILIDADE RIGOROSA, DETURPAÇÃO E OUTROS ATOS ILÍCITOS.

As informações contidas neste documento estão sujeitas a alteração sem aviso prévio. Se você encontrar quaisquer problemas nesta documentação, informe-os em infa_documentation@informatica.com.

Os produtos Informatica apresentam garantias segundo os termos e condições dos acordos em que são fornecidos. A INFORMATICA FORNECE AS INFORMAÇÕES NESTE DOCUMENTO "COMO ESTÃO" SEM GARANTIA DE QUALQUER TIPO, EXPRESSA OU IMPLÍCITA, INCLUINDO, SEM QUAISQUER GARANTIAS DE COMERCIALIZAÇÃO, ADEQUAÇÃO A UM DETERMINADO FIM E QUALQUER GARANTIA OU CONDIÇÃO DE NÃO-VIOLAÇÃO.

Data da Publicação: 2018-06-08

Conteúdo

Prefácio.....	7
Recursos da Informatica.	7
Rede da Informatica.	7
Base de Dados de Conhecimento da Informatica.	7
Documentação da Informatica.	7
Matrizes de Disponibilidade de Produto Informatica.	8
Informatica Velocity.	8
Informatica Marketplace.	8
Suporte global a clientes Informatica.	8
 Capítulo 1: Introdução aos Dados de Referência.....	 9
Visão Geral dos Dados de Referência.	9
Dados de Referência da Informatica.	10
Dados de Referência Definidos pelo Usuário.	10
Tabelas de Referência.	11
Estrutura de Tabelas de Referência.	11
Privilégios do Data Warehouse de Referência.	12
Parâmetros e Tabelas de Referência.	12
Objetos de Dados de Referência e Controle de Versão.	13
 Capítulo 2: Tabelas de Referência na Ferramenta Analyst.....	 14
Visão Geral das Tabelas de Referência da Ferramenta Analyst.	14
Propriedades da Tabela de Referência.	15
Propriedades Gerais da Tabela de Referência.	15
Propriedades de Coluna da Tabela de Referência.	16
Criar uma Tabela de Referência no Editor da Tabela de Referência.	16
Criar uma Tabela de Referência a partir de Dados do Perfil.	17
Criando uma Tabela de Referência a partir de Dados de Colunas de Perfil.	17
Criando uma Tabela de Referência a partir de Padrões de Valor.	18
Criar uma Tabela de Referência a partir um Arquivo Simples.	20
Propriedades de Arquivo Simples da Ferramenta Analyst.	20
Criação de uma tabela de referência de um arquivo simples.	20
Criar uma Tabela de Referência a partir de uma Tabela de Banco de Dados.	21
Criar uma Tabela de Referência a partir de uma Tabela de Banco de Dados.	21
Trabalhando com Tabelas de Referência em um Repositório do Modelo com Controle de Versão.	22
Atualizações da Tabela de Referência.	23
Gerenciar Colunas.	23
Gerenciar Linhas.	24
Localizar e Substituir Valores.	25
Exportando Dados da Tabela de Referência.	25

Ativar e Desativar Edições em uma Tabela de Referência não Gerenciada.	26
Atualizar os Valores da Tabela de Referência.	26
Eventos de Trilha de Auditoria.	27
Exibir Eventos de Trilha de Auditoria.	27
Regras e Diretrizes para Tabela de Referência.	28

Capítulo 3: Dados de Referência na Ferramenta Developer..... 29

Visão Geral dos Dados de Referência da Ferramenta Developer.	29
Transformações e Dados de Referência.	30
Trabalhando com Objetos de Dados de Referência em um Repositório do Modelo com Controle de Versão.	30
Fazendo Check-Out de Objeto de Dados de Referência.	30
Fazendo Check-In de Objeto de Dados de Referência.	31
Tabelas de Referência.	31
Propriedades de Dados da Tabela de Referência.	32
Criando um Objeto de Tabela de Referência.	32
Criar uma Tabela de Referência a partir de um Arquivo Simples.	33
Criar uma Tabela de Referência de uma Origem Relacional.	34
Conjuntos de Conteúdo.	36
Conjuntos de Caracteres.	36
Modelos do Classificador.	37
Conjuntos de Padrões.	37
Modelos Probabilísticos.	37
Expressões Regulares.	38
Conjuntos de Tokens.	39
Regras e Diretrizes dos Modelos Probabilísticos e do Classificador.	40
Criar um Conjunto de Conteúdo.	41
Criando um Objeto de Dados de Referência em um Conjunto de Conteúdo.	42

Capítulo 4: Modelos do Classificador..... 43

Visão Geral dos Modelos do Classificador.	43
Estrutura do Modelo do Classificador.	44
Pontuações do Classificador.	44
Exemplo de Transformação do Classificador.	44
Opções do Modelo do Classificador.	45
Dados de Referência do Modelo do Classificador.	47
Dados do Rótulo do Modelo do Classificador.	47
Gerenciamento do Rótulo do Modelo do Classificador.	48
Configuração do Modelo do Classificador.	49
Criando um Modelo do Classificador.	49
Anexando os Dados de uma Fonte de Dados a um Modelo do Classificador	50
Adicionando uma Linha de Dados de Referência a um Modelo do Classificador.	51
Adicionar um Rótulo a um Modelo do Classificador.	51

Atribuindo um Rótulo às Linhas de Dados de Referência.	51
Identificando Valores de Rótulo Não Utilizados.	52
Excluindo Linhas de um Modelo do Classificador.	52
Excluir um Rótulo de um Modelo do Classificador.	52
Compilar um Modelo do Classificador.	53
Operações Filtrar e Localizar.	53
Usando um Valor de Dados para Filtrar as Linhas de Dados de Referência.	53
Usando um Valor de Rótulo para Filtrar as Linhas de Dados de Referência.	53
Localizando um Valor em uma Linha de Dados de Referência.	54
Operações de Copiar e Colar.	54
Copiar um Modelo do Classificador para Outro Conjunto de Conteúdo.	54
Importar um Modelo do Classificador de Outro Conjunto de Conteúdo.	54
Capítulo 5: Modelos Probabilísticos.	56
Visão Geral de Modelos Probabilísticos.	56
Estrutura de Modelo Probabilístico.	57
Exemplo de Transformação do Rotulador.	58
Exemplo de Transformação do Analisador.	59
Opções do Modelo Probabilístico.	59
Exibição de Dados do Modelo Probabilístico.	59
Exibição de Rótulo do Modelo Probabilístico.	61
Dados de Referência do Modelo Probabilístico.	62
Dados de Rótulo do Modelo Probabilístico.	62
Estouro de Rótulo.	63
Propriedades do Modelo Probabilístico.	64
Configuração de Modelo Probabilístico.	64
Criar um Modelo Probabilístico Vazio.	65
Criar um Modelo Probabilístico a partir de Objeto de Dados.	65
Anexando os Dados de uma Fonte de Dados a um Modelo Probabilístico.	66
Adicionar uma Linha de Dados de Referência a um Modelo Probabilístico.	67
Adicionar um Rótulo a um Modelo Probabilístico.	67
Atribuindo um Rótulo a um Valor de Dados de Referência.	68
Atribuindo um Rótulo a Vários Valores de Dados	68
Excluindo Linhas de um Modelo Probabilístico.	69
Excluir um Rótulo de um Modelo Probabilístico.	69
Compilar o Modelo Probabilístico.	69
Localizando Linhas de Dados em um Modelo Probabilístico.	70
Filtrando Valores de Dados de Referência por Atribuição de Rótulo.	70
Localizando Valores de Rótulo Não Utilizados.	70
Operações de Copiar e Colar.	71
Copiar um Modelo Probabilístico para Outro Conjunto de Conteúdo.	71
Importar um Modelo Probabilístico de Outro Conjunto de Conteúdo.	71
Copiando Linhas de Dados de Referência para a Área de Transferência.	71

Índice.....	73
-------------	----

Prefácio

O *Guia de Dados de Referência da Informatica* inclui informações sobre os objetos e os arquivos de dados de referência que você pode usar no Informatica Developer e no Informatica Analyst. Ele foi desenvolvido para analistas de dados, administradores de dados e outras pessoas que usam dados de referência para verificar e aprimorar a precisão e a capacidade de uso dos dados da organização.

Recursos da Informatica

Rede da Informatica

A Rede da Informatica hospeda o Suporte Global a Clientes da Informatica, a Base de Dados de Conhecimento da Informatica e outros recursos de produtos. Para acessar a Rede da Informatica, visite <https://network.informatica.com>.

Como membro, você pode:

- Acessar todos os seus recursos Informatica em um só lugar.
- Pesquisar a Base de Dados de Conhecimento em busca de recursos de produtos, incluindo documentações, perguntas frequentes e práticas recomendadas.
- Visualizar informações sobre disponibilidade de produtos.
- Revisar seus casos de suporte.
- Encontrar a sua Rede de Grupo de Usuários da Informatica local e colaborar com seus colegas.

Base de Dados de Conhecimento da Informatica

Use a Base de Dados de Conhecimento da Informatica para pesquisar a Rede da Informatica em busca de recursos de produtos, como documentações, artigos de instruções, práticas recomendadas e PAMs.

Para acessar a Base de Dados de Conhecimento, visite <https://kb.informatica.com>. Em caso de dúvidas, comentários ou ideias sobre a Base de Dados de Conhecimento, entre em contato com a equipe da Base de Dados de Conhecimento da Informatica em KB_Feedback@informatica.com.

Documentação da Informatica

Para obter a documentação mais recente do seu produto, navegue pela Base de Dados de Conhecimento da Informatica em https://kb.informatica.com/_layouts/ProductDocumentation/Page/ProductDocumentSearch.aspx

Em caso de dúvidas, comentários ou ideias sobre esta documentação, entre em contato com a equipe de Documentação da Informatica pelo e-mail infa_documentation@informatica.com.

Matrizes de Disponibilidade de Produto Informatica

As Matrizes de Disponibilidade de Produto (PAMs) indicam as versões dos sistemas operacionais, os bancos de dados e outros tipos de fontes e destinos de dados com os quais uma versão de produto é compatível. Se você for membro da Rede da Informatica, poderá acessar PAMs em <https://network.informatica.com/community/informatica-network/product-availability-matrices>.

Informatica Velocity

O Informatica Velocity é uma coleção de dicas e práticas recomendadas desenvolvidas pelos Serviços Profissionais da Informatica. Desenvolvido com base na experiência no mundo real de centenas de projetos de gerenciamento de dados, o Informatica Velocity representa o conhecimento coletivo de nossos consultores, que trabalharam com organizações de todo o mundo para planejar, desenvolver, implantar e manter soluções de gerenciamento de dados bem-sucedidas.

Se você for membro da Rede da Informatica, poderá acessar os recursos do Informatica Velocity em <http://velocity.informatica.com>.

Se você tiver dúvidas, comentários ou ideias sobre o Informatica Velocity, entre em contato com os Serviços Profissionais da Informatica em ips@informatica.com.

Informatica Marketplace

O Informatica Marketplace é um fórum onde você pode encontrar soluções que aumentam, ampliam ou aprimoram suas implementações da Informatica. Aproveitando qualquer uma das centenas de soluções fornecidas por desenvolvedores e parceiros da Informatica, você pode melhorar sua produtividade e agilizar o tempo de implementação nos seus projetos. Você pode acessar o Informatica Marketplace através do link <https://marketplace.informatica.com>.

Suporte global a clientes Informatica

Você pode entrar em contato com um Centro de Suporte Global por telefone ou via Suporte Online na Rede da Informatica.

Para descobrir o número de telefone local do Suporte Global a Clientes da Informatica, visite o site da Informatica no seguinte link:

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

Se você for membro da Rede da Informatica, poderá usar o Suporte Online em <http://network.informatica.com>.

CAPÍTULO 1

Introdução aos Dados de Referência

Este capítulo inclui os seguintes tópicos:

- [Visão Geral dos Dados de Referência, 9](#)
- [Dados de Referência da Informatica, 10](#)
- [Dados de Referência Definidos pelo Usuário, 10](#)
- [Tabelas de Referência, 11](#)
- [Objetos de Dados de Referência e Controle de Versão, 13](#)

Visão Geral dos Dados de Referência

As transformações da Informatica podem usar dados de referência para analisar e atualizar dados. É possível criar objetos de dados de referência nas ferramentas Developer e Analyst. Também é possível importar objetos e arquivos de dados de referência para o repositório do Modelo e para o sistema de arquivos. Você pode usar o instalador de Conteúdo do Data Quality para importar objetos de dados de referência e para instalar os arquivos de dados de referência.

É possível criar ou editar os seguintes tipos de dados de referência:

Tabelas de referência

Uma tabela de referência contém as versões padrão e alternativa de um conjunto de valores de dados. Adicione uma tabela de referência a uma transformação na Developer tool para verificar se os valores de dados de origem estão precisos e formatados corretamente.

A maioria das tabelas de referência contém pelo menos duas colunas. Uma coluna contém a versão padrão ou preferencial de um valor e as outras colunas contêm versões alternativas. Quando você adicionar uma tabela de referência a uma transformação, a transformação pesquisa os dados da porta de entrada para os valores que também são exibidos na tabela. Você pode criar tabelas com quaisquer dados que sejam úteis para o projeto de dados no qual você está trabalhando.

Conjuntos de conteúdo

Um conjunto de conteúdo é um objeto de repositório do Modelo que especifica valores de dados de referência no repositório ou em um arquivo. Quando você adiciona um conjunto de conteúdo a uma transformação, essa transformação pesquisa os dados de entrada em busca de valores que correspondem aos padrões de dados no conjunto de conteúdo.

O instalador de Conteúdo do Data Quality pode instalar os seguintes tipos de dados de referência:

Tabelas de referência da Informatica

Objetos de repositório e arquivos de dados desenvolvidos pela Informatica. Você importa tabelas de referência da Informatica ao importar objetos aceleradores para o repositório do Modelo. Os tipos de informações de referência incluem códigos de área de telefone, formatos de código postal, nomes, ocupações e acrônimos. É possível editar tabelas de referência da Informatica.

Conjuntos de conteúdo da Informatica

Objetos de repositório e arquivos de dados desenvolvidos pela Informatica. Você importa conjuntos de conteúdo ao importar objetos aceleradores para o repositório do Modelo. Um conjunto de conteúdo inclui diferentes tipos de dados de referência que você pode usar para realizar operações de pesquisa com transformações de qualidade de dados.

Arquivos de dados de referência de endereço

Arquivos de dados de referência que contêm dados para os endereços que podem receber entregas em um país. A transformação de Validador de Endereço lê os dados de referência. Não é possível criar ou editar arquivos de dados de referência de endereço.

Os dados de referência de endereço são atuais para um determinado período e você deve atualizar os dados regularmente, por exemplo, a cada trimestre.

Arquivos de preenchimento de identidade

Arquivos de dados de referência que contêm informações sobre identidades pessoais, familiares e corporativas. As transformações de Correspondência e Comparação usam arquivos de preenchimento para encontrar possíveis identidades em dados de entrada. Você não pode criar nem editar arquivos de preenchimento de identidade.

Dados de Referência da Informatica

Você pode comprar e baixar dados de referência de endereço e dados de preenchimento de identidade da Informatica.

É possível comprar uma assinatura anual para os dados de endereço em um país e fazer download dos dados de endereço mais recentes da Informatica a qualquer momento durante o período da assinatura.

O usuário do Instalador de Conteúdo faz download e instala os dados de referência separadamente dos aplicativos. Entre em contato com o administrador do usuário para obter informações sobre os dados de referência instalados no sistema

Dados de Referência Definidos pelo Usuário

Os valores em um objeto de dados podem ser usados para criar um objeto de dados de referência.

Por exemplo, você pode selecionar um objeto de dados ou coluna de perfil que contenha valores que são específicos para um projeto ou organização. Crie objetos de dados de referência personalizados com base nos valores de coluna.

É possível criar um objeto de dados de referência de uma coluna de dados para verificar o seguinte:

- As linhas de dados na coluna contêm o mesmo tipo de informações.

- Se um valor de origem é válido. O objeto de referência pode incluir uma lista de valores válidos ou inválidos.

A tabela a seguir lista os exemplos comuns de projeto de colunas de dados que podem conter dados de referência:

Informações	Exemplo de Dados de Referência
Códigos de Unidade de Manutenção de Estoque (SKU)	Use uma coluna SKU para criar uma tabela de referência de código SKU válido para uma organização. Use a tabela de referência para localizar os códigos SKU corretos ou incorretos em um conjunto de dados.
Códigos de funcionário	Use uma coluna de código ou ID de funcionário para criar uma tabela de referência dos códigos de funcionários válidos. Use a tabela de referência para localizar erros nos dados de funcionários.
Números de conta de cliente	Execute um perfil em uma coluna de conta de cliente para identificar os padrões de número de conta. Use o perfil para criar um conjunto de token de padrões de dados incorretos. Use o conjunto de tokens para localizar números de conta que não estão em conformidade com a estrutura correta de número da conta.
Nomes de cliente	Quando uma coluna de nome de cliente contém o primeiro nome, o nome do meio e sobrenome, você pode criar um modelo probabilístico que define a estrutura esperada das strings na coluna. Use o modelo probabilístico para localizar strings de dados que não pertençam à coluna.

Tabelas de Referência

Crie e atualize as tabelas de referência na ferramenta Analyst e na Developer tool.

As tabelas de referência armazenam metadados no repositório do Modelo. As tabelas de referência podem armazenar dados da coluna no data warehouse de referência ou em outro banco de dados. Quando o data warehouse de referência armazena os dados da coluna, os Serviços Informatica identificam a tabela como uma tabela de referência gerenciada. Quando outro banco de dados armazena os dados da coluna, os Serviços Informatica identificam a tabela como uma tabela de referência não gerenciada.

O Serviço do Gerenciamento de Conteúdo armazena a conexão de banco de dados do data warehouse de referência. Você pode especificar um banco de dados do IBM DB2, do Microsoft SQL Server ou do Oracle como um data warehouse de referência.

Quando você importar dados para o data warehouse de referência de outro banco de dados, use uma conexão nativa ou uma conexão ODBC para importá-los. Quando você especificar um banco de dados não gerenciado como a fonte de dados de uma tabela de referência, use uma conexão nativa para se conectar ao banco de dados.

Estrutura de Tabelas de Referência

A maioria das tabelas de referência contém pelo menos duas colunas. Uma coluna contém a versão correta ou obrigatória dos valores de dados. As outras colunas contêm diferentes versões dos valores, incluindo versões alternativas que podem ser exibidas nos dados de origem.

A coluna que contém os valores corretos ou obrigatórios é chamada de coluna válida. Quando uma transformação lê uma tabela de referência em um mapeamento, a transformação procura valores nas colunas inválidas. Quando a transformação encontra um valor inválido, ela retorna o valor correspondente da

coluna válida. Você também pode configurar uma transformação para retornar um único valor comum, em vez dos valores válidos.

A coluna válida pode conter dados que estão formalmente corretos, como CEPs. Ela pode conter dados relevantes para um projeto, como números de Unidade de Manutenção de Estoque (SKU) que são únicos para uma organização. Também é possível criar uma coluna válida de dados inválidos, como valores que contêm erros de dados conhecidos que você deseja pesquisar.

Por exemplo, você cria uma tabela de referência incluindo uma lista de números de SKU válidos em uma organização de varejo. Você adiciona a tabela de referência à transformação de Rotulador e cria um mapeamento com a transformação. Você executa o mapeamento com uma tabela de banco de dados de produtos. Quando o mapeamento é executado, o Rotulador cria uma coluna que identifica os registros do produto não contêm números de SKU válidos.

Tabelas de Referência e a Transformação do Analisador

Crie uma tabela de referência com uma única coluna para usar a tabela de dados em uma operação de análise baseada em padrões. Configure a transformação do Analisador para executar a análise baseada em padrões e importe os dados de referência para a configuração da transformação.

Privilégios do Data Warehouse de Referência

O Serviço do Gerenciamento de Conteúdo usa os privilégios para restringir as ações do usuário em tabelas de referência. Use as opções de Segurança na ferramenta Administrator para revisar ou atualizar os privilégios do serviço.

Para trabalhar com as tabelas de referência, você deve ter os seguintes privilégios no Serviço do Gerenciamento de Conteúdo:

- Criar Tabelas de Referência
- Editar Dados de Tabela de Referência
- Editar Metadados de Tabela de Referência

Para editar dados em uma tabela de referência não gerenciada, verifique também se você configurou o objeto de tabela de referência de forma a permitir edições.

Nota: Se você editar os metadados para uma tabela de referência não gerenciada em um aplicativo de banco de dados, use a ferramenta Analyst para sincronizar o repositório do Modelo com a tabela. Você deve sincronizar o repositório do Modelo e a tabela antes de usar a tabela de referência não gerenciada na Developer tool.

Parâmetros e Tabelas de Referência

Você pode usar parâmetros para identificar as tabelas de referência no repositório do Modelo. Você pode criar um parâmetro na Developer tool que identifica a tabela de referência. Ou você pode adicionar a localização da tabela de referência a um arquivo de parâmetro.

Quando você criar um parâmetro na Developer tool, adicione-o a uma transformação em um mapeamento. Quando você adicionar a localização da tabela de referência a um arquivo de parâmetro, especifique o arquivo ao executar um mapeamento no prompt de comando. Nesse caso, o Serviço de Integração de Dados lerá a tabela de referência que o parâmetro identifica quando você executar o mapeamento.

Você pode adicionar um parâmetro que identifica uma tabela de referência das seguintes transformações:

- Conversor de Maiúsculas/Minúsculas transformação
- transformação de Rotulador
- transformação de Analisador no modo de análise de token

- transformação de Padronizador

Nota: Use o comando *infacmd ms runMapping* para executar um mapeamento no prompt de comando.

Objetos de Dados de Referência e Controle de Versão

Se o repositório do Modelo que armazena os objetos de dados de referência integrar-se a um aplicativo de controle de versão, você poderá aplicar controle de versão aos objetos. É possível aplicar controle de versão a tabelas de referência e conjuntos de conteúdo.

Você pode fazer o check-in e o check-out de objetos de dados de referência em um repositório do Modelo com suporte para controle de versão. É possível desfazer um check-out, recuperar uma versão anterior de um objeto e restaurar um objeto para uma versão anterior. Quando os objetos de dados de referência não estão sob controle de versão, o repositório do Modelo bloqueia um objeto de dados de referência que você edita. Outros usuários não podem editar um objeto bloqueado no qual você esteja trabalhando. Quando você fechar o objeto, o repositório do Modelo liberará o bloqueio, e outros usuários poderão editar esse objeto.

Nota: O controle de versão aplica-se aos metadados que o repositório do Modelo armazena para um objeto de tabela de referência não gerenciada. O controle de versão não se aplica aos dados em uma tabela de referência não gerenciada. Não é possível exibir ou restaurar os dados de referência de uma versão anterior de uma tabela de referência não gerenciada.

CAPÍTULO 2

Tabelas de Referência na Ferramenta Analyst

Este capítulo inclui os seguintes tópicos:

- [Visão Geral das Tabelas de Referência da Ferramenta Analyst, 14](#)
- [Propriedades da Tabela de Referência, 15](#)
- [Criar uma Tabela de Referência no Editor da Tabela de Referência, 16](#)
- [Criar uma Tabela de Referência a partir de Dados do Perfil, 17](#)
- [Criar uma Tabela de Referência a partir um Arquivo Simples, 20](#)
- [Criar uma Tabela de Referência a partir de uma Tabela de Banco de Dados, 21](#)
- [Trabalhando com Tabelas de Referência em um Repositório do Modelo com Controle de Versão, 22](#)
- [Atualizações da Tabela de Referência, 23](#)
- [Eventos de Trilha de Auditoria, 27](#)
- [Regras e Diretrizes para Tabela de Referência, 28](#)

Visão Geral das Tabelas de Referência da Ferramenta Analyst

Crie tabelas de referência no espaço de trabalho Design da ferramenta Analyst.

Você pode criar uma tabela de referência com base em um arquivo simples, em uma fonte de dados no repositório do Modelo e em uma tabela em outro banco de dados.

Você pode criar uma tabela de referência com base em uma coluna de perfil ou em um subconjunto dos dados de uma coluna de perfil. É possível também criar uma tabela de referência de padrões de coluna escolhidos de um perfil.

Ao criar ou atualizar uma tabela de referência, você configura as propriedades da tabela e as colunas de dados que ela contém.

Propriedades da Tabela de Referência

As propriedades da tabela de referência podem ser exibidas e atualizadas na ferramenta Analyst. Uma tabela de referência exibe as propriedades gerais e as propriedades de coluna. As propriedades gerais incluem o nome da tabela de referência, a data de criação, o nome da conexão de banco de dados e o nome da coluna válida. As propriedades da coluna incluem os nomes de colunas, os valores de precisão e os valores de escala.

Você pode exibir as propriedades no modo somente leitura. Para atualizar as propriedades, edite ou faça check-out da tabela de referência.

Propriedades Gerais da Tabela de Referência

As propriedades gerais contêm informações sobre o objeto de tabela de referência.

A tabela a seguir descreve as propriedades gerais:

Propriedade	Descrição
Nome	O nome da tabela de referência.
Descrição	Qualquer descrição que um usuário inseriu para a tabela de referência.
Localização	A localização do objeto de tabela de referência no repositório do Modelo.
Coluna Válida	O nome da coluna válida na tabela de referência.
Criado em	A data e hora de criação para o nome da tabela de referência.
Criado por	O nome de logon do usuário que criou a tabela de referência.
Última Modificação	A data e hora da atualização mais recente na tabela de referência.
Última Modificação por	O nome de login do usuário que fez a atualização mais recente.
Nome da Conexão	O nome da conexão do banco de dados que armazena os valores de dados de referência.
Tipo	O tipo de tabela de referência. A tabela de referência pode ser gerenciada ou não gerenciada.

Propriedades de Coluna da Tabela de Referência

As propriedades de coluna contêm informações sobre a coluna de metadados.

A tabela a seguir descreve as propriedades de coluna

Propriedade	Descrição
Nome	O nome da coluna.
Tipo de dados	O tipo dos dados em cada coluna. É possível selecionar um dos seguintes tipos de dados: <ul style="list-style-type: none">- bigint- data/hora- decimal- duplo- número inteiro- string Você não pode selecionar um tipo de dados duplo quando cria uma tabela de referência vazia ou cria uma tabela de referência de um arquivo simples.
Precisão	A precisão de cada coluna. A precisão é o número máximo de dígitos ou o número máximo de caracteres que a coluna pode acomodar. Os valores de precisão que você configura dependem do tipo de dados.
Escala	A escala de cada coluna. A escala é o número máximo de dígitos que uma coluna pode acomodar à direita do ponto decimal. Aplica-se às colunas decimais. A escala de valores que você configura depende do tipo de dados.
Descrição	Uma descrição opcional de cada coluna.
Anulável	Indica se a coluna pode conter valores nulos.
Chave	Identifica uma coluna de chave primária em uma tabela que você importa de uma origem de banco de dados. Se você atualizar uma tabela de referência não gerenciada, não poderá adicionar valores duplicados a uma coluna que se origina de uma coluna de chave primária no banco de dados de origem. Você pode adicionar valores duplicados a qualquer coluna em uma tabela de referência gerenciada.

Criar uma Tabela de Referência no Editor da Tabela de Referência

Defina a estrutura da tabela e adicione dados à tabela de referência no editor de tabelas de referência.

1. Clique em **Novo > Tabela de Referência**.
O assistente para **Nova Tabela de Referência** é aberto.
2. Selecione a opção para **Usar o editor de tabela de referência** e clique em **Avançar**.
3. Use a opção **Adicionar Nova Coluna** para adicionar colunas à tabela.
4. Configure as propriedades para cada coluna.
As propriedades incluem o nome da coluna, o tipo de dados, a precisão e a escala.

Se a coluna contiver dados que uma transformação pode retornar em uma pesquisa de dados de referência, selecione a opção **Válida**.

5. Opcionalmente, adicione uma coluna para incluir descrições de baixo nível como metadados na tabela de referência.
6. Opcionalmente, insira uma nota de auditoria para a tabela.
A nota de auditoria é exibida no log da trilha de auditoria.
7. Clique em **Avançar**.
8. Insira um nome para a tabela de referência e selecione uma localização para o objeto de tabela de referência no repositório do Modelo.
9. Clique em **Concluir**.

Criar uma Tabela de Referência a partir de Dados do Perfil

Você pode usar dados do perfil para criar tabelas de referência que se relacionam aos dados de origem no perfil. Use as tabelas de referência para localizar diferentes tipos de informações nos dados de origem.

Você pode usar um perfil para criar ou atualizar uma tabela de referência das seguintes maneiras:

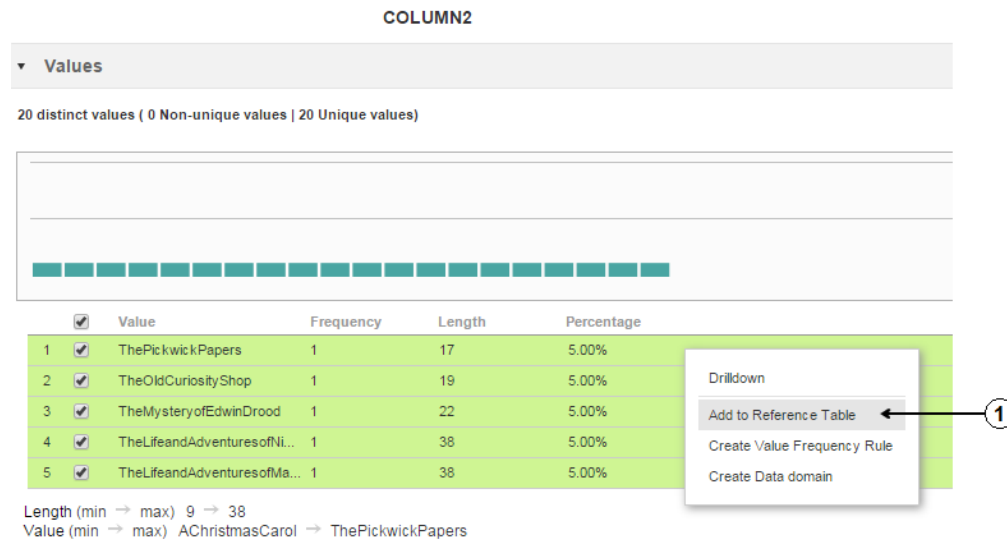
- Selecione uma coluna no perfil e a adicione a uma tabela de referência.
- Procure uma coluna de perfil e adicione um subconjunto dos dados da coluna a uma tabela de referência.
- Selecione uma coluna no perfil e adicione os valores de padrão a essa coluna para uma tabela de referência.

Criando uma Tabela de Referência a partir de Dados de Colunas de Perfil

Você pode criar uma tabela de referência a partir de um ou mais valores em uma coluna de dados de perfil. Selecione uma coluna em um perfil e escolha os valores de coluna a serem adicionados à tabela de referência.

1. Abra o espaço de trabalho **Biblioteca** na ferramenta Analyst.
2. Selecione a categoria de ativo **Perfis**.
A biblioteca exibe uma lista dos perfis no repositório do Modelo.
3. Abra o perfil que contém a coluna a ser adicionada à tabela de referência.
A visão geral do perfil lista os nomes de colunas de perfil.
4. Reveja os dados da coluna.
Para exibir os dados da coluna, clique no nome dessa coluna.
5. Na exibição de perfil detalhada, selecione os valores de dados a serem adicionados à tabela de referência. Você pode selecionar valores um por um ou pode selecionar todos eles.
6. Clique com o botão direito do mouse no nome da coluna e selecione **Adicionar à Tabela de Referência**.

A seguinte imagem mostra uma coluna de dados na exibição de perfil detalhada:



O número 1 identifica a opção **Adicionar à Tabela de Referência** na imagem.

7. O assistente para **Adicionar à Tabela de Referência** é aberto.

Selecione a opção para **Criar uma tabela de referência**.

Nota: Você também pode selecionar uma opção para adicionar os dados a uma tabela de referência atual.

8. Clique em **Avançar**.

O nome da coluna é exibido por padrão como o nome da tabela de referência. Opcionalmente, atualize o nome.

9. Opcionalmente, insira uma descrição e o valor padrão.

A ferramenta Analyst usa o valor padrão para qualquer registro de tabela que não contém um valor.

10. Clique em **Avançar**.

11. Verifique as propriedades da coluna.

Opcionalmente, opte por criar uma coluna para metadados descritivos de baixo nível.

12. Clique em **Avançar**.

13. Reveja o nome e a descrição da tabela de referência.

Opcionalmente, insira uma nota de auditoria.

14. Selecione uma localização do repositório do Modelo para o objeto de tabela de referência.

15. Clique em **Concluir**.

Criando uma Tabela de Referência a partir de Padrões de Valor

Você pode criar uma tabela de referência a partir dos padrões de coluna em uma coluna do perfil. Os padrões representam a composição dos valores de dados em um ou mais campos de coluna. Selecione uma coluna no perfil e selecione os padrões a serem adicionados à tabela de referência que você criar.

1. Abra o espaço de trabalho **Biblioteca** na ferramenta Analyst.
2. Selecione a categoria de ativo **Perfis**.

A biblioteca exibe uma lista dos perfis no repositório do Modelo.

3. Abra o perfil que contém os padrões de valor a serem adicionados à tabela de referência.

A visão geral do perfil lista os nomes de colunas de perfil.

4. Selecione a coluna que define os dados de padrão que você deseja adicionar à tabela de referência.

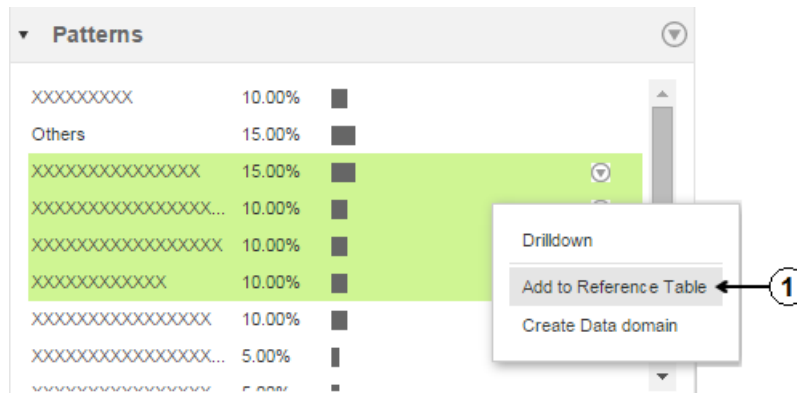
5. Reveja os padrões de dados da coluna.

Para exibir os dados da coluna, clique no nome dessa coluna.

6. Na exibição de perfil detalhada, selecione os padrões de coluna que você deseja adicionar.

7. Clique com o botão direito do mouse nos padrões que você selecionou e escolha **Adicionar à Tabela de Referência**.

A seguinte imagem mostra os padrões de dados para uma coluna na exibição de perfil detalhada:



O número 1 identifica a opção **Adicionar à Tabela de Referência** na imagem.

8. O Assistente para **Adicionar à Tabela de Referência** é aberto.

Selecione a opção para **Criar uma tabela de referência**.

Nota: Você também pode selecionar uma opção para adicionar os dados a uma tabela de referência atual.

9. Clique em **Avançar**.

O nome da coluna é exibido por padrão como o nome da tabela de referência. Opcionalmente, atualize o nome.

10. Opcionalmente, insira uma descrição e o valor padrão.

A ferramenta Analyst usa o valor padrão para qualquer registro de tabela que não contém um valor.

11. Clique em **Avançar**.

12. Verifique as propriedades da coluna.

Opcionalmente, opte por criar uma coluna para metadados descritivos de baixo nível.

13. Clique em **Avançar**.

14. Reveja o nome e a descrição da tabela de referência.

Opcionalmente, insira uma nota de auditoria.

15. Selecione uma localização do repositório do Modelo para o objeto de tabela de referência.

16. Clique em **Concluir**.

Criar uma Tabela de Referência a partir um Arquivo Simples

Os dados de referência são podem ser importados de um arquivoCSV. Use o assistente de **Nova Tabela de Referência** para importar o arquivo de dados.

Configure as propriedades para cada arquivo simples que você usar para criar uma tabela de referência.

Propriedades de Arquivo Simples da Ferramenta Analyst

Quando você importar um arquivo simples como uma tabela de referência, configure as propriedades para cada coluna no arquivo. As opções que você configura determinam como a ferramenta Analyst lê os dados do arquivo.

A tabela a seguir descreve as propriedades que podem ser configuradas no arquivo de importação de dados para uma tabela de referência:

Propriedades	Descrição
Delimitadores	Caractere usado para separar colunas de dados. Use o campo Outro para inserir um delimitador diferente. Os delimitadores deverão ser caracteres imprimíveis diferentes do caractere de escape e de aspas, se estiverem selecionados. Não é possível selecionar caracteres multibyte não imprimíveis como delimitadores.
Qualificador de Texto	O caractere de aspas que define os limites de cadeias de texto. Selecione Sem Aspas, Aspas Simples ou Aspas Duplas. Se você selecionar um caractere de aspas, o assistente ignora os delimitadores entre as aspas.
Nomes das Colunas	Importa nomes de coluna da primeira linha. Selecione-a se os nomes de coluna aparecerem na primeira linha. O assistente usa os dados na primeira linha na visualização para nomes de coluna. O padrão não está ativado.
Valores	Opção para iniciar a importação de valor de uma linha. Indica o número da linha na visualização em que o assistente começa a ler quando importa o arquivo.

Criação de uma tabela de referência de um arquivo simples

Quando você cria dados de tabela de referência de um arquivo simples, a tabela usa a estrutura de coluna do arquivo e importa os dados do arquivo.

1. Clique em **Novo > Tabela de Referência**.
O Assistente de **Nova Tabela de Referências** é exibido.
2. Selecione a opção para **Importar um arquivo simples**.
3. Clique em **Avançar**.
4. Clique em **Escolher Arquivo** para selecionar o arquivo simples.
5. Selecione uma página de código que corresponda aos dados no arquivo simples.
6. Clique em **Upload** para fazer upload dos dados desse arquivo.
7. Clique em **Avançar**.

8. Configure as propriedades do arquivo simples.
As propriedades identificam o delimitador que o arquivo utiliza e se a primeira linha desse arquivo contém nomes de coluna.
9. Para visualizar as propriedades que você configurou, atualize o painel **Visualização**.
10. Clique em **Avançar**.
11. Configure as propriedades para cada coluna.
As propriedades incluem o nome da coluna, o tipo de dados, a precisão e a escala.
Se a coluna contiver dados que uma transformação pode retornar em uma pesquisa de dados de referência, selecione a opção **Válida**.
12. Opcionalmente, adicione uma coluna para incluir descrições de baixo nível como metadados na tabela de referência.
13. Opcionalmente, insira uma nota de auditoria para a tabela.
A nota de auditoria é exibida no log da trilha de auditoria.
14. Clique em **Avançar**.
15. Insira um nome para a tabela de referência e selecione uma localização para o objeto de tabela de referência no repositório do Modelo.
16. Opcionalmente, insira uma descrição da tabela.
17. Clique em **Concluir**.

Criar uma Tabela de Referência a partir de uma Tabela de Banco de Dados

Quando você cria uma tabela de referência com base em uma tabela de banco de dados, cria um objeto de metadados no repositório do Modelo. Opcionalmente, você pode importar os dados da tabela para o data warehouse de referência.

Quando você cria uma tabela de referência gerenciada, importa os dados da coluna para o data warehouse de referência. Quando você cria uma tabela de referência não gerenciada, identifica a tabela de banco de dados que armazena os dados da coluna. Você pode criar uma tabela de referência gerenciada de uma conexão ODBC ou uma conexão nativa. Você pode criar uma tabela de referência não gerenciada de uma conexão nativa.

Antes de criar a tabela de referência, verifique se o domínio Informatica contém uma conexão ao banco de dados que contém os dados de referência. Se o domínio não contém uma conexão com o banco de dados, defina uma na ferramenta Analyst.

Para definir uma conexão de banco de dados, clique em **Gerenciar > Conexões**.

Criar uma Tabela de Referência a partir de uma Tabela de Banco de Dados

Para criar a tabela de referência, conecte a um banco de dados e selecione a tabela que contém os dados de referência.

1. Selecione **Nova > Tabela de Referência**.
O assistente de **Nova Tabela de Referência** é exibido.

2. Selecione a opção para **Conectar a uma tabela relacional**.
Para criar uma tabela de referência que não armazena dados no data warehouse de referência, selecione **Tabela não gerenciada**.
Para permitir que os usuários editem uma tabela de referência não gerenciada, selecione a opção **Editável**.
Clique em **Avançar**.
3. Selecione a conexão de banco de dados na lista de conexões.
Clique em **Avançar**.
4. No painel **Tabelas**, selecione uma tabela.
5. Reveja as propriedades da tabela no painel **Propriedades**.
Opcionalmente, clique em **Visualização de Dados** para exibir os dados da tabela.
Clique em **Avançar**.
6. No painel **Atributos da Coluna**, selecione a coluna válida.
Se você criar uma tabela de referência gerenciada, as seguintes ações no painel **Atributos da Coluna**:
 - Editar os nomes de coluna da tabela de referência.
 - Adicionar uma coluna de metadados em descrições de nível de linha.
7. Opcionalmente, adicione uma coluna para incluir descrições de baixo nível como metadados na tabela de referência.
8. Opcionalmente, insira uma nota de auditoria para a tabela.
A nota de auditoria é exibida no log da trilha de auditoria.
9. Clique em **Avançar**.
10. Insira um nome para a tabela de referência e selecione uma localização para o objeto de tabela de referência no repositório do Modelo.
11. Opcionalmente, insira uma descrição para a tabela de referência.
12. Clique em **Concluir**.

Trabalhando com Tabelas de Referência em um Repositório do Modelo com Controle de Versão

Você abre uma tabela de referência no modo somente leitura. Para trabalhar na tabela de referência, você deve entrar no modo de edição ou fazer check-out dessa tabela no repositório do Modelo.

1. Na barra de ferramentas do Informatica, clique em **Abrir**.
A biblioteca de ativos é aberta.
2. Selecione a categoria de ativos **Tabelas de Referência** e selecione um nome de tabela de referência.
A tabela de referência é aberta no modo somente leitura.
3. Para editar a versão atual da tabela de referência, clique em **Editar**.
Para editar a tabela de referência em um repositório do Modelo com controle de versão, faça check-out dessa tabela.

4. Quando concluir o trabalho na tabela de referência, clique em **Concluir**. A ferramenta Analyst salva as alterações na tabela de referência.

Se você tiver feito check-out da tabela de referência em um repositório do Modelo com controle de versão, faça check-in desse objeto. Um repositório do Modelo com controle de versão não atualizará a versão da tabela de referência até que você faça check-in do objeto.

Atualizações da Tabela de Referência

Os dados que uma tabela de referência contém podem se alterar ao longo do tempo. Revise e atualize os dados e metadados em uma tabela de referência para verificar se a tabela contém informações precisas. Atualize as tabelas de referência na ferramenta Analyst. Os dados e metadados podem ser atualizados em uma tabela de referência gerenciada e em uma não gerenciada.

Você pode executar as seguintes operações nos dados da tabela de referência e metadados:

Gerenciar colunas

É possível adicionar e excluir colunas e editar propriedades de colunas.

Gerenciar linhas

É possível adicionar linhas de dados a uma tabela de referência.

Editar os valores de dados de referência

É possível editar um valor de dados de referência.

Substituir valores de dados

Use a opção **Localizar e Substituir** para substituir os valores de dados que não são mais precisos ou relevantes para a organização. Localize um valor em uma coluna e substitua por outro. Também é possível substituir todos os valores em uma coluna com um único valor.

Exportar uma tabela de referência

Exporte uma tabela de referência para um arquivo de valores separados por vírgula (CSV), arquivo de dicionário ou arquivo do Excel.

Ativar ou desativar edições em uma tabela não gerenciada

Atualize uma tabela de referência não gerenciada para ativar ou desativar edições para dados de tabela e metadados.

Atualizar os dados da tabela de referência

Recarregue os dados da tabela de referência na ferramenta Analyst para exibir as últimas alterações feitas.

Gerenciar Colunas

Você pode adicionar colunas a uma tabela de referência e atualizar as propriedades dessas colunas. Também pode atualizar o status editável de uma tabela de referência não gerenciada.

1. Clique em **Abrir**.
A biblioteca de ativos é aberta.
2. Selecione a categoria de ativos **Tabelas de Referência** e selecione um nome de tabela de referência.
A tabela de referência é aberta no modo somente leitura.

3. Para editar a versão atual da tabela de referência, clique em **Editar**.
Para editar a tabela de referência em um repositório do Modelo com controle de versão, faça check-out dessa tabela.
4. Abra o menu **Ações** e selecione **Alterar Propriedades da Coluna**.
A caixa de diálogo **Alterar propriedades da coluna** é aberta. Use as opções da caixa de diálogo para realizar as seguintes operações:
 - Adicionar uma coluna.
 - Alterar a coluna válida na tabela.
 - Alterar o nome da coluna.
 - Atualizar o texto descritivo para uma coluna.
 - Atualizar o status editável de uma tabela de referência não gerenciada.
 - Atualizar a nota de auditoria para a tabela.
5. Quando concluir as operações, clique em **OK**.

Gerenciar Linhas

É possível adicionar, editar ou excluir linhas em uma tabela de referência.

1. Clique em **Abrir**.
A biblioteca de ativos é aberta.
2. Selecione a categoria de ativos **Tabelas de Referência** e selecione um nome de tabela de referência.
A tabela de referência é aberta no modo somente leitura.
3. Para editar a versão atual da tabela de referência, clique em **Editar**.
Para editar a tabela de referência em um repositório do Modelo com controle de versão, faça check-out dessa tabela.
4. Edite as linhas de dados. Você pode editar as linhas de dados das seguintes maneiras:
 - Para adicionar uma linha, selecione **Ações > Adicionar Linha**.
Na caixa de diálogo **Adicionar Linha**, insira um valor na coluna válida e em pelo menos outra coluna. Opcionalmente, insira uma nota de auditoria.
Clique em **OK** para adicionar a linha.
 - Para atualizar um único valor de dados, clique nesse valor e atualize os dados.
Depois de atualizar os dados, use as opções em nível de linha para aceitar ou rejeitar esses dados. Não é possível inserir uma nota de auditoria ao inserir dados diretamente na linha de dados.
 - Para atualizar os valores de dados em uma linha, selecione **Ações > Editar Linha**.
Na caixa de diálogo **Editar Linha**, insira um valor em uma ou mais colunas. Opcionalmente, insira uma nota de auditoria.
Clique em **Aplicar** para atualizar os dados nas colunas que você selecionou.
 - Para atualizar os valores em várias linhas, selecione as linhas a serem editadas e escolha **Ações > Editar Linha**.
Na caixa de diálogo **Editar Várias Linhas**, insira um valor em uma ou mais colunas. Opcionalmente, insira uma nota de auditoria.
Clique em **OK** para atualizar os dados nas colunas que você selecionou.

- Para excluir linhas, selecione as linhas para excluir e clique em **Ações > Excluir**.

Na caixa de diálogo **Excluir Linhas**, insira opcionalmente uma nota de auditoria.

Clique em **OK** para excluir as linhas.

Nota: Use a ferramenta Developer para editar dados de linha em uma grande tabela de referência. Por exemplo, se uma tabela de referência contém mais de 500 linhas, edite a tabela na ferramenta Developer.

Localizar e Substituir Valores

É possível localizar e substituir valores de dados em uma tabela de referência. Use as opções de localizar e substituir quando uma tabela contiver uma ou mais instâncias de um valor de dados que você precisa atualizar.

1. Clique em **Abrir**.
A biblioteca de ativos é aberta.
2. Selecione a categoria de ativos **Tabelas de Referência** e selecione um nome de tabela de referência.
A tabela de referência é aberta no modo somente leitura.
3. Para editar a versão atual da tabela de referência, clique em **Editar**.
Para editar a tabela de referência em um repositório do Modelo com controle de versão, faça check-out dessa tabela.
4. Clique em **Ações > Localizar e Substituir**.
A barra de ferramentas **Localizar e Substituir** é exibida.
5. Insira os critérios de pesquisa na barra de ferramentas:
 - Insira um valor de dados no campo **Localizar**.
 - Selecione as colunas para pesquisa. Por padrão, a operação pesquisa todas as colunas.
 - Insira um valor de dados no campo **Substituir por**.
6. Use as seguintes opções para substituir valores um por um ou para substituir todos os valores:
 - Use as opções **Próximo** e para localizar valores um por um.
 - Para substituir um valor, selecione **Substituir**.
 - Para exibir todas as instâncias do valor, selecione **Realçar Todos**.
 - Para substituir todas as instâncias do valor, selecione **Substituir Todos**.

Exportando Dados da Tabela de Referência

Exporte os dados em uma tabela de referência para um arquivo separado por vírgulas, um arquivo de dicionário ou um arquivo do Microsoft Excel. Você pode exportar os dados no modo somente leitura.

1. Clique em **Abrir**.
A biblioteca de ativos é aberta.
2. Selecione a categoria de ativos **Tabelas de Referência** e selecione um nome de tabela de referência.
A tabela de referência é aberta no modo somente leitura.
3. Clique em **Ações > Exportar Dados**.
A caixa de diálogo **Exportar os dados para um arquivo** é exibida.

A seguinte tabela descreve as opções da caixa de diálogo:

Opção	Descrição
Nome do Arquivo	Nome do arquivo que conterá os dados. A operação de exportação cria o arquivo.
Formato de Arquivo	Formato do arquivo que conterá os dados. Selecione um dos seguintes formatos: <ul style="list-style-type: none">• csv. Arquivo separado por vírgula. Formato padrão.• xls. Arquivo do Microsoft Excel.• dic. Arquivo de dicionário da Informatica.
Exportar nomes de campo como primeira linha	Opção de nome da coluna. Selecione a opção para indicar que a primeira linha do arquivo contém os nomes de coluna.
Página de Código	Página de código dos dados de referência. A página de código padrão é UTF-8.

4. Clique em **OK** para exportar o arquivo.

Ativar e Desativar Edições em uma Tabela de Referência não Gerenciada

Você pode ativar ou desativar atualizações nos valores e colunas de dados em uma tabela de referência não gerenciada.

Antes de alterar o status editável da tabela de referência, salve a tabela.

1. Clique em **Abrir**.
A biblioteca de ativos é aberta.
2. Selecione a categoria de ativos **Tabelas de Referência** e selecione um nome de tabela de referência.
A tabela de referência é aberta no modo somente leitura.
3. Para editar a versão atual da tabela de referência, clique em **Editar**.
Para editar a tabela de referência em um repositório do Modelo com controle de versão, faça check-out dessa tabela.
4. Abra o menu **Ações** e selecione **Alterar Propriedades da Coluna**.
A caixa de diálogo **Alterar propriedades da coluna** é aberta.
5. Selecione ou desmarque a opção **Editável**.

Nota: As seguintes condições se aplicam a uma tabela de referência não gerenciada que permite atualizações do usuário:

- A tabela de referência deve usar tipos de dados simples, como cadeia e número .
- Não defina restrições nos metadados da tabela de referência ou especifique um valor padrão para qualquer coluna.

Atualizar os Valores da Tabela de Referência

Talvez você precise atualizar os valores que a ferramenta Analyst exibe para a tabela de referência.

Para recarregar os valores da tabela de referência, clique em **Ações > Atualizar**. A ferramenta Analyst recupera as versões atuais dos valores de dados do banco de dados.

Eventos de Trilha de Auditoria

É possível exibir uma trilha de auditoria das alterações feitas pelos usuários em uma tabela de referência. Use a exibição Trilha de Auditoria na tabela de referência para exibir os eventos de trilha de auditoria. Você pode filtrar os eventos de trilha de auditoria que são exibidos pela ferramenta Analyst.

A seguinte tabela descreve as opções de filtro que você pode especificar:

Opção	Descrição
Data	Datas de início e de término para as ações a serem exibidas. Use as opções de calendário para definir as datas.
Tipo	<p>Tipo de evento de trilha de auditoria. É possível exibir os seguintes tipos de eventos:</p> <ul style="list-style-type: none">- Dados. Eventos que se relacionam aos valores de dados na tabela de referência. Os eventos incluem operações para adicionar, excluir e atualizar uma linha.- Metadados. Os eventos que se relacionam aos metadados da tabela de referência. Eventos incluem operações para criar a tabela de referência, adicionar ou excluir uma coluna e fazer check-in da tabela de referência. <p>Nota: Não é possível exibir eventos de dados e metadados simultaneamente.</p>
Usuário	Usuário que editou a tabela de referência. O filtro exibe o nome completo e o nome de logon do usuário.
Status	O status dos eventos de log da trilha de auditoria. O status corresponde à ação executada no editor de tabela de referência. Por exemplo, ele pode indicar que um usuário criou a tabela de referência ou adicionou uma linha.

Os eventos de log da trilha de auditoria também incluem os comentários da trilha de auditoria e os valores de coluna que foram inseridos, atualizados ou excluídos.

Exibir Eventos de Trilha de Auditoria

Exiba eventos de trilha de auditoria para descobrir mais sobre as atualizações que os usuários fizeram em uma tabela de referência. Você pode exibir os eventos de trilha de auditoria no modo somente leitura.

1. Clique em **Abrir**.
A biblioteca de ativos é aberta.
2. Selecione a categoria de ativos **Tabelas de Referência** e selecione um nome de tabela de referência.
A tabela de referência é aberta no modo somente leitura.
3. Clique na **Trilha de Auditoria**.
4. Configure as opções de filtro.
É possível filtrar por data de atualização, tipo de atualização, status da atualização e nome do usuário que realizou a atualização.
5. Clique em **Mostrar**.
Os eventos de log aparecem para as opções de filtro que você especificou.

Regras e Diretrizes para Tabela de Referência

Use as seguintes regras e diretrizes ao trabalhar com tabelas de referência na ferramenta Analyst:

- Quando você importar uma tabela de referência do Oracle, IBM DB2 ou banco de dados Microsoft SQL Server, a ferramenta Analyst não poderá exibir a visualização se os nomes de tabela, exibição, esquema, sinônimo ou coluna contiverem caracteres com letras minúsculas ou com letras maiúsculas e minúsculas.

Para visualizar os dados em tabelas que residem em bancos de dados com diferenciação de maiúsculas e minúsculas, defina o atributo Suporte a Identificadores de Letras Maiúsculas e Minúsculas na conexão de banco de dados como verdadeiro.

- Quando você criar uma tabela de referência de padrões de coluna inferidos em um formato, a ferramenta Analyst preenche a tabela de referência com padrões de coluna em um formato diferente.

Por exemplo, quando você criar uma tabela de referência para a coluna padrão X (5), a ferramenta Analyst exibirá o seguinte formato para a coluna padrão na tabela de referência: XXXXX.

- Quando você importar uma tabela de banco de dados Oracle, verifique o tamanho de qualquer coluna VARCHAR2 na tabela. A ferramenta Analyst não pode importar uma tabela de banco de dados Oracle que contenha uma coluna VARCHAR2 com um tamanho maior que 1000.

- Para ler uma tabela de referência, execute as permissões na conexão ao banco de dados que armazena os valores dos dados da tabela. Por exemplo, se o data warehouse de referência armazena os valores de dados, você precisa executar as permissões na conexão ao data warehouse de referência. Execute as permissões para acessar a tabela de referência no modo de leitura ou gravação. As permissões de conexão de banco de dados se aplicam a todos os dados de referência no banco de dados.

- Quando você executar um mapeamento com uma transformação que especifica uma tabela de referência, o mapeamento usará a versão atual da tabela de referência no repositório do modelo. Não é possível selecionar uma versão histórica da tabela de referência ao configurar a transformação.

Se outro usuário restaurar a tabela de referência para uma versão anterior em uma sessão simultânea da Developer tool, as versões da tabela de referência não serão mais idênticas nas sessões. Se você configurar e executar um mapeamento que usa a tabela de referência, o mapeamento poderá falhar porque a sessão atual não identificará a versão atual da tabela de referência. Para garantir que o mapeamento use a tabela de referência atual, atualize o repositório do modelo antes de executar o mapeamento.

- Quando você configura uma tabela de referência não gerenciada para permitir edições, verifique se a tabela de referência usa tipos de dados simples, como string e número. Além disso, verifique se a tabela de referência não define qualquer restrição nos metadados da tabela de referência ou use valores padrão para colunas.

CAPÍTULO 3

Dados de Referência na Ferramenta Developer

Este capítulo inclui os seguintes tópicos:

- [Visão Geral dos Dados de Referência da Ferramenta Developer, 29](#)
- [Transformações e Dados de Referência, 30](#)
- [Trabalhando com Objetos de Dados de Referência em um Repositório do Modelo com Controle de Versão, 30](#)
- [Tabelas de Referência, 31](#)
- [Conjuntos de Conteúdo, 36](#)

Visão Geral dos Dados de Referência da Ferramenta Developer

Você pode criar, atualizar e exibir as propriedades de configuração de objetos de dados de referência na ferramenta Developer.

Use a ferramenta Developer para criar e atualizar os seguintes tipos de objeto:

Tabelas de referência

Uma tabela de referência contém as versões padrão e alternativa de um conjunto de valores de dados. Adicione uma tabela de referência a uma transformação na ferramenta Developer para verificar se os valores de dados de origem estão precisos e formatados corretamente.

Conjuntos de Conteúdo

Um conjunto de conteúdo é um objeto de repositório do Modelo que especifica valores de dados de referência no repositório ou em um arquivo. Um conjunto de conteúdo inclui diferentes tipos de dados de referência que você pode usar para realizar operações de pesquisa em transformações de qualidade de dados.

Você também pode trabalhar com arquivos de dados de referência de endereço e arquivos de preenchimento de identidade na ferramenta Developer. Você seleciona arquivos de dados de referência de endereço ao configurar uma transformação de Validador de Endereço. Você seleciona arquivos de preenchimento de identidade ao configurar uma transformação de Correspondência para a análise de correspondência de identidades.

Transformações e Dados de Referência

Várias transformações leem os dados de referência para executar tarefas de qualidade de dados.

As seguintes transformações podem ler dados de referência:

- Validador de Endereço. Lê dados de referência de endereço para verificar a precisão dos endereços.
- Conversor de Maiúsculas/Minúsculas. Lê tabelas de dados de referência para identificar as strings que devem alterar maiúsculas e minúsculas.
- Classificador. Lê dados do conjunto de conteúdo para identificar o tipo de informação em uma string.
- Comparação. Lê dados de preenchimento de identidade durante a análise duplicada.
- Rotulador. Lê dados do conjunto de conteúdo para identificar e rotular strings.
- Correspondência. Lê dados de preenchimento de identidade durante a análise duplicada.
- Analisador. Lê dados do conjunto de conteúdo para analisar as strings com base no que as informações contêm.
- Padronizador. Lê tabelas de dados de referência para padronizar strings em um formato comum.

O conjunto de arquivos do Instalador de Conteúdo do Data Quality inclui objetos de dados de referência da Informática que podem ser importados.

Trabalhando com Objetos de Dados de Referência em um Repositório do Modelo com Controle de Versão

Se você trabalhar com tabelas de referência ou conjuntos de conteúdo em um repositório do Modelo com controle de versão, esse repositório poderá aplicar controle de versão aos objetos. Para aplicar controle de versão a um objeto, um usuário faz check-in desse objeto no repositório do Modelo.

Se uma tabela de referência ou um conjunto de conteúdo não estiver sob controle de versão, você poderá abrir e atualizar o objeto fora do sistema de controle de versão. Quando você abre o objeto, o repositório do Modelo o bloqueia para que outro usuário não possa trabalhar nele.

Se uma tabela de referência ou um conjunto de conteúdo estiver sob controle de versão, você o abrirá no modo somente leitura. Para trabalhar no objeto, faça o seu check-out do repositório do Modelo. Como alternativa, confira o objeto e abra-o em seguida. Faça check-in do objeto para criar uma versão dele que contenha as suas alterações mais recentes.

Fazendo Check-Out de Objeto de Dados de Referência

Para trabalhar em uma tabela de referência ou um conjunto de conteúdo que um usuário submeteu a check-in no repositório do Modelo, faça check-out desse objeto no repositório.

1. No Object Explorer, navegue até uma tabela de referência ou um conjunto de conteúdo.
2. Clique com o botão direito do mouse no nome do objeto e depois em **Abrir**.
O objeto é aberto no modo somente leitura.
3. Clique com o botão direito do mouse no nome do objeto e depois em **Fazer Check-Out**.
Você pode editar o objeto.

Fazendo Check-In de Objeto de Dados de Referência

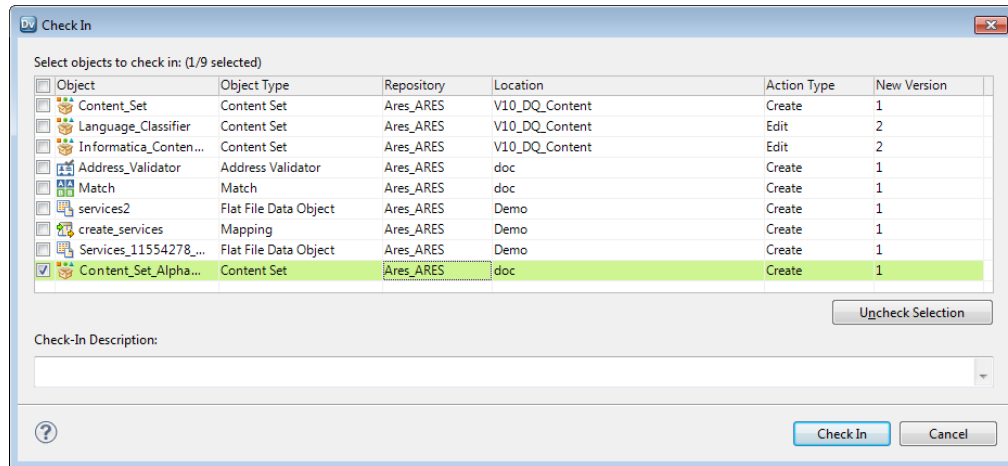
Quando terminar o trabalho em uma tabela de referência ou em um conjunto de conteúdo que você colocou em check-out do repositório do Modelo, faça o check-in do objeto.

Para exibir a lista de objetos atualmente em estado de check-out, abra a guia **Objetos com Check-out** abaixo do editor de tabela de referência.

1. Salve qualquer alteração feita na tabela de referência ou conjunto de conteúdo.
2. No Object Explorer, navegue até a tabela de referência ou o conjunto de conteúdo.
3. Clique com o botão direito do mouse no nome do objeto e depois em **Fazer Check-In**.

A caixa de diálogo **Fazer Check-In** é aberta.

A seguinte imagem mostra a caixa de diálogo:



4. Selecione um ou mais objetos para check-in no repositório.

Nota: Você pode fazer check-in de um objeto que não está aberto na sessão atual. Você pode fazer check-in de qualquer objeto em estado de check-out.

5. Opcionalmente, insira uma descrição para a operação.
6. Clique em **Fazer Check-In**.

A operação de check-in atualiza o número de versão do objeto. Se você fizer check-in do objeto pela primeira vez, o repositório do Modelo criará versão um (1) desse objeto.

Tabelas de Referência

Adicione uma tabela de referência a uma transformação na Developer tool. Você configura a transformação para encontrar valores da tabela de referência em dados de entrada e para gravar os valores válidos correspondentes da tabela de referência como saída.

Para criar uma tabela de referência na Developer tool, use um dos seguintes métodos:

- Crie uma tabela de referência vazia e insira os valores de dados.
- Crie uma tabela de referência de dados a partir um arquivo simples.
- Crie uma tabela de referência de dados em uma tabela de banco de dados, um sinônimo ou uma exibição.

Propriedades de Dados da Tabela de Referência

Você pode exibir propriedades para os dados da tabela de referência e metadados na ferramenta Developer. A ferramenta Developer exibe as propriedades quando você abre a tabela de referência do repositório do Modelo.

Uma tabela de referência exibe as propriedades gerais e as propriedades de coluna. Você pode exibir as propriedades da tabela de referência na ferramenta Developer. As propriedades da tabela de referência podem ser exibidas e editadas na ferramenta Analyst.

A tabela a seguir descreve as propriedades gerais de uma tabela de referência:

Propriedade	Descrição
Nome	Nome da tabela de referência.
Descrição	Descrição opcional da tabela de referência.

A tabela a seguir descreve as propriedades de coluna de uma tabela de referência:

Propriedade	Descrição
Válido	Identifica a coluna que contém os dados de referência válidos.
Nome	Nome de cada coluna.
Tipo de Dados	O tipo de dados dos dados em cada coluna.
Precisão	A precisão de cada coluna.
Escala	A escala de cada coluna.
Descrição	A descrição do conteúdo da coluna. A descrição pode ser adicionada quando você cria a tabela de referência.
Inclui uma coluna para descrições de nível baixo	Indica que a tabela de referência contém uma coluna para descrições dos dados de coluna.
Valor padrão	O valor padrão dos campos na coluna. Opcionalmente, você pode adicionar um valor padrão quando cria a tabela de referência.
Nome da Conexão	O nome da conexão com o banco de dados que contém os valores de dados de tabelas de referência.

Criando um Objeto de Tabela de Referência

Selecione essa opção quando você desejar criar uma tabela de referência vazia e adicionar valores manualmente.

1. Selecione **Arquivo > Novo > Tabela de Referência** no menu da ferramenta Developer.
2. No assistente de nova tabela, selecione **Tabela de Referência como Vazia**.
3. Insira um nome para a tabela.
4. Selecione um projeto para armazenar os metadados da tabela.

No campo Localização, clique em **Procurar**. A caixa de diálogo **Selecionar Localização** se abre e exibe os projetos no repositório. Selecione o projeto.

Clique em **Avançar**.

5. Adicione duas ou mais colunas à tabela. Clique na opção **Novo** para criar uma coluna.

A tabela a seguir descreve as propriedades de cada coluna:

Propriedade	Valor Padrão
Nome	coluna
Tipo de dados	string
Precisão	10
Escala	0
Descrição	Vazio. Propriedade opcional.

6. Selecione a coluna que contém os valores válidos. A ordem das colunas que você criar pode ser alterada.

7. A tabela a seguir descreve as propriedades opcionais:

Propriedade	Valor Padrão
Incluir uma coluna de descrições de nível de linha	Limpo
Nota de auditoria	Vazio
Valor padrão	Vazio

Clique em **Concluir**.

A tabela de referência é aberta no espaço de trabalho da ferramenta Developer.

Criar uma Tabela de Referência a partir de um Arquivo Simples

É possível criar uma tabela de referência a partir dos dados armazenados em um arquivo simples.

1. Selecione **Arquivo > Novo > Tabela de Referência** no menu da ferramenta Developer.
2. No assistente de nova tabela, selecione **Tabela de Referência a partir de um Arquivo Simples**.
3. Navegue para o arquivo que você deseja usar como a fonte de dados para a tabela.
4. Insira um nome para a tabela.
5. Selecione um projeto para armazenar os metadados da tabela.

No campo Localização, clique em **Procurar**. A caixa de diálogo **Selecionar Localização** se abre e exibe os projetos no repositório. Selecione o projeto.

Clique em **Avançar**.

6. Defina UTF-8 como a página de código.
7. Especifique o delimitador que o arquivo simples usa.
8. Se o arquivo simples contém nomes de coluna, selecione a opção para importar os nomes de coluna da primeira linha do arquivo.

9. A tabela a seguir descreve as propriedades opcionais de tabela:

Propriedade	Valor Padrão
Qualificador de texto	Sem aspas
Iniciar importação na linha	Linha 1
Delimitador de Linhas	\012 LF (\n)
Tratar delimitadores consecutivos como um	Limpo
Caractere de escape	Vazio
Manter caractere de escape nos dados	Limpo
Máximo de linhas para visualização	500

Clique em **Avançar**.

10. Selecione a coluna que contém os valores válidos.
11. A tabela a seguir descreve as propriedades opcionais:

Propriedade	Valor Padrão
Incluir uma coluna de descrições de nível de linha	Limpo
Nota de auditoria	Vazio
Valor padrão	Vazio
Máximo de linhas para visualização	500

Clique em **Concluir**.

A tabela de referência é aberta no espaço de trabalho da ferramenta Developer.

Criar uma Tabela de Referência de uma Origem Relacional

Você pode criar uma tabela de referência com base em uma tabela relacional, um sinônimo ou uma exibição.

Quando você cria uma tabela de referência gerenciada, importa os dados da coluna para o data warehouse de referência. Quando você cria uma tabela de referência não gerenciada, identifica a tabela de banco de dados que armazena os dados da coluna. Você pode criar uma tabela de referência gerenciada de uma conexão ODBC ou uma conexão nativa. Você pode criar uma tabela de referência não gerenciada de uma conexão nativa.

Antes de criar a tabela de referência, verifique se o domínio Informatica contém uma conexão ao banco de dados que contém os dados de referência.

É possível configurar uma conexão de banco de dados no Explorador de Conexões. Se a Developer tool não mostrar o Explorador de Conexões, selecione **Janela > Mostrar Exibição > Explorador de Conexões** no menu da Developer tool.

Criando uma Tabela de Referência de uma Origem Relacional

Para criar a tabela de referência, conecte a um banco de dados e selecione a tabela que contém os dados de referência.

1. Selecione **Arquivo > Novo > Tabela de Referência** no menu da Developer tool.
2. No assistente de nova tabela, selecione **Tabela de Referência com base em uma Origem Relacional**.
Clique em **Avançar**.
3. Selecione uma conexão de banco de dados.
No campo de conexão, clique em **Procurar**. A caixa de diálogo **Escolher Conexão** é aberta e exibe as conexões de banco de dados disponíveis.
Clique em **OK** quando você selecionar uma conexão.
4. Selecione um recurso do banco de dados.
No campo de Recurso, clique em **Procurar**. A caixa de diálogo **Selecionar um Recurso** é aberta e exibe os recursos na conexão de banco de dados. Explore o banco de dados e selecione uma tabela de banco de dados, um sinônimo ou uma exibição.
Opcionalmente, você pode visualizar as informações de entidade no recurso.
5. Insira um nome para a tabela.
6. Selecione uma localização para o objeto de tabela de referência.
No campo Localização, clique em **Procurar**. A caixa de diálogo **Selecionar Localização** se abre e exibe os projetos no repositório.
Selecione uma localização e clique em **Avançar**.
7. Para criar uma tabela de referência que não armazena dados no data warehouse de referência, selecione **Tabela não gerenciada**.
Para possibilitar que os usuários editem uma tabela de referência não gerenciada, selecione a opção **Editável**.
Clique em **Avançar**.
8. Selecione a coluna que contém os valores válidos.
9. A tabela a seguir descreve propriedades opcionais que você pode especificar:

Propriedade	Valor Padrão
Incluir uma coluna de descrições de nível de linha	Limpo
Descrição	Limpo
Valor padrão	Vazio
Nota de auditoria	Vazio
Máximo de linhas para visualização	500

10. Clique em **Concluir**.

Conjuntos de Conteúdo

Um conjunto de conteúdo é um objeto de repositório do Modelo que armazena dados ou metadados de outros objetos de dados de referência. Um conjunto de conteúdo pode incluir conjuntos de caracteres, conjuntos de padrões, conjuntos de tokens, expressões regulares, modelos probabilísticos e modelos de classificador. Use um conjunto de conteúdo para definir e organizar objetos de dados de referência que estão relacionados a um único projeto, tipo de informações ou propósito de negócios.

A ferramenta Developer inclui conjuntos de caracteres e conjuntos de tokens definidos pelo sistema que não aparecem no repositório do Modelo. Para exibir e usar os objetos definidos pelo sistema, configure uma estratégia na transformação de Rotulador, Analisador ou Padronizador.

Conjuntos de Caracteres

Os conjuntos de caracteres contêm expressões que identificam caracteres específicos e intervalos de caracteres. Os conjuntos de caracteres podem ser usados em transformações do Rotulador que usam o modo de rotulagem de caractere.

As faixas de caractere especificam uma faixa sequencial de códigos de caractere. Por exemplo, a faixa de caractere "[A-C]" corresponde aos caracteres maiúsculos "A," "B" e "C." Essa faixa de caractere não corresponde aos caracteres minúsculos "a," "b" ou "c."

Use conjuntos de caracteres para identificar um caractere específico ou faixas de caracteres como parte de operações de rotulagem. Por exemplo, você poderá rotular todos os número de uma coluna que contém números de telefone. Depois de rotular os números, identifique os padrões com uma transformação de Analisador e grave padrões problemáticos em portas de saída separadas.

Propriedades do Conjunto de Caracteres

Configure as propriedades que determinam as operações de rotulagem de caractere para um conjunto de caracteres.

A tabela a seguir descreve as propriedades definidas pelo usuário para um conjunto de caracteres:

Propriedade	Descrição
Rótulo	Define o rótulo que uma transformação de Rotulador aplica aos dados que correspondem ao conjunto de caracteres.
Modo Padrão	Habilita uma exibição de edição simples que inclui campos para as faixas inicial e final.
Faixa Inicial	Especifica o primeiro caractere em uma faixa de caractere.
Faixa Final	Especifica o último caractere em uma faixa de caractere. Para uma faixa com um único caractere, deixe este campo em branco.
Modo Avançado	Habilita uma exibição de edição avançada onde você pode inserir manualmente as faixas de caractere usando caracteres de faixa e caracteres delimitadores.
Caractere da Faixa	Altera temporariamente o símbolo que significa uma faixa de caractere. A faixa de caractere reverterá para o caractere padrão quando você fechar o conjunto de caracteres.
Caractere Delimitador	Altera temporariamente o símbolo que separa as faixas de caractere. O caractere delimitador reverterá para o caractere padrão quando você fechar o conjunto de caracteres.

Modelos do Classificador

Um modelo do classificador analisa as cadeias de entrada e determina os tipos de informações provavelmente contidos nas cadeias. Use um modelo do classificador em uma transformação do Classificador.

Um modelo do classificador contém linhas de dados de referência e valores de rótulos. As linhas representam os dados de entrada na porta que você pode conectar à transformação do Classificador. Os valores de rótulos descrevem os tipos de informações que as linhas de dados contêm. Ao configurar um modelo do classificador, atribua um rótulo a cada linha de dados de referência no modelo.

Para vincular as linhas de dados de referência aos rótulos em um modelo do classificador, compile o modelo. O processo de compilação gera uma série de associações lógicas entre as linhas de dados e os valores de rótulo. Quando você executa um mapeamento que lê o modelo, o Serviço de Integração de Dados aplica a lógica do modelo aos dados de entrada da transformação do Classificador. O Serviço de Integração de Dados retorna os rótulos que descrevem mais com precisão as informações em cada campo de dados de entrada.

Você pode criar um modelo do classificador na Developer tool. O repositório do Modelo armazena o objeto do modelo do classificador. A Developer tool grava as linhas de dados, os rótulos e a compilação de dados em um arquivo na estrutura de diretórios da Informatica.

Conjuntos de Padrões

Um conjunto de padrão contém expressões que identificam padrões de dados na saída de uma operação de rotulagem de token. Você pode usar conjuntos de padrão para analisar a porta de saída de Dados com Token e gravar a correspondência de strings em uma ou mais portas de saída. Use os conjuntos de padrão nas transformações do Analisador que usem o modo de análise de padrão.

Por exemplo, você pode configurar uma transformação do Analisador para usar conjuntos de padrão que identifiquem nomes e iniciais. Essa transformação usa os conjuntos de padrão para analisar a saída de uma transformação do Rotulador no modo de rotulagem de token. Você pode configurar a transformação do Analisador para gravar os nomes e as iniciais na saída para portas diferentes.

Propriedades do Conjunto de Padrão

Configure as propriedades que determinam os padrões em um conjunto de padrão.

A tabela a seguir descreve a propriedade para um conjunto de dados definido pelo usuário:

Propriedade	Descrição
Padrão	Define os padrões que o analisador de padrão pesquisa. Você pode inserir vários padrões para um conjunto de padrão. Insira padrões criados a partir de uma combinação de curingas, caracteres e strings.

Modelos Probabilísticos

Um modelo probabilístico analisa os valores de dados de entrada e determina os tipos de informações provavelmente contidos nos valores. Use um modelo probabilístico em uma transformação de Rotulador e em uma transformação do Analisador.

Um modelo probabilístico contém valores de dados de referência e de rótulos. Os valores de dados de referência representam os dados em uma porta de entrada que você conectar à transformação. Os valores

de rótulos descrevem os tipos de informações contidas nos valores de dados de referência. Atribua um rótulo para cada valor de dados de referência no modelo.

Para vincular os valores de dados de referência aos rótulos em um modelo probabilístico, compile o modelo. O processo de compilação gera uma série de associações lógicas entre os valores de dados e os rótulos. Quando você executa um mapeamento que lê o modelo, o Serviço de Integração de Dados aplica a lógica de modelo aos dados de entrada da transformação. O Serviço de Integração de Dados retorna o rótulo descreve os valores de dados de entrada com mais precisão.

Você pode criar um modelo probabilístico na Developer tool. O repositório do Modelo armazena o objeto do modelo probabilístico. A Developer tool grava os valores de dados, os rótulos e os dados de compilação em um arquivo na estrutura de diretórios da Informatica.

Expressões Regulares

No contexto de conjuntos de conteúdo, uma expressão regular é uma expressão que você pode usar em operações de análise e rotulagem. Use expressões regulares para identificar uma ou mais strings em dados de entrada. Você pode usar expressões regulares nas transformações de Analisador que usam o modo de análise de token. Você também pode usar expressões regulares nas transformações do Rotulador que usam o modo de rotulagem de token.

As transformações do Analisador usam expressões regulares para combinar padrões em dados de entrada e analisar todas as strings correspondentes para uma ou mais saídas. Por exemplo, é possível usar uma expressão regular para identificar todos os e-mails em dados de entrada e analisar cada componente de e-mail para uma saída diferente.

As transformações do Rotulador usam expressões regulares para corresponder a um padrão de entrada e criar um único rótulo. Expressões regulares que têm várias saídas não geram vários rótulos.

Propriedades de Expressão Regular

Configure as propriedades que determinam como uma expressão regular identifica e grava a saída de strings.

A tabela a seguir descreve as propriedades definidas pelo usuário para uma expressão regular:

Propriedade	Descrição
Número de Saídas	Define o número de portas de saída que a expressão regular grava.
Expressão Regular	Define um padrão que a transformação do Analisador usa para corresponder strings.
Expressão de Teste	Contém dados que você insere para testar a expressão regular. Como o tipo de dados nesse campo, o campo realça as strings que correspondem à expressão regular.
Próxima Expressão	Move para a próxima string que corresponda à expressão regular e altera a fonte da string para negrito.
Expressão Anterior	Move para a string anterior que corresponda à expressão regular e altera a fonte da string para negrito.

Conjuntos de Tokens

Um conjunto de token contém as expressões que identificam tokens específicos. Você pode usar conjuntos de tokens em transformações do Rotulador que usam o modo de rotulagem de token. Também pode usar conjuntos de tokens em transformações do Analisador que usam o modo de análise de token.

Use os conjuntos de tokens para identificar tokens específicos como parte das operações de rotulagem e análise. Por exemplo, você pode usar um conjunto de tokens para rotular todos os e-mails que usem um formato "AccountName@DomainName". Depois da rotulagem de tokens, use a transformação do Analisador para gravar e-mail para as portas de saída que você especificar.

Propriedades do Conjunto de Token

Configure propriedades que especificam as operações de rotulagem de para um conjunto de token.

A tabela a seguir descreve as propriedades definidas pelo usuário para um conjunto de caracteres:

Propriedade	Modo de Conjunto de Token	Descrição
Nome	N/A	Define o nome do conjunto de token.
Descrição	N/A	Descreve o conjunto de token.
Opções de Definição de Token	N/A	Define se o conjunto de token usa o modo de expressão regular ou modo de caractere.
Rótulo	Expressão Regular	Define o rótulo que uma transformação do Rotulador aplica aos dados que correspondem ao conjunto de token.
Expressão Regular	Expressão Regular	Define um padrão que a transformação do Rotulador usa para corresponder strings.
Expressão de Teste	Expressão Regular	Contém dados que você insere para testar a expressão regular. Como o tipo de dados nesse campo, o campo realça as strings que correspondem à expressão regular.
Próxima Expressão	Expressão Regular	Move para a próxima string que corresponda à expressão regular e altera a fonte da string para negrito.
Expressão Anterior	Expressão Regular	Move para a string anterior que corresponda à expressão regular e altera a fonte da string para negrito.

Propriedade	Modo de Conjunto de Token	Descrição
Rótulo	Caractere	Define o rótulo que uma transformação de Rotulador aplica aos dados que correspondem ao conjunto de caracteres.
Modo Padrão	Caractere	Habilita uma exibição de edição simples que inclui campos para as faixas inicial e final.
Faixa Inicial	Caractere	Especifica o primeiro caractere em uma faixa de caractere.
Faixa Final	Caractere	Especifica o último caractere em uma faixa de caractere. Para faixas com um único caractere, deixe este campo em branco.
Modo Avançado	Caractere	Habilita uma exibição de edição avançada onde você pode inserir manualmente as faixas de caractere usando caracteres de faixa e caracteres delimitadores.
Caractere da Faixa	Caractere	Altera temporariamente o símbolo que significa uma faixa de caractere. A faixa de caractere reverterá para o caractere padrão quando você fechar o conjunto de caracteres.
Caractere Delimitador	Caractere	Altera temporariamente o símbolo que separa as faixas de caractere. O caractere delimitador reverterá para o caractere padrão quando você fechar o conjunto de caracteres.

Regras e Diretrizes dos Modelos Probabilísticos e do Classificador

Cada modelo probabilístico e do classificador no repositório do Modelo identifica um arquivo na estrutura de diretórios da Informatica. Os arquivos contêm os valores de dados e os rótulos que você adiciona ao modelo na ferramenta Developer. A lógica de compilação que define as associações entre os valores de dados e os rótulos também estão contidas nos arquivos.

Considere as seguintes regras e diretrizes ao trabalhar com modelos probabilísticos ou do classificador:

- Quando você executa um mapeamento que inclui um modelo, o Serviço de Integração de Dados aplica a lógica de modelo compilado aos dados de entrada da transformação. O Serviço de Integração de Dados não lê os valores de dados ou os rótulos no modelo quando o mapeamento é executado.

- Opcionalmente, você pode remover os valores de dados e os rótulos de um modelo probabilístico ou do classificador. Por exemplo, você pode decidir remover os dados confidenciais ou patenteados de um modelo. Você pode remover os valores de dados individuais e rótulos na ferramenta Developer. Você pode remover todos os valores de dados e os rótulos quando exporta um modelo do repositório do Modelo.

Nota: Se você remover todos os valores de dados e os rótulos de um modelo, não poderá compilar o modelo.

- Quando você remove um ou mais valores de dados ou rótulos de um modelo, a lógica de modelo compilado não representa mais os dados atuais no arquivo de modelo. Para sincronizar a lógica de modelo, dos valores de dados e dos rótulos, compile o modelo novamente. Não compile o modelo se você deseja manter a lógica de modelo atual.
- Para proteger os dados em um modelo do classificador ou probabilístico, faça backup do arquivo de modelo na estrutura de diretórios da Informatica. Faça backup do arquivo antes de remover todos os valores de dados e os rótulos de um modelo.
- Localize os arquivos de modelo na máquina host do Serviço do Gerenciamento de Conteúdo.

Os arquivos de modelo probabilístico têm a seguinte localização padrão e a extensão de nome de arquivo:

```
<Diretório_Instalação_Informatica>/tomcat/bin/ner/<nome_de_arquivo>.ner
```

Os arquivos do modelo do classificador têm a seguinte localização padrão e extensão de nome de arquivo:

```
<Diretório_Instalação_Informatica>/tomcat/bin/classifier/<nome_de_arquivo>.classifier
```

- Se você atualizar a instalação da Informatica, talvez seja necessário compilar os modelos probabilístico e do classificador antes de usar os modelos em um mapeamento. Se um modelo não contiver dados, substitua o arquivo atual na estrutura de diretórios da Informatica com o arquivo de backup que contém os dados.

Gerenciando Rótulos em Modelos do Classificador e Modelos Probabilísticos

Para revisar e atualizar os rótulos em um modelo probabilístico ou um modelo do classificador, use a caixa de diálogo **Gerenciar Rótulos**.

1. Abra o conjunto de conteúdo que contém o modelo do classificador.
2. Selecione o nome do modelo e clique em **Editar**.
3. Abra a caixa de diálogo **Gerenciar Rótulos**.
A caixa de diálogo lista os rótulos no modelo.

TÓPICOS RELACIONADOS:

- [“Gerenciamento do Rótulo do Modelo do Classificador” na página 48](#)
- [“Configuração de Modelo Probabilístico” na página 64](#)

Criar um Conjunto de Conteúdo

Crie um conjunto de conteúdo para gerenciar objetos de dados de referência que fazem referência a um único projeto, tipo de informações ou propósito de negócios.

1. Na exibição **Object Explorer**, selecione um projeto ou pasta para armazenar o conjunto de conteúdo.
2. Clique em **Arquivo > Novo > Conjunto de Conteúdo**.
3. Insira um nome para o conjunto de conteúdo.

4. Opcionalmente, selecione **Procurar** para alterar a localização do repositório do Modelo do conjunto de conteúdo.
5. Clique em **Concluir**.

Criando um Objeto de Dados de Referência em um Conjunto de Conteúdo

Você pode criar um conjunto de caracteres, um conjunto de padrões, um conjunto de tokens, uma expressão regular, um modelo probabilístico e um modelo de classificador em um conjunto de conteúdo.

1. Abra um conjunto de conteúdo no editor e selecione a exibição **Conteúdo**.
2. Selecione um tipo de objeto de dados de referência.
3. Clique em **Adicionar**.
4. Insira um nome para o objeto de dados de referência.
Opcionalmente, insira uma descrição do objeto.
5. Configure as propriedades do objeto de dados de referência.
6. Clique em **Concluir**.

Sugestão: Você pode copiar objetos de dados de referência de um conjunto de conteúdo para outro. Use as opções **Copiar para** e **Colar de** para criar uma cópia de um objeto em um conjunto de conteúdo. Use a tecla **CTRL** para selecionar vários objetos de conjunto de conteúdo.

CAPÍTULO 4

Modelos do Classificador

Este capítulo inclui os seguintes tópicos:

- [Visão Geral dos Modelos do Classificador, 43](#)
- [Estrutura do Modelo do Classificador, 44](#)
- [Pontuações do Classificador, 44](#)
- [Exemplo de Transformação do Classificador, 44](#)
- [Opções do Modelo do Classificador, 45](#)
- [Dados de Referência do Modelo do Classificador, 47](#)
- [Dados do Rótulo do Modelo do Classificador, 47](#)
- [Configuração do Modelo do Classificador, 49](#)
- [Operações Filtrar e Localizar, 53](#)
- [Operações de Copiar e Colar, 54](#)

Visão Geral dos Modelos do Classificador

Um modelo de classificador é um objeto de dados de referência em um conjunto de conteúdo. Use um modelo de classificador para analisar strings de texto longas que contenham vários valores. Um modelo do classificador identifica o tipo mais comum de informações de cada string.

Adicione um modelo do classificador para fazer uma transformação do Classificador. A transformação procura valores comuns entre o modelo de dados do classificador e os dados em cada linha de entrada. A transformação usa os valores comuns para categorizar o tipo de informações que cada linha representa.

Use um modelo do classificador quando os dados de entrada tiverem as seguintes características:

- Os dados de entrada contêm texto. Os modelos do classificador aplicam processos de idioma natural para dados de texto para identificar os tipos de informações no texto. O processo de idioma natural detecta palavras relevantes na string de entrada. Os processos de idioma natural desconsideram as palavras que não estão relacionadas.
- As strings de dados de entrada contêm vários valores. Por exemplo, você pode criar uma coluna de dados que contenha o conteúdo de uma mensagem de e-mail em cada campo.

A transformação do classificador lê os tipos de dados da string. A transformação não impõe limite ao tamanho das strings de entrada.

Os modelos do classificador são compilados na ferramenta Developer. Ao compilar um modelo, você cria associações entre os valores de dados similares no modelo. A transformação do classificador usa os dados compilados para pesquisar informações nos dados de entrada.

Estrutura do Modelo do Classificador

Um modelo do classificador contém valores de dados de referência e valores de rótulo. Os valores de dados de referência representam os dados que você deseja classificar. Os valores de rótulo especificam os tipos de informações que uma transformação do Classificador pode identificar nos dados.

Um modelo do classificador também contém dados de compilação. A transformação do Classificador usa os dados de compilação para medir as similaridades entre os dados de referência no modelo e os dados de entrada da transformação. Ao compilar um modelo do classificador, você cria ou atualiza os dados de compilação. Quando uma transformação do Classificador compara os dados de entrada com os dados de modelo, ela retorna os valores de rótulo que descrevem cada linha dos dados de entrada.

A Developer tool grava os valores de dados de referência, os valores de rótulo e os dados de compilação em um arquivo na estrutura de diretórios da Informatica. O objeto do modelo do classificador no repositório do Modelo armazena o nome do arquivo. Quando você salva um modelo do classificador, grava os valores de dados de referência e os valores de rótulo no arquivo. Ao compilar o modelo, você atualiza os dados de compilação no arquivo. Você pode ler o nome de arquivo nas propriedades do modelo na Developer tool.

Pontuações do Classificador

Uma transformação do classificador compara cada linha dos dados de entrada com cada linha dos dados de referência em um modelo do classificador. A transformação calcula uma pontuação para cada comparação. As pontuações representam os graus de similaridade entre a linha de entrada e as linhas dos dados de referência.

Quando executar um mapeamento que contém uma transformação do classificador, o mapeamento retorna o rótulo que identifica a linha dos dados de referência com a pontuação mais alta. A faixa de pontuação é de 0 até 1. Uma pontuação alta indica uma forte correspondência entre os dados de entrada e os dados do modelo.

Consulte as pontuações do classificador para verificar se a saída do rótulo descreve de forma precisa cada linha dos dados de entrada. Ou, analise as pontuações para verificar se o modelo do classificador é adequado para os dados de entrada. Se a saída de transformação contiver um grande percentual de pontuações baixas, o modelo do classificador pode ser inadequado. Para melhorar as comparações, compile o modelo novamente. Se o modelo compilado não melhorar as pontuações, substitua o modelo na transformação.

Exemplo de Transformação do Classificador

É possível um modelo do classificador e uma transformação do classificador para categorizar mensagens de e-mail com base no texto que elas contêm.

Por exemplo, você tem um administrador de dados no centro de suporte ao cliente de um fabricante de software. Você revisa as mensagens de e-mail que o centro de suporte recebe dos clientes. A organização tem clientes em vários países e o centro de suporte recebe e-mails em vários idiomas. Você deseja classificar os e-mails por idioma para que possa enviar cada e-mail para o departamento que pode oferecer a melhor resposta ao cliente.

Para classificar os e-mails, realize as seguintes etapas:

1. Grave as mensagens de e-mail em um único arquivo ou tabela de banco de dados.
2. Crie um objeto de dados no repositório do Modelo que lê o arquivo ou a tabela de banco de dados.
3. Crie objetos de dados no repositório do Modelo para cada idioma que uma mensagem usa.
4. Crie um modelo do classificador que contenha um texto de amostra para cada idioma.

Nota: Você pode usar dados de amostra dos dados das mensagens de e-mail como dados de origem para o modelo.

5. Adicione o modelo do classificador a uma transformação do Classificador reutilizável.
6. Configure um mapeamento para aplicar a transformação do Classificador aos dados da mensagem.

Para configurar o mapeamento, realize as seguintes etapas:

- Adicionar a transformação do Classificador e os objetos de dados ao mapeamento.
- Conecte uma porta de entrada da transformação do Classificador ao objeto de dados de origem.
- Conecte as portas de saída da transformação do Classificador aos objetos de dados de destino.

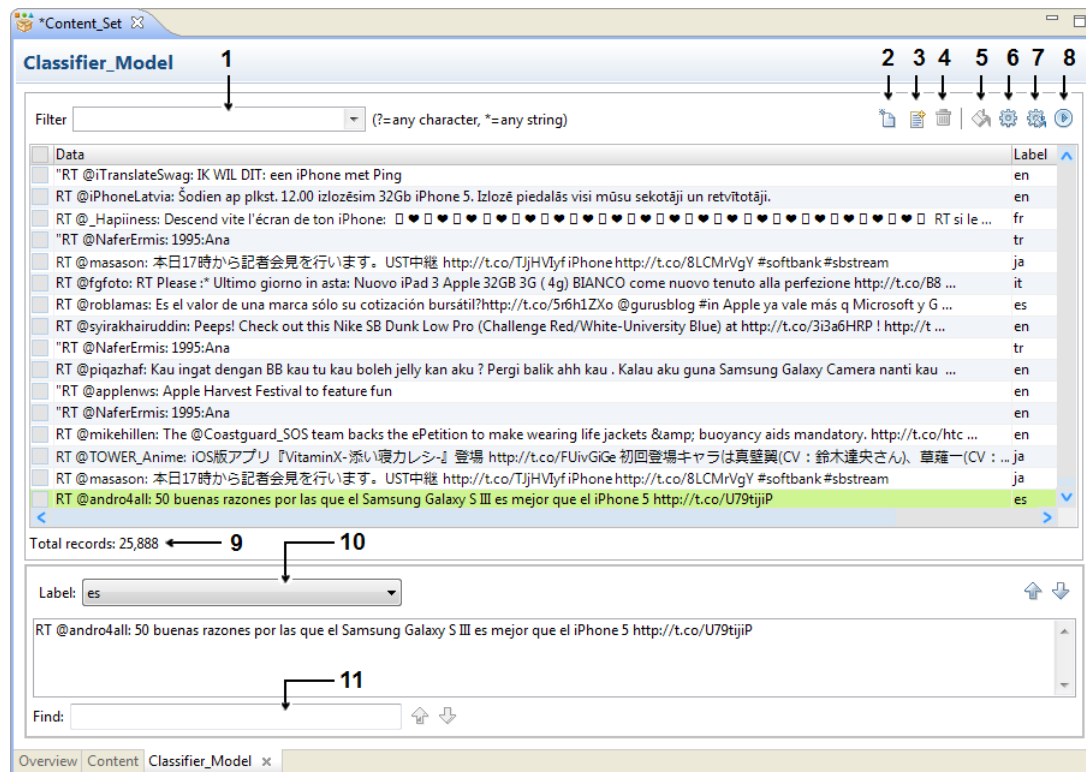
Quando você executa o mapeamento, a transformação do classificador analisa as mensagens de e-mail e grava o texto do e-mail para o destino de dados correto. Você pode compartilhar os destinos de dados com os membros da equipe em cada departamento.

Opções do Modelo do Classificador

A Developer tool exibe os dados do modelo do classificador em um editor com um painel superior e um painel inferior. O painel superior exibe cada linha dos dados de referência e todos os rótulos que você atribuir aos dados. O painel inferior exibe o conteúdo de uma linha que você selecionar.

Use o painel superior para analisar as linhas de dados de referência e identificar uma linha que não usa um rótulo. Use o painel inferior para revisar o conteúdo de uma linha e atribuir um rótulo à linha. O painel superior exibe aproximadamente 100 caracteres de dados em cada linha. O painel inferior exibe todos os dados da linha que você selecionar.

A seguinte imagem mostra o editor de modelo do classificador:



O editor inclui as seguintes opções:

1. Campo Filtrar
Filtra a lista de linhas de dados de referência com base no valor de dados ou no rótulo que você especificar.
2. Adicionar Linha
Insere uma linha de dados de referência em branco.
3. Anexar dados
Importa os dados de um objeto de dados no repositório do Modelo.
4. Excluir
Exclui as linhas de dados de referência que você selecionar. Use as caixas de seleção para selecionar as linhas.
5. Atribuir Rótulo
Atribui um rótulo a uma ou mais linhas de dados de referência que você selecionar. Use as caixas de seleção para selecionar as linhas.
6. Editar Propriedades
Exibe as propriedades do modelo do classificador.
7. Gerenciar Rótulos
Abre a caixa de diálogo **Gerenciar Rótulos**. Use a caixa de diálogo para adicionar ou excluir valores de rótulo do modelo do classificador.
8. Compilar
Compila o modelo de classificador.

9. Total de registros
Indica o número de linhas de dados de referência do modelo do classificador.
10. Campo Rótulo
Exibe um valor de rótulo que você pode aplicar à linha de dados de referência atual.
11. Campo Localizar
Localiza um valor de dados que você especificar na linha de dados de referência atual.

Dados de Referência do Modelo do Classificador

Um modelo do classificador contém uma coluna de dados de referência que podem incluir frases, parágrafos ou páginas de texto. Os dados de referência representam os tipos diferentes de entrada de texto que uma transformação de classificador pode ler em um mapeamento. Ao criar um modelo, verifique se os dados de referência incluem os tipos de texto que você pretende localizar quando executar o mapeamento.

Você pode usar o mapeamento de dados de origem para criar um modelo do classificador. Selecione uma amostra de dados de origem e copie os dados de amostra para o modelo.

Considere as seguintes regras e diretrizes ao trabalhar com dados de referência do modelo do classificador:

- Um campo de dados de referência pode ter qualquer tamanho. Páginas de texto podem ser inseridas em cada campo de dados.
- Os dados de referência são importados de um objeto de dados.
- Não é possível editar os valores de dados de referência. No entanto, você pode excluir uma linha de dados.
- Quando você compila um modelo de classificador, o processo de compilação desconsidera quaisquer valores numéricos nos dados de referência.

Dados do Rótulo do Modelo do Classificador

Um modelo do classificador contém um ou mais rótulos descritivos que resumem os tipos de informações nas linhas de dados de referência. Atribua um rótulo a cada linha de dados de referência.

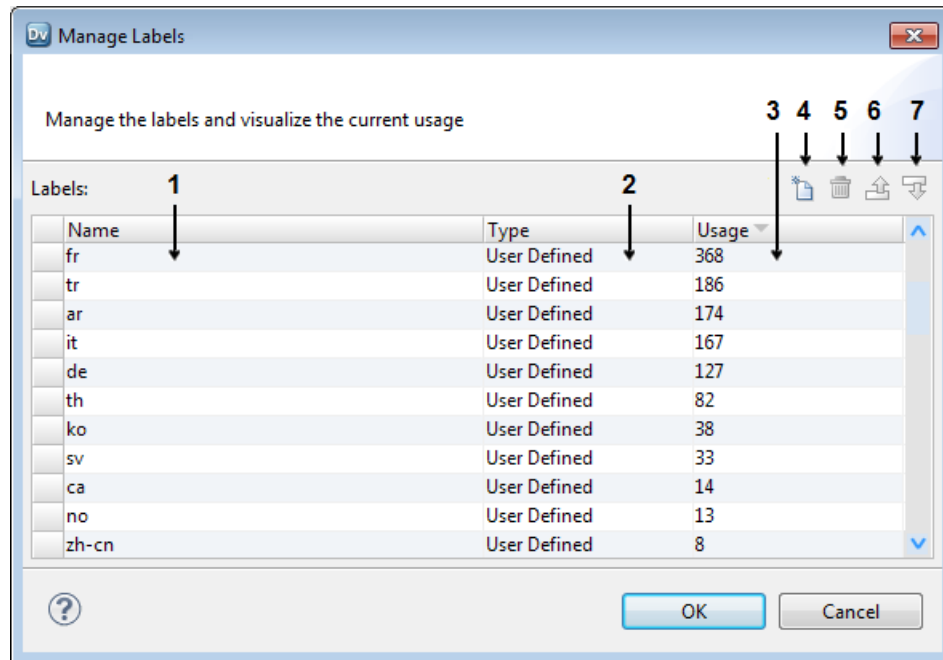
Quando você adiciona dados de uma fonte de dados a um modelo do classificador, pode especificar uma coluna como uma coluna de dados de rótulo. Você também pode criar rótulos no modelo.

Os rótulos são independentes dos valores dos dados de referência que eles descrevem. Se você excluir as linhas de dados de referência que usam um rótulo, não exclua o rótulo do modelo. Se você excluir um rótulo, os valores dos dados de referência associados ao rótulo não são excluídos.

Gerenciamento do Rótulo do Modelo do Classificador

Use a caixa de diálogo **Gerenciar Rótulos** para revisar e atualizar os valores de rótulo em um modelo do classificador. Você também pode classificar e atualizar os valores de rótulo.

A seguinte imagem mostra a caixa de diálogo **Gerenciar Rótulos**:



A caixa de diálogo **Gerenciar Rótulos** contém os seguintes elementos:

1. Coluna Nome.
Contém os valores de rótulo que a transformação do Classificador pode aplicar às linhas de dados de entrada. Você pode classificar os rótulos por nome.
2. Coluna Tipo.
Identifica a fonte dos valores de rótulo. O modelo do classificador identifica todos os rótulos como valores definidos pelo usuário.
3. Coluna Uso.
Indica o número de linhas de dados de referência que usam cada rótulo. Você pode classificar os rótulos pelo número de linhas.
4. Botão Adicionar.
Adiciona um rótulo ao modelo do classificador. Insira um valor de rótulo na coluna Nome na linha.
Nota: Para atualizar um valor de rótulo, clique duas vezes nele e insira o valor que você precisa.
5. Botão Excluir.
Exclui um rótulo do modelo do classificador.
6. Seta para cima.
Move o rótulo uma única linha para cima na caixa de diálogo.
7. Seta para baixo.
Move o rótulo uma única linha para baixo na caixa de diálogo.

Configuração do Modelo do Classificador

As etapas para configurar um modelo do classificador começam com os dados que você deseja classificar. O conteúdo dos dados de referência que você adicionar ao modelo deve refletir os dados que você conecta à transformação do Classificador. A transformação compara os valores e os padrões de dados dos dados de entrada com os valores e os padrões de dados no modelo do classificador.

Para criar um modelo do classificador que você possa usar em uma transformação do Classificador, realize as seguintes tarefas:

1. Identifique os valores de dados de referência e de rótulo a serem adicionados ao modelo.
Você pode usar um fragmento dos dados que deseja classificar. Crie um objeto de dados no repositório do Modelo que lê o fragmento de dados.
2. Crie um conjunto de conteúdo e adicione um modelo do classificador ao conteúdo.
3. Adicione os valores de dados de referência ao modelo.
4. Adicione os valores de rótulo ao modelo.
Você pode importar os dados do objeto de dados no repositório do Modelo. Você também pode inserir uma única linha de dados de referência ou um único rótulo.
5. Atribua um rótulo a cada linha de dados de referência.
Você pode atribuir um rótulo a várias linhas em uma única operação.
6. Compile o modelo.

Depois de compilar o modelo do classificador, você poderá usar o modelo em uma transformação do Classificador.

Criando um Modelo do Classificador

Use um objeto de dados como a fonte dos dados do modelo do classificador.

Um modelo do classificador funciona de modo mais eficiente quando você usa os dados de entrada para a transformação do Classificador como a fonte dos dados de referência do modelo.

1. No Object Explorer, abra ou crie um conjunto de conteúdo.
2. Selecione a exibição de **Conteúdo**.
3. Selecione **Modelos do Classificador** e clique em **Adicionar**.
O Assistente de Modelo do Classificador é aberto.
4. Insira um nome para o modelo do classificador.
Opcionalmente, insira uma descrição de texto para o modelo.
5. Procure o repositório do Modelo e selecione o objeto de dados que contém os dados a serem importados.
Não selecione um objeto de dados de mídia social.
Clique em **Avançar**.

6. Revise as colunas no objeto de dados e selecione uma ou mais colunas a serem adicionadas ao modelo. Você pode adicionar colunas de dados de referência e uma coluna de rótulo na mesma operação.

- Para importar uma coluna de dados como dados de referência, selecione o nome de coluna e clique em **Dados**.

Várias colunas de dados podem ser selecionadas. A Developer tool mescla o conteúdo das colunas que você selecionar em uma única coluna.

- Para importar uma coluna de dados como valores de rótulo, selecione o nome da coluna e clique em **Rótulo**.

Quando você importa dados de referência e valores de rótulo, a Developer tool atribui o rótulo em cada linha à cadeia de dados de referência na mesma linha. Você pode visualizar os dados antes de selecionar as colunas. Você poderá alterar as atribuições de rótulo depois de criar o modelo.

Clique em **Avançar**.

7. Selecione o número de linhas a serem copiadas da fonte de dados.

Por padrão, a Developer tool importará todas as linhas da fonte de dados. Se você inserir um número, o modelo conta as linhas a partir do início do conjunto de dados.

8. Clique em **Concluir** e salve o modelo.

Depois de criar o modelo, verifique as atribuições de rótulo e compile o modelo.

Anexando os Dados de uma Fonte de Dados a um Modelo do Classificador

Você pode importar várias linhas de valores de dados de referência ou valores de rótulo para um modelo do classificador em uma única operação.

1. Abra o conjunto de conteúdo que contém o modelo do classificador.

2. Selecione o nome do modelo e clique em **Editar**.

3. Clique em **Anexar Dados**.

O Assistente de Modelo do Classificador é aberto.

4. Procure o repositório do Modelo e selecione o objeto de dados que contém os dados a serem importados.

Não selecione um objeto de dados de mídia social.

Clique em **Avançar**.

5. Revise as colunas no objeto de dados e selecione uma ou mais colunas a serem adicionadas ao modelo. Você pode adicionar colunas de dados de referência e uma coluna de rótulo na mesma operação.

- Para importar uma coluna de dados como dados de referência, selecione o nome de coluna e clique em **Dados**.

Várias colunas de dados podem ser selecionadas. A Developer tool mescla o conteúdo das colunas que você selecionar em uma única coluna.

- Para importar uma coluna de dados como valores de rótulo, selecione o nome da coluna e clique em **Rótulo**.

Quando você importa dados de referência e valores de rótulo, a Developer tool atribui o rótulo em cada linha à cadeia de dados de referência na mesma linha. Você pode visualizar os dados antes de selecionar as colunas. Você poderá alterar as atribuições de rótulo depois de criar o modelo.

Clique em **Avançar**.

6. Selecione o número de linhas a serem copiadas da fonte de dados.

Por padrão, a Developer tool importará todas as linhas da fonte de dados. Se você inserir um número, o modelo conta as linhas a partir do início do conjunto de dados.

7. Clique em **Concluir** e salve o modelo.

Adicionando uma Linha de Dados de Referência a um Modelo do Classificador

Você pode adicionar uma única linha de dados de referência a um modelo do classificador.

1. Abra o conjunto de conteúdo que contém o modelo do classificador.
2. Selecione o nome do modelo e clique em **Editar**.
3. Clique em **Adicionar Linha**.
A Developer tool adiciona uma linha abaixo da linha atual nos dados de referência.
4. Insira os valores de dados de referência na linha.
Você pode usar os atalhos do Windows para colar dados na linha.

Adicionar um Rótulo a um Modelo do Classificador

Você pode adicionar um único rótulo a um modelo probabilístico.

1. Abra o conjunto de conteúdo que contém o modelo do classificador.
2. Selecione o nome do modelo e clique em **Editar**.
3. Abra a caixa de diálogo **Gerenciar Rótulos**.
A caixa de diálogo lista os rótulos no modelo.
4. Clique em **Novo**.
A Developer tool adiciona uma linha na parte inferior da lista de rótulos.
5. Clique duas vezes no valor padrão na coluna Nome e insira um nome de rótulo.
6. Clique em **OK**.

Depois de criar o rótulo, você poderá atribuí-lo a uma ou mais linhas dos dados de referência. A coluna **Uso** na caixa de diálogo **Gerenciar Rótulos** indica o número de linhas que usam o rótulo.

Atribuindo um Rótulo às Linhas de Dados de Referência

Você pode atribuir um rótulo a uma ou mais linhas de dados de referência em uma única operação.

1. Abra o conjunto de conteúdo que contém o modelo.
2. Selecione o nome do modelo e clique em **Editar**.
3. Selecione uma ou mais linhas de dados de referência. Use as opções da caixa de seleção para selecionar as linhas.
Nota: Você pode usar a opção de filtro para mostrar todas as linhas que contêm um valor de dados que você especificar. Use a opção da caixa de seleção **Selecionar Tudo** para selecionar todas as linhas que contêm o valor.
4. Clique em **Atribuir Rótulo**.
A Developer tool exibe a lista de rótulos no modelo do classificador.
5. Selecione um valor de rótulo e clique em **Atribuir**.
A Developer tool atualiza as linhas de dados de referência que você selecionou com o valor de rótulo.

Compile o modelo para adicionar os nomes de rótulo à lógica do modelo do classificador.

Identificando Valores de Rótulo Não Utilizados

Use a caixa de diálogo **Gerenciar Rótulos** para localizar todos os valores de rótulo que permanecerem não utilizados no modelo do classificador. A caixa de diálogo **Gerenciar Rótulos** exibe os dados de uso dos valores de rótulo no modelo do classificador. Utilize os dados de uso para verificar o número de linhas de dados de referência que usam um valor de rótulo e para localizar os valores de rótulo não utilizados.

1. Abra o conjunto de conteúdo que contém o modelo do classificador.
2. Selecione o nome do modelo e clique em **Editar**.
3. Abra a caixa de diálogo **Gerenciar Rótulos**. A caixa de diálogo lista os rótulos no modelo do classificador.
4. Revise os dados da coluna Uso de cada rótulo.

A coluna Uso lista o número de linhas de dados de referência que usam o rótulo. Se um valor de rótulo não for usado, a coluna Uso terá um valor de zero.

Excluindo Linhas de um Modelo do Classificador

Você pode excluir uma ou mais linhas de dados de referência de um modelo do classificador em uma única ação.

1. Abra o conjunto de conteúdo que contém o modelo.
2. Selecione o nome do modelo e clique em **Editar**.
3. Selecione uma ou mais linhas de dados de referência. Use as opções da caixa de seleção para selecionar as linhas.
4. Clique em **Excluir**.

A Developer tool remove as linhas que você selecionou do modelo do classificador.

Para desfazer a operação, pressione as teclas Ctrl + Z no teclado.

Excluir um Rótulo de um Modelo do Classificador

Use a caixa de diálogo **Gerenciar Rótulos** para excluir um rótulo de um modelo do classificador.

1. Abra o conjunto de conteúdo que contém o modelo.
2. Selecione o nome do modelo e clique em **Editar**.
3. Abra a caixa de diálogo **Gerenciar Rótulos**.
4. Clique em **Excluir**.
5. Clique em **Sim** para confirmar a operação.

A Developer tool exclui o rótulo do modelo. A Developer tool não exclui nenhuma linha de dados de referência que usa o rótulo.

6. Clique em **OK** para fechar a caixa de diálogo.

Para desfazer a operação, pressione as teclas Ctrl + Z no teclado.

Compilar um Modelo do Classificador

Cada vez que você editar um rótulo ou valores dos dados de referência em um modelo do classificador, é necessário compilar o modelo. Ao compilar o modelo, você atualiza a compilação de dados no modelo.

- Para atualizar a compilação de dados, abra o modelo na ferramenta Developer e clique em **Compilar**.

Operações Filtrar e Localizar

Use as opções de filtro para mostrar ou ocultar as linhas de dados de referência que atendem aos critérios que você especificar. Quando você aplica um filtro, pode realizar mais ações nas linhas de dados que o modelo do classificador exibe. Por exemplo, você pode aplicar um valor de rótulo a todas as linhas de dados.

Use a opção de filtro para realizar as seguintes tarefas:

- Localizar as linhas de dados de referência que contêm um valor que você inserir.
- Localizar as linhas de dados de referência que usam um rótulo que você selecionar.
- Localizar as linhas de dados de referência que não usam um rótulo.

Você também pode pesquisar por um valor de dados em uma linha dos dados de referência.

Usando um Valor de Dados para Filtrar as Linhas de Dados de Referência

Use o filtro para verificar se uma ou mais linhas de dados de referência contêm os valores de dados esperados.

1. Abra o conjunto de conteúdo que contém o modelo do classificador.
2. Selecione o nome do modelo e clique em **Editar**.
3. Insira um valor no campo Filtro.

Você pode incluir caracteres curinga no valor que inserir.

A Developer tool exibe as linhas de dados que contêm o texto do filtro.

Usando um Valor de Rótulo para Filtrar as Linhas de Dados de Referência

Use o filtro para mostrar ou ocultar as linhas de dados de referência que usam um rótulo que você selecionar.

1. Abra o conjunto de conteúdo que contém o modelo.
2. Selecione o nome do modelo e clique em **Editar**.
3. Selecione um valor de rótulo no menu Filtro.

A Developer tool exibe as linhas de dados de referência que usam o valor de rótulo.

Nota: Para localizar qualquer linha de dados de referência que não usa um rótulo, selecione a opção **Nenhum Rótulo** no menu Filtro.

Localizando um Valor em uma Linha de Dados de Referência

Use o campo Localizar para pesquisar um valor de dados em uma linha que você selecionar.

1. Abra o conjunto de conteúdo que contém o modelo.
2. Selecione o nome do modelo e clique em **Editar**.
3. Selecione uma linha de dados de referência.
4. Insira um valor no campo Localizar.
O modelo realça as instâncias do valor na linha de dados de referência.
5. Use a seta para cima ou para baixo para encontrar mais instâncias do valor na linha.

Operações de Copiar e Colar

O modelo do classificador pode ser copiado de um conjunto de conteúdo para outro em um repositório do Modelo. Copie um modelo do classificador para compartilhar recursos com outros usuários da ferramenta Developer.

Você pode copiar um modelo para outro conjunto de conteúdo, ou pode importar um modelo para o conjunto de conteúdo atual. Também pode importar vários modelos de vários conjuntos de conteúdo no repositório em uma única operação.

Quando você copia um modelo, o Serviço de Gerenciamento de Conteúdo criará uma cópia do arquivo de dados do modelo na máquina do serviço. Cada modelo usa um arquivo de dados de diferentes.

Copiar um Modelo do Classificador para Outro Conjunto de Conteúdo

O modelo do classificador pode ser copiado de um conjunto de conteúdo para outro em um repositório do Modelo. Quando você copia um modelo do classificador, o objeto do modelo e os conjuntos de conteúdo de origem e destino são especificados.

1. Abra o conjunto de conteúdo que contém o modelo do classificador.
2. Selecione um modelo do classificador e clique em **Copiar para**.
3. Procure o repositório do Modelo e selecione um conjunto de conteúdo.
É possível copiar o modelo do classificador para um conjunto de conteúdo no projeto atual ou em outro projeto.
4. Clique em **OK**.
A ferramenta Developer copia o modelo do classificador para o conjunto de conteúdo selecionado.

Importar um Modelo do Classificador de Outro Conjunto de Conteúdo

O modelo do classificador pode ser importado de um conjunto de conteúdo para outro em um repositório do Modelo. Quando você copia um modelo do classificador, um ou mais objetos do modelo e os conjuntos de conteúdo de origem e destino são especificados.

1. Abra o conjunto de conteúdo para incluir o modelo classificador.
2. Selecione um modelo do classificador e clique em **Colar de**.

3. Procure o repositório do Modelo e selecione um modelo do classificador.
É possível colar o modelo do classificador de um conjunto de conteúdo no projeto atual ou de outro projeto.
4. Clique em **OK**.
A ferramenta Developer cola o modelo do classificador no conjunto de conteúdo atual.

CAPÍTULO 5

Modelos Probabilísticos

Este capítulo inclui os seguintes tópicos:

- [Visão Geral de Modelos Probabilísticos, 56](#)
- [Estrutura de Modelo Probabilístico, 57](#)
- [Exemplo de Transformação do Rotulador, 58](#)
- [Exemplo de Transformação do Analisador, 59](#)
- [Opções do Modelo Probabilístico, 59](#)
- [Dados de Referência do Modelo Probabilístico, 62](#)
- [Dados de Rótulo do Modelo Probabilístico, 62](#)
- [Propriedades do Modelo Probabilístico, 64](#)
- [Configuração de Modelo Probabilístico, 64](#)
- [Operações de Copiar e Colar, 71](#)

Visão Geral de Modelos Probabilísticos

Um modelo probabilístico é um objeto de dados de referência que você cria em um conjunto de conteúdo. Use um modelo probabilístico para analisar uma string de dados que contém vários valores de dados. Um modelo probabilístico identifica o tipo de informações em cada valor na string. Você pode adicionar um modelo probabilístico a uma transformação de Rotulador e a uma transformação de Analisador.

Use um modelo probabilístico em uma transformação de Rotulador para aplicar um rótulo descritivo a cada valor em uma cadeia de entrada. A transformação de Rotulador grava os rótulos em uma única porta de saída. Use um modelo probabilístico em uma transformação de Analisador para gravar cada valor de uma string de entrada em uma porta que representa as informações no valor. A transformação de Analisador cria uma porta de saída para cada tipo de informações.

É possível projetar e compilar um modelo probabilístico na Developer tool. Ao definir um modelo probabilístico, você adiciona uma série de linhas de dados ao modelo e atribui um rótulo a cada valor em cada linha. Quando você compila um modelo probabilístico, a Developer tool cria associações entre os valores de dados e os rótulos que foram adicionados. A transformação de Rotulador e a transformação de Analisador usam processos de idioma natural para comparar os dados do modelo probabilístico com os dados da porta de entrada.

Os processos de idioma natural usam as seguintes técnicas para identificar os tipos de informações em valores de dados:

- Os processos de idioma natural podem reconhecer valores de dados semelhantes e aplicar o mesmo rótulo aos valores.
- Os processos de idioma natural podem comparar um valor de dados com os valores adjacentes na string. Os processos de idioma natural analisam a sequência de valores para compreender o uso de cada string e verificar os tipos de informações que essas strings representam.

Estrutura de Modelo Probabilístico

Um modelo probabilístico contém os valores de dados de referência e de rótulo. Os valores de dados de referência representam os diferentes valores que podem ser exibidos nos dados de entrada da transformação. Os valores de rótulo identificam os tipos de informações que você espera que os dados de entrada contenham.

Um modelo probabilístico também contém dados de compilação. A transformação do Rotulador e a transformação do Analisador usam os dados de compilação para medir as similaridades entre os dados de referência no modelo e os dados de entrada da transformação. Ao compilar um modelo probabilístico, você cria ou atualiza os dados de compilação.

Uma linha de dados pode conter um único valor ou vários valores. Cada linha de entrada pode ter uma estrutura diferente. Você pode atribuir o mesmo rótulo a diferentes valores em uma linha de dados. Como alternativa, você pode atribuir um rótulo diferente a valores idênticos que são exibidos em diferentes posições em uma linha. O Serviço de Integração de Dados considera as posições relativas dos valores na cadeia de entrada quando o mapeamento é executado. Atribua cada rótulo a pelo menos um valor de dados antes de compilar o modelo probabilístico.

A Developer tool grava os valores de dados de referência, os valores de rótulo e os dados de compilação em um arquivo na estrutura de diretórios da Informatica. O objeto de modelo probabilístico no repositório do Modelo armazena o nome do arquivo. Ao salvar um modelo probabilístico, você grava os valores de dados de referência e os valores de rótulo no arquivo. Ao compilar o modelo, você atualiza os dados de compilação no arquivo. Você pode ler o nome de arquivo nas propriedades do modelo na Developer tool.

Nota: Para otimizar os recursos do modelo probabilístico, verifique se cada linha de dados contém vários valores de dados de referência. A ordem dos valores em cada linha deve corresponder o máximo possível à ordem na qual esses valores ocorrem nos dados de entrada da transformação. Se o as linhas de dados contiverem valores únicos de dados de referência, a transformação do Rotulador ou a transformação do Analisador não poderão aplicar os processos de idioma natural durante a análise probabilística.

Exemplo de Transformação do Rotulador

O banco de dados do cliente em uma organização seguradora contém vários erros de entrada de dados. Você é o administrador de dados na organização seguradora. Você configura um mapeamento com uma transformação do Rotulador para determinar os diferentes tipos de dados que cada coluna contém.

A tabela a seguir descreve os dados de amostra do banco de dados de clientes:

ID de linha	Campo 1	Campo 2	Campo 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

Quando você executa o mapeamento, a transformação do Rotulador compara os dados de entrada com os dados de referência do modelo probabilístico. A transformação de Rotulador seleciona um rótulo para os dados em cada porta de entrada. A transformação grava os rótulos em uma porta de saída. Cada linha de saída contém um conjunto de rótulos que define a estrutura de dados na linha de entrada correspondente.

A tabela a seguir descreve os rótulos que a transformação de Rotulador adiciona à porta de saída:

ID de Linha	Rótulos de Saída
1	número de contato da organização
2	número da organização de contato
3	número de contato da organização
4	número da organização de contato
5	número de contato da organização
6	contato do número da organização
7	contato do número da organização
8	número da organização de contato

Exemplo de Transformação do Analisador

Um supermercado armazena as descrições de produto em uma única coluna em uma tabela de banco de dados. As descrições do produto contêm vários valores de dados que representam diferentes tipos de informações. Você é o administrador de dados no supermercado. Você deseja criar colunas para os diferentes tipos de informações nas descrições de produto.

Um mapeamento com uma transformação do Analisador foi configurado para organizar os valores de dados nos campos corretos.

O seguinte fragmento de dados contém a descrição do produto para um suco de laranja:

Suco de Laranja sem Açúcar Sunnydream 350 ml

A tabela a seguir descreve os dados de saída que a transformação do Analisador cria com base nos dados de entrada:

Nome do Produto	Tipo de Produto	Detalhes do Produto	Tamanho do Produto
Sunnydream	Suco de Laranja	Sem Açúcar	350 ml

Opções do Modelo Probabilístico

Ao editar um modelo probabilístico, você pode trabalhar na exibição Dados ou Rótulo.

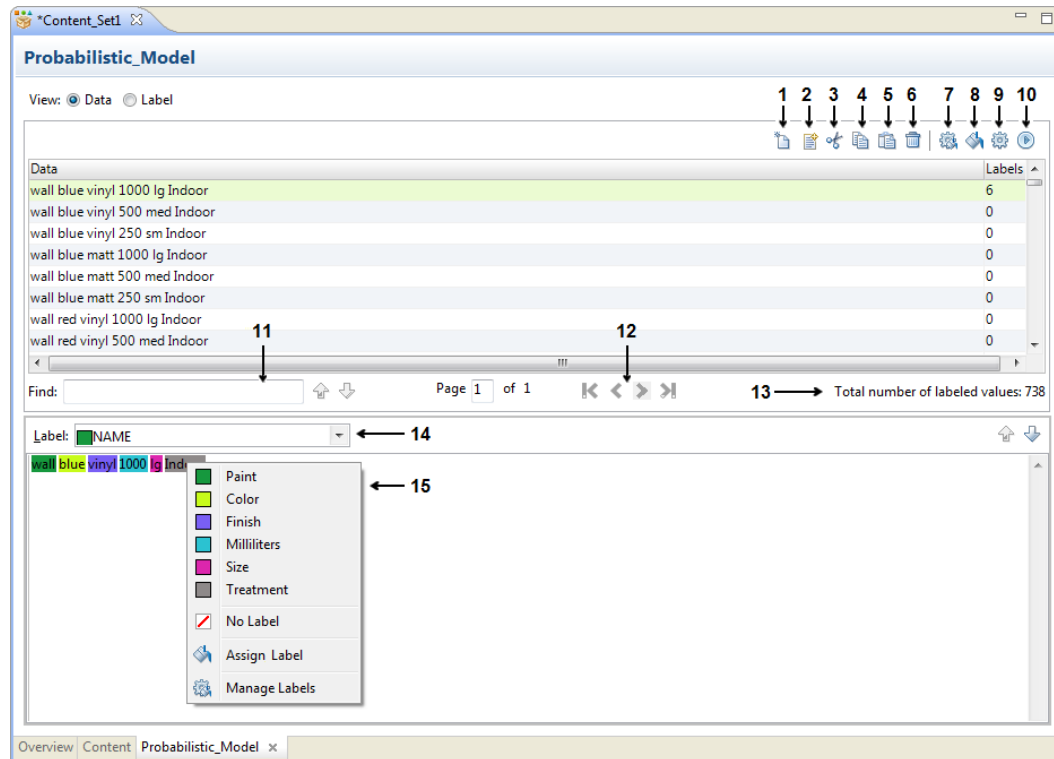
Use a exibição Dados para adicionar linhas de dados ao modelo e para atribuir rótulos aos valores de dados em cada linha. Use a exibição Rótulo para revisar as informações detalhadas sobre o uso dos valores de rótulo no modelo. Você pode adicionar rótulos ao modelo probabilístico nas exibições Dados e Rótulo.

Exibição de Dados do Modelo Probabilístico

A exibição Dados mostra as linhas de dados de referência no modelo probabilístico e o número de valores de rótulo que você atribui a cada linha. A exibição Dados também mostra o número total de rótulos atribuídos aos valores no modelo atual.

Quando você seleciona uma linha de dados de referência, os valores na linha aparecem em um editor abaixo do campo Localizar. Para atribuir um rótulo a um valor de dados de referência em uma linha, clique com o botão direito do mouse no valor no editor e selecione um valor de rótulo.

A seguinte imagem mostra as opções do modelo probabilístico que você pode usar ao selecionar a exibição Dados:



A exibição Dados inclui as seguintes opções:

1. Adicionar Linha
Insere uma linha de dados em branco.
2. Anexar dados.
Importa os dados de um objeto de dados no repositório do Modelo.
3. Recortar
Remove uma linha de dados do modelo probabilístico e a adiciona à área de transferência.
4. Copiar
Copia uma linha de dados para a área de transferência.
5. Colar
Cola uma linha de dados da área de transferência no modelo probabilístico.
6. Excluir
Exclui uma linha de dados do modelo probabilístico.
7. Gerenciar Rótulos
Abre a caixa de diálogo **Gerenciar Rótulos**. Use a caixa de diálogo para adicionar ou excluir os valores de rótulo do modelo probabilístico.
8. Atribuir Rótulo
Atribui um rótulo para um ou mais valores de dados de referência que você selecionar. Você pode usar a opção para atribuir um rótulo a todas as instâncias de um valor de dados de referência no modelo.
9. Editar Propriedades

Exibe as propriedades do modelo probabilístico.

10. Compilar

Compila o modelo probabilístico.

11. Campo Localizar

Localiza as linhas no modelo que contêm o valor de dados de valor que você inseriu. Use a seta Para Cima e Para Baixo para mover até as linhas que contêm o valor.

12. Setas para frente e para trás

Mova para frente e para trás nas linhas de valores de dados no modelo.

13. Número total de valores rotulados

Indica o número de valores de dados de referência que usam um rótulo.

14. Campo Rótulo

Exibe um valor de rótulo que você pode aplicar ao valor de dados de referência que selecionar.

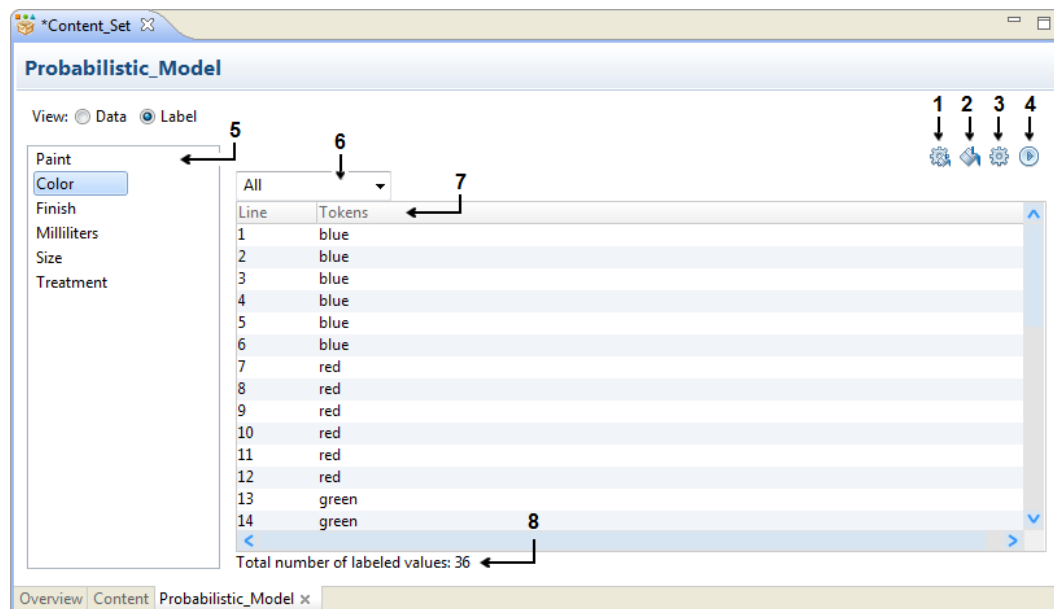
15. Menu Rótulo

Exibe uma lista de opções que você pode usar para atribuir um rótulo a um ou mais valores de dados de referência. Para abrir o menu, clique em um valor de dados de referência no editor de dados de referência.

Exibição de Rótulo do Modelo Probabilístico

A exibição Rótulo lista os rótulos que você define no modelo probabilístico. Quando você seleciona um rótulo, a exibição Rótulo mostra os valores de dados que são atribuídos ao rótulo em cada linha.

A seguinte imagem mostra as opções do modelo probabilístico que você pode usar ao selecionar a exibição Dados:



A exibição Rótulo inclui as seguintes opções:

1. Gerenciar Rótulos

Abre a caixa de diálogo **Gerenciar Rótulos**. Use a caixa de diálogo para adicionar ou excluir os valores de rótulo do modelo probabilístico.

2. Atribuir Rótulo

Atribui um rótulo para um ou mais valores de dados de referência que você selecionar.

Você pode atribuir um rótulo a um único valor de dados ou atribuir um rótulo a vários valores em uma única operação.

3. Editar Propriedades

Exibe as propriedades do modelo probabilístico.

4. Compilar

Compila o modelo probabilístico.

5. Lista de valores de rótulo

Lista os rótulos de que você pode atribuir aos valores de dados de referência no modelo.

6. Filtro de atribuição

Filtra a lista de valores de dados de referência que usam o rótulo que você seleciona. As opções de filtro mostram ou ocultam os valores de dados de referência com base no método usado para atribuir o rótulo aos valores de dados.

Quando você aplica um filtro, o número total de valores rotulados na exibição Rótulo reflete o número de valores que atendem à condição de filtragem.

7. Coluna de valor de dados de referência

Lista os valores de dados de referência que usam o rótulo atual.

8. Número total de valores rotulados

Indica o número de valores de dados de referência que usam o rótulo atual.

Dados de Referência do Modelo Probabilístico

Os valores de dados de referência em um modelo probabilístico representam os tipos de dados de entrada que você pode conectar a uma transformação em um mapeamento.

As linhas de dados de referência podem ser adicionadas, editadas e excluídas na Developer tool. Você pode colar dados da área de transferência e importar dados de uma fonte de dados. Depois de adicionar os valores de dados de referência, atribua um rótulo a cada valor de dados em cada linha.

Dados de Rótulo do Modelo Probabilístico

Os valores de rótulos em um modelo probabilístico representam os tipos de informações que os valores de dados de referência podem conter. Ao adicionar linhas de dados de referência a um modelo, atribua um rótulo a cada valor em cada linha. Os rótulos que você adiciona ao modelo aparecem na exibição Rótulo e nas opções de menu na exibição Dados.

Você pode atribuir qualquer rótulo no modelo para qualquer valor de dados de referência. Se o mesmo valor tiver diferentes significados em diferentes linhas de dados de referência, você poderá atribuir um rótulo diferente a cada valor em cada linha.

O intervalo de valores de rótulo pode corresponder ao intervalo de portas de entrada que a transformação do Rotulador ou do Analisador lê durante a análise probabilística. O modelo probabilístico deve conter pelo

menos um valor de rótulo que a transformação possa aplicar aos valores de dados em cada porta de entrada.

Por exemplo, um depósito pode armazenar dados de inventário em um arquivo delimitado por vírgulas que define oito colunas. Projete um mapeamento que analise os dados do inventário para uma tabela de banco de dados. Crie um modelo probabilístico com um valor de rótulo para cada coluna de dados. Quando você executa o mapeamento, a transformação do Analisador grava cada valor nos dados de entrada na coluna correta na tabela de destino.

A seguinte tabela mostra as colunas dos dados de inventário e dos valores de rótulo que você pode criar em um modelo probabilístico:

Nome da Coluna do Inventário	Nome do Rótulo
Product_Name	Product_Name
Quantidade	Quantidade
Localização	Localização
Código de Barras	Código de Barras
SKU	Stock_Keeping_Unit
Arrival_Date	Arrival_Date
Cost_Price	Cost_Price

Nota: Você pode usar os nomes de coluna de entrada ou outros nomes. Os nomes não precisam coincidir.

Estouro de Rótulo

Quando uma transformação não pode aplicar um rótulo a um valor de dados de entrada, a transformação trata o valor de dados como dados de overflow. A transformação do Rotulador atribui um rótulo de overflow a qualquer valor de dados que não conseguir identificar. A transformação do Analisador grava todos os valores de dados que não conseguir identificar em uma porta de overflow.

A seguinte tabela mostra como uma transformação do Analisador pode usar uma porta de overflow para analisar os elementos de dados de endereço que um modelo probabilístico não reconhece:

Dados de Entrada	Porta Street_Name	Porta Street_Descriptor	Porta de estouro
Park Place	Park	Place	Sem dados de overflow
Park Avenue	Park	Avenue	Sem dados de overflow
Madison Avenue	Madison	Avenue	Sem dados de overflow
Central Park	Central	Park	Sem dados de overflow
Washington Square Park	Washington	Square	Park
Madison Square Garden	Madison	Square	Garden

A transformação do Analisador também grava os valores em uma porta de overflow quando o número de valores de entrada é maior do que o número de rótulos no modelo. Antes de usar um modelo probabilístico em uma transformação, analise os dados de entrada e verifique se o modelo contém o número correto de valores de rótulo.

Propriedades do Modelo Probabilístico

Você pode revisar as propriedades gerais e as propriedades avançadas de um modelo probabilístico.

Para abrir o editor de propriedades, selecione a opção **Editar Propriedades** na exibição Dados ou Rótulo.

As propriedades gerais exibem o nome do modelo probabilístico, todas as descrições do modelo e o nome do arquivo de dados do modelo. As propriedades avançadas exibem as propriedades computacionais que a Developer tool utiliza para compilar o modelo probabilístico.

O elemento básico na compilação de modelos probabilísticos é *n-gram*. Um *n-gram* é uma série de letras que seguem ou precedem outras letras para completar uma palavra. Quando um mapeamento é executado, a transformação de Rotulador ou Analisador cria vários *n-gramas* para cada valor na coluna de dados de referência do modelo probabilístico. A transformação compara os valores de dados de entrada com os valores de dados de referência e *n-grams*. As propriedades avançadas em um modelo probabilístico determinam como o modelo probabilístico trata *n-grams* e outros recursos do modelo.

Nota: Os valores padrão nas propriedades avançadas representam as configurações preferenciais para a análise probabilística e a compilação do modelo probabilístico. Se você editar uma propriedade avançada, a precisão da análise probabilística poderá ser negativamente afetada. Não edite as propriedades avançadas, a menos que você entenda os efeitos das alterações feitas.

TÓPICOS RELACIONADOS:

- [“Regras e Diretrizes dos Modelos Probabilísticos e do Classificador” na página 40](#)

Configuração de Modelo Probabilístico

As etapas para configurar um modelo probabilístico começam com o tipo de análise que você deseja executar. Use um modelo probabilístico em uma transformação do Rotulador para identificar os tipos de informações em cada valor em uma cadeia de entrada. Use um modelo probabilístico em uma transformação do Analisador para analisar os valores de dados em uma cadeia de entrada para diferentes portas de saída.

Você pode usar o mesmo modelo probabilístico para rotular e analisar dados. Quando você usa o modelo em uma transformação do Rotulador, a transformação cria uma única porta de saída para cada porta de entrada que você selecionar. Quando você usa o modelo em uma transformação do Analisador, a transformação cria uma porta de saída para cada tipo de dados de entrada que identificar.

Para criar um modelo probabilístico, realize as seguintes tarefas:

1. Identifique os valores de dados de referência e de rótulo a serem adicionados ao modelo.
Você pode usar um fragmento dos dados que deseja analisar. Crie um objeto de dados no repositório do Modelo que lê o fragmento de dados.
2. Crie um conjunto de conteúdo e adicione um modelo probabilístico ao conteúdo.

3. Adicione os valores de dados de referência ao modelo.

4. Adicione os valores de rótulo ao modelo.

Você pode importar os dados do objeto de dados no repositório do Modelo. Você também pode inserir uma única linha de dados de referência ou um único rótulo.

Para usar o modelo probabilístico para analisar dados, verifique se o modelo contém um valor de rótulo para cada porta de saída que a transformação deve criar.

5. Atribua um rótulo para fazer referência a cada valor de dados em cada linha.

Você pode atribuir um rótulo a vários valores de dados de referência em uma única operação.

6. Compile o modelo.

Depois de compilar o modelo probabilístico, você poderá usá-lo em uma transformação.

Criar um Modelo Probabilístico Vazio

Você pode criar um objeto de modelo probabilístico que não contenha dados de referência nem dados de rótulo. Crie o modelo vazio e adicione ou importe dados para o modelo.

1. No Object Explorer, abra ou crie um conjunto de conteúdo.

2. Selecione a exibição **Conteúdo**.

3. Selecione **Modelos Probabilísticos** e clique em **Adicionar**.

O assistente de Modelo Probabilístico é aberto.

4. Selecione a opção **Modelo Probabilístico**.

Clique em **Avançar**.

5. Insira um nome para o modelo probabilístico.

Opcionalmente, insira uma descrição de texto para o modelo.

6. Clique em **Concluir**.

Criar um Modelo Probabilístico a partir de Objeto de Dados

Você pode usar um objeto de dados como a fonte dos dados do modelo probabilístico.

Um modelo probabilístico funciona de modo mais eficiente quando você usa os dados de entrada para a transformação do Rotulador ou do Analisador como a fonte dos dados de referência do modelo.

1. No Object Explorer, abra ou crie um conjunto de conteúdo.

2. Selecione a exibição **Conteúdo**.

3. Selecione **Modelos Probabilísticos** e clique em **Adicionar**.

O assistente de Modelo Probabilístico é aberto.

4. Selecione a opção **Modelo Probabilístico a partir de Objetos de Dados**.

Clique em **Avançar**.

5. Insira um nome para o modelo probabilístico.

Opcionalmente, insira uma descrição de texto para o modelo.

6. Procure o repositório do Modelo e selecione o objeto de dados que contém os dados a serem importados.

Não selecione um objeto de dados de mídia social.

Clique em **Avançar**.

7. Revise as colunas no objeto de dados e selecione uma ou mais colunas a serem adicionadas ao modelo. Você pode adicionar colunas de dados de referência e uma coluna de rótulo na mesma operação.

- Para importar uma coluna de dados como dados de referência, selecione o nome de coluna e clique em **Dados**.

Várias colunas de dados podem ser selecionadas. A Developer tool mescla o conteúdo das colunas que você selecionar em uma única coluna.

- Para importar uma coluna de dados como valores de rótulo, selecione o nome da coluna e clique em **Rótulo**.

Quando você importa dados de referência e valores de rótulo, a Developer tool atribui o rótulo em cada linha à cadeia de dados de referência na mesma linha. Você pode visualizar os dados antes de selecionar as colunas. Você poderá alterar as atribuições de rótulo depois de criar o modelo.

Clique em **Avançar**.

8. Selecione o número de linhas a serem copiadas da fonte de dados.

Por padrão, a Developer tool importará todas as linhas da fonte de dados. Se você inserir um número, o modelo conta as linhas a partir do início do conjunto de dados.

9. Especifique os delimitadores dos valores de dados que você importar.

Você pode especificar diferentes delimitadores para os valores de dados de referência e de rótulo. O delimitador padrão é um caractere de espaço.

10. Clique em **Concluir** e salve o modelo.

Depois de criar o modelo probabilístico, verifique as atribuições de rótulo e compile o modelo.

Anexando os Dados de uma Fonte de Dados a um Modelo Probabilístico

Você pode importar várias linhas de valores de dados de referência de rótulo para um modelo probabilístico em uma única operação.

1. Abra o conjunto de conteúdo que contém o modelo probabilístico.
2. Selecione o nome do modelo e clique em **Editar**.
3. Clique em **Anexar Dados**.

O assistente de Modelo Probabilístico é aberto.

4. Procure o repositório do Modelo e selecione o objeto de dados que contém os dados a serem importados.

Não selecione um objeto de dados de mídia social.

Clique em **Avançar**.

5. Revise as colunas no objeto de dados e selecione uma ou mais colunas a serem adicionadas ao modelo. Você pode adicionar colunas de dados de referência e uma coluna de rótulo na mesma operação.

- Para importar uma coluna de dados como dados de referência, selecione o nome de coluna e clique em **Dados**.

Várias colunas de dados podem ser selecionadas. A Developer tool mescla o conteúdo das colunas que você selecionar em uma única coluna.

- Para importar uma coluna de dados como valores de rótulo, selecione o nome da coluna e clique em **Rótulo**.

Quando você importa dados de referência e valores de rótulo, a Developer tool atribui o rótulo em cada linha à cadeia de dados de referência na mesma linha. Você pode visualizar os dados antes de selecionar as colunas. Você poderá alterar as atribuições de rótulo depois de criar o modelo.

Clique em **Avançar**.

6. Selecione o número de linhas a serem copiadas da fonte de dados.
Por padrão, a Developer tool importará todas as linhas da fonte de dados. Se você inserir um número, o modelo conta as linhas a partir do início do conjunto de dados.
7. Especifique os delimitadores dos valores de dados que você importar.
Você pode especificar diferentes delimitadores para os valores de dados de referência e de rótulo. O delimitador padrão é um caractere de espaço.
8. Clique em **Concluir** e salve o modelo.

Adicionar uma Linha de Dados de Referência a um Modelo Probabilístico

Use a exibição Dados para adicionar uma linha vazia a um modelo probabilístico.

1. Abra o conjunto de conteúdo que contém o modelo.
Selecione o nome do modelo e clique em **Editar**.
2. Selecione a exibição Dados.
3. Para adicionar uma linha vazia ao modelo, clique em **Novo**.
4. Selecione a linha que você adicionou e insira um ou mais valores de dados de referência na linha.
5. Salve o modelo probabilístico.

Depois de salvar o modelo, atribua um rótulo a cada valor na linha. Opcionalmente, compile o modelo.

Adicionar um Rótulo a um Modelo Probabilístico

Você pode adicionar um único rótulo em um modelo probabilístico. Adicione um rótulo para cada tipo de informações que os valores de dados do modelo representam. Se o modelo probabilístico em uma transformação do Analisador for usado, adicione um rótulo para cada porta de saída que você espera que a transformação crie.

1. Abra o conjunto de conteúdo que contém o modelo.
2. Selecione o nome do modelo e clique em **Editar**.
3. Na exibição Dados ou Rótulo, clique em **Gerenciar Rótulos**.
A caixa de diálogo **Gerenciar Rótulos** é exibida.
4. Na caixa de diálogo **Gerenciar Rótulos**, clique em **Novo**.
Um rótulo é exibido na primeira linha vazia da caixa de diálogo.
5. Edite o nome do rótulo. Opcionalmente, atualize a cor do rótulo.
6. Clique em **OK** para adicionar o rótulo ao modelo.
7. Salve o modelo probabilístico.

Depois de adicionar o rótulo, atribua-o a pelo menos um valor de dados.

Atribuindo um Rótulo a um Valor de Dados de Referência

Você pode atribuir um rótulo a um único valor de dados em uma linha de dados de referência.

Você poderá atribuir diferentes rótulos ao mesmo valor de dados se o valor de dados aparecer em diferentes locais na linha ou em linhas diferentes.

1. Abra o conjunto de conteúdo que contém o modelo.
2. Selecione o nome do modelo e clique em **Editar**.
3. Selecione a exibição Dados.
4. Localize o valor de dados sem um rótulo ou com um rótulo incorreto. Os valores de dados que usam um rótulo são codificados por cor.
5. Selecione a linha de dados que contém o valor de dados.

A linha é exibida no editor.

6. Clique com o botão direito do mouse em um valor de dados no editor e selecione um rótulo no menu de contexto.

A Developer tool atribui o rótulo ao valor de dados.

7. Salve o modelo probabilístico.

Opcionalmente, depois de salvar o modelo probabilístico, compile-o.

Atribuindo um Rótulo a Vários Valores de Dados

Você pode atribuir um rótulo a vários valores de dados de referência em uma única operação.

1. Abra o conjunto de conteúdo que contém o modelo.
2. Selecione o nome do modelo e clique em **Editar**.
3. Clique em **Atribuir Rótulo**.

A caixa de diálogo **Atribuir um Rótulo a Vários Valores** é aberta.

4. Insira um ou mais caracteres no campo Localizar.

Você pode inserir caracteres curinga no campo Localizar.

5. Opcionalmente, selecione outro critério de pesquisa.

Você pode selecionar ou desmarcar as seguintes opções:

- Corresponder maiúsculas e minúsculas.

Especifica que a operação de pesquisa faz distinção entre maiúsculas e minúsculas. Não use caracteres curinga com a opção.

- Corresponder a cadeia completa Especifica que a operação de pesquisa procura por uma correspondência completa entre os caracteres no valor de dados de referência e os caracteres que você inserir. Não use caracteres curinga com a opção.

- Ignorar valores de rótulo.

Especifica que a operação de pesquisa ignorará todos os valores de dados de referência que usarem um rótulo.

6. Selecione um rótulo a ser atribuído aos valores de dados de referência que correspondem aos critérios de pesquisa.

Você também pode selecionar a opção **Nenhum Rótulo**. Selecione a opção para remover o rótulo dos valores de dados de referência que incluem os caracteres que você inserir.

7. Clique em **Iniciar**.

A Developer tool atribui o rótulo a todos os valores de dados de referência que correspondem aos critérios de pesquisa que você definir.

Nota: Para exibir os valores de dados de referência que você rotulou em uma única operação, use o filtro **Atribuído em massa** na exibição Rótulo.

Excluindo Linhas de um Modelo Probabilístico

Você pode excluir uma ou mais linhas de dados de referência de um modelo probabilístico em uma única ação.

1. Abra o conjunto de conteúdo que contém o modelo.
2. Selecione o nome do modelo e clique em **Editar**.
3. Na exibição Dados, selecione uma ou mais linhas de dados de referência.
4. Clique em **Excluir**.

A Developer tool remove as linhas que você selecionou do modelo do classificador.

Para desfazer a operação, pressione as teclas Ctrl + Z no teclado.

Excluir um Rótulo de um Modelo Probabilístico

Quando você exclui um valor de rótulo de um modelo, qualquer valor de dados de referência que usou o rótulo permanece no modelo. Atribua outro valor de rótulo a cada valor de dados de referência.

1. Abra o conjunto de conteúdo que contém o modelo.
2. Selecione o nome do modelo e clique em **Editar**.
3. Na exibição Dados ou Rótulo, clique em **Gerenciar Rótulos**.
4. Na caixa de diálogo **Gerenciar Rótulos**, selecione um valor de rótulo.
5. Clique em **Excluir**.
6. Clique em **OK** para excluir o rótulo.
7. Salve o modelo probabilístico.

Nota: Um rótulo é um elemento estrutural de um modelo probabilístico. Se você adicionar ou remover um rótulo depois de adicionar o modelo a uma transformação, a operação que usa esse modelo será invalidada. Para usar o modelo que você atualizou, exclua e crie a operação de transformação.

Compilar o Modelo Probabilístico

Ao atualizar os dados ou as atribuições de rótulo em um modelo probabilístico, você pode compilar o modelo. Compilar o modelo para atualizar a lógica do modelo com as associações entre os valores de dados de referência atuais e os valores de rótulo atuais.

Antes de compilar o modelo probabilístico, verifique se cada valor de rótulo identifica pelo menos um valor de dados de referência.

- Para compilar o modelo, abra o modelo na Developer tool e clique em **Compilar**.

Localizando Linhas de Dados em um Modelo Probabilístico

Use a exibição Dados para localizar as linhas de dados de referência que contêm um valor que você inserir.

1. Abra o conjunto de conteúdo que contém o modelo probabilístico.
2. Selecione o nome do modelo e clique em **Editar**.
3. Selecione a exibição Dados.
4. Insira um ou mais caracteres no campo **Localizar**.

A exibição de Dados exibe a primeira linha no modelo que contém o valor inserido.

5. Use a seta para cima ou para baixo para mover até outras linhas que contêm o valor.

Filtrando Valores de Dados de Referência por Atribuição de Rótulo

Use a exibição Rótulo para localizar os valores de dados de referência que utilizam o rótulo que você especificar. Filtre os resultados com base no método que você utilizou para atribuir o rótulo.

1. Abra o conjunto de conteúdo que contém o modelo probabilístico.
2. Selecione o nome do modelo e clique em **Editar**.
3. Na exibição Rótulo, selecione um valor de rótulo.

O modelo probabilístico exibe uma lista dos valores de dados de referência que usam o rótulo. O modelo também mostra o número de valores de dados que usam o rótulo.

4. Aplique um filtro à lista de valores de dados de referência que utilizam o rótulo.

Selecione um dos seguintes filtros:

- Todos. Exibe os valores de dados de referência que usam o rótulo. Todos é a opção padrão.
- Atribuído pelo usuário. Exibe qualquer valor de dados de referência que você selecionou individualmente quando atribuiu o rótulo.
- Atribuído em massa. Exibe os valores de dados de referência aos quais você atribuiu um rótulo como parte de uma operação de atribuição em massa.

O modelo probabilístico exibe os valores de dados de referência que atendem à condição de filtragem.

Localizando Valores de Rótulo Não Utilizados

Use a exibição Rótulo para localizar qualquer valor de rótulo que você não tiver atribuído a um valor de dados de referência. Você deve atribuir cada rótulo a pelo menos um valor de dados de referência.

1. Abra o conjunto de conteúdo que contém o modelo probabilístico.
2. Selecione o nome do modelo e clique em **Editar**.
3. Na exibição Rótulo, selecione um valor de rótulo.

O modelo probabilístico exibe uma lista dos valores de dados de referência que usam o rótulo. O modelo também mostra o número total de valores de dados que usam o rótulo.

Se o número total de valores de dados for zero, você não atribuiu o rótulo a nenhum valor de dados de referência no modelo probabilístico.

Operações de Copiar e Colar

O modelo probabilístico pode ser copiado de um conjunto de conteúdo para outro em um repositório do Modelo. Copie um modelo probabilístico para compartilhar recursos com outros usuários da ferramenta Developer.

Você pode copiar um modelo para outro conjunto de conteúdo, ou pode importar um modelo para o conjunto de conteúdo atual. Também pode importar vários modelos de vários conjuntos de conteúdo no repositório em uma única operação.

Quando você copia um modelo, o Serviço do Gerenciamento de Conteúdo cria uma cópia do arquivo de dados de modelo na máquina host dos serviços Informatica. Cada modelo usa um arquivo de dados de diferentes.

Copiar um Modelo Probabilístico para Outro Conjunto de Conteúdo

O modelo probabilístico pode ser copiado de um conjunto de conteúdo para outro em um repositório do Modelo. Quando você copia um modelo probabilístico, o objeto do modelo e os conjuntos de conteúdo de origem e destino são especificados.

1. Abra o conjunto de conteúdo que contém o modelo probabilístico.
2. Selecione um modelo probabilístico e clique em **Copiar para**.
3. Procure o repositório do Modelo e selecione um conjunto de conteúdo.

É possível copiar o modelo probabilístico para um conjunto de conteúdo no projeto atual ou em outro projeto.

4. Clique em **OK**.

A ferramenta Developer copia o modelo probabilístico para o conjunto de conteúdo selecionado.

Importar um Modelo Probabilístico de Outro Conjunto de Conteúdo

O modelo probabilístico pode ser importado de um conjunto de conteúdo para outro em um repositório do Modelo. Quando você importa um modelo probabilístico, um ou mais objetos do modelo e os conjuntos de conteúdo de origem e destino são especificados.

1. Abra o conjunto de conteúdo para incluir o modelo probabilístico.
2. Selecione um modelo probabilístico e clique em **Colar de**.
3. Procure o repositório do Modelo e selecione um modelo probabilístico.

É possível colar o modelo probabilístico de um conjunto de conteúdo no projeto atual ou de outro projeto.

4. Clique em **OK**.

A ferramenta Developer cola o modelo probabilístico no conjunto de conteúdo atual.

Copiando Linhas de Dados de Referência para a Área de Transferência

Você pode copiar um ou mais linhas de dados de referência de um modelo probabilístico para a área de transferência. Você pode colar as linhas em outro modelo probabilístico.

1. Abra o conjunto de conteúdo que contém o modelo probabilístico.
2. Selecione o nome do modelo e clique em **Editar**.

3. Na exibição Dados, selecione uma ou mais linhas de dados de referência.
4. Use as teclas Ctrl + C para copiar as linhas para a área de transferência.

A operação copia os dados de referência e os valores de rótulo que você atribuiu aos dados de referência.

Você pode usar as teclas Ctrl + V para colar as linhas em um editor de texto ou na exibição Dados de outro modelo probabilístico.

ÍNDICE

C

- conjuntos de caracteres [36](#)
- conjuntos de conteúdo
 - conjuntos de caracteres [36](#)
 - conjuntos de padrões [37](#)
 - conjuntos de tokens [39](#)
 - controle de versão [13](#), [22](#), [30](#)
 - expressões regulares [38](#)
 - modelos do classificador [37](#)
 - modelos probabilísticos [37](#)
- conjuntos de padrões [37](#)
- conjuntos de tokens [39](#)
- controle de versão
 - conjuntos de conteúdo [13](#), [30](#)
 - tabelas de referência [13](#)
 - tabelas de referência na Developer tool [30](#)
 - tabelas de referência na ferramenta Analyst [22](#)
- criando uma tabela de referência a partir de dados de colunas de perfil
 - tabelas de referência [17](#)
- criar uma tabela de referência a partir de padrões de coluna
 - tabelas de referência [18](#)
- criar uma tabela de referência manualmente
 - tabelas de referência [16](#)

E

- exibir eventos de tabela de auditoria
 - tabelas de referência [27](#)
- exportar uma tabela de referência
 - tabelas de referência [25](#)
- expressões regulares [38](#)

F

- ferramenta Analyst
 - localizar e substituir valores de dados de referência [25](#)

G

- gerenciar colunas
 - tabelas de referência [23](#)
- gerenciar linhas
 - tabelas de referência [24](#)

I

- importando uma tabela de referências
 - tabelas de referência [20](#)

M

- modelos do classificador
 - em conjuntos de conteúdo [37](#)
 - regras e diretrizes [40](#)
- modelos probabilísticos
 - em conjuntos de conteúdo [37](#)
 - regras e diretrizes [40](#)

P

- privilégios
 - Serviço do Gerenciamento de Conteúdo [12](#)

S

- Serviço do Gerenciamento de Conteúdo
 - privilégios da tabela de referência [12](#)

T

- tabelas de referência
 - atualizando na ferramenta Analyst [26](#)
 - controle de versão [13](#), [22](#), [30](#)
 - criar uma tabela de referência a partir das colunas de perfil [17](#)
 - criar uma tabela de referência a partir de padrões de coluna [18](#)
 - criar uma tabela de referência manualmente [16](#)
 - data warehouse de referência [11](#)
 - exibir as tabelas de trilha de auditoria [27](#)
 - exportar uma tabela de referência [25](#)
 - gerenciada e não gerenciada [11](#)
 - gerenciar colunas [23](#)
 - gerenciar linhas [24](#)
 - importando uma tabela de referências [20](#)
 - localizar e substituir valores na ferramenta Analyst [25](#)
 - na análise baseada em padrões [11](#)
 - privilégios [12](#)
 - propriedades da ferramenta Developer [32](#)
 - propriedades na ferramenta Analyst [15](#)
 - Serviço do Gerenciamento de Conteúdo [11](#)
 - tabelas de referência gerenciadas [11](#)
 - tabelas de referência não gerenciadas [11](#)
 - Visão geral da Developer tool [31](#)
 - Visão geral da ferramenta Analyst [14](#)
- tabelas de referência gerenciadas [11](#)
- tabelas de referência não gerenciadas
 - ativar e desativar edições [26](#)
 - definição [11](#)
 - sincronizando com o repositório do Modelo [12](#)