



Informatica® Data Quality
10.4.0

참조 데이터 가이드

Informatica Data Quality 참조 데이터 가이드

10.4.0

2019년 12월

© 저작권 Informatica LLC 2013, 2020

이 소프트웨어와 설명서는 사용 및 공개에 대한 제한 사항이 포함되어 있는 별도의 사용권 계약에 따라서만 제공됩니다. 본 문서의 어떤 부분도 Informatica LLC의 사전 통지 없이 어떠한 형태나 수단(전자적, 사진 복사, 녹음 등)으로 복제되거나 전송될 수 없습니다.

미국 정부 권한. 미국 정부 고객에게 제공되는 프로그램, 소프트웨어, 데이터베이스, 관련 문서 및 기술 데이터는 해당하는 연방 입수 규정 및 기관별 보안 규정에 따라 "상용 컴퓨터 소프트웨어" 또는 "상용 기술 데이터"입니다. 따라서 사용, 복제, 공개, 수정 및 조정은 해당하는 정부 계약에 규정된 제한 사항 및 라이선스 조건을 따르며, 정부 계약 조건에 의해 적용 가능한 한도 내에서, FAR 52.227-19, 상용 소프트웨어 라이선스에 규정된 추가 권한이 적용됩니다.

Informatica 및 Informatica 로고는 미국과 전 세계 여러 관할 국가에서 Informatica LLC의 상표 또는 등록 상표입니다. Informatica 상표의 현재 목록은 <https://www.informatica.com/trademarks.html> 웹에서 확인할 수 있습니다. 다른 회사 및 제품명은 해당 소유자의 상표 또는 등록 상표일 수 있습니다.

이 소프트웨어 및/또는 설명서의 일부에는 타사의 저작권이 적용될 수 있습니다. 필요한 타사 고지 사항은 제품에 포함되어 있습니다.

이 설명서의 정보는 예고 없이 변경될 수 있습니다. 이 문서에서 문제가 발견되는 경우 infa_documentation@informatica.com으로 보고해 주십시오.

Informatica 제품은 제품이 제공될 당시의 계약 조건에 따라 보증됩니다. Informatica는 상품성과 특정 목적에의 적합성에 대한 보증 그리고 비침해에 대한 보증 또는 조건을 포함하여 어떠한 종류의 명시적이거나 묵시적인 보증 없이 이 문서의 정보를 "있는 그대로" 제공합니다.

발행 날짜: 2020-02-07

목차

서문	7
Informatica 리소스	7
Informatica 네트워크	7
Informatica 기술 자료	7
Informatica 설명서	7
Informatica Product Availability Matrix	8
Informatica Velocity	8
Informatica Marketplace	8
Informatica 글로벌 고객 지원 센터	8
장 1: 참조 데이터 소개	9
참조 데이터 개요	9
Informatica 참조 데이터	10
사용자 정의 참조 데이터	10
참조 테이블	11
참조 테이블 구조	11
참조 데이터 웨어하우스 권한	12
매개 변수 및 참조 테이블	12
참조 데이터 개체 및 버전 제어	12
장 2: Analyst 도구의 참조 테이블	13
Analyst 도구 참조 테이블 개요	13
참조 테이블 속성	13
참조 테이블 일반 속성	14
참조 테이블 열 속성	14
참조 테이블 편집기에서 참조 테이블 작성	15
프로필 데이터를 사용하여 참조 테이블 작성	15
프로필 열 데이터에서 참조 테이블 작성	16
값 패턴에서 참조 테이블 작성	17
플랫 파일을 사용하여 참조 테이블 작성	18
Analyst 도구 플랫 파일 속성	18
플랫 파일을 사용하여 참조 테이블 작성	18
데이터베이스 테이블을 사용하여 참조 테이블 작성	19
데이터베이스 테이블을 사용하여 참조 테이블 작성	19
버전이 지정된 모델 리포지토리에서 참조 테이블로 작업	20
참조 테이블 업데이트	21
열 관리	21
행 관리	22
값 찾기 및 바꾸기	22

참조 테이블 데이터 내보내기.....	23
관리되지 않는 참조 테이블에서 편집 활성화 및 비활성화.....	24
참조 테이블 값 새로 고침.....	24
감사 내역 이벤트.....	24
감사 내역 이벤트 보기.....	25
참조 테이블 관련 규칙 및 지침.....	25

장 3: Developer 도구의 참조 데이터..... 27

Developer tool 참조 데이터 개요.....	27
참조 데이터와 변환.....	27
버전이 지정된 모델 리포지토리에서 참조 데이터 개체로 작업.....	28
참조 데이터 개체 체크 아웃.....	28
참조 데이터 개체 체크 인.....	28
참조 테이블.....	29
참조 테이블 데이터 속성.....	29
참조 테이블 개체 작성.....	30
플랫 파일을 사용하여 참조 테이블 작성.....	31
관계형 소스에서 참조 테이블 작성.....	32
콘텐츠 집합.....	33
문자 집합.....	34
분류자 모델.....	34
패턴 집합.....	35
확률 모델.....	35
정규식.....	35
토큰 집합.....	36
확률 모델 및 분류자 모델에 대한 규칙 및 지침.....	37
콘텐츠 집합 작성.....	38
콘텐츠 집합에 참조 데이터 개체 작성.....	39

장 4: 분류자 모델..... 40

분류자 모델 개요.....	40
분류자 모델 구조.....	41
분류자 점수.....	41
분류자 변환 예.....	41
분류자 모델 옵션.....	42
분류자 모델 참조 데이터.....	43
분류자 모델 레이블 데이터.....	44
분류자 모델 레이블 관리.....	44
분류자 모델 구성.....	45
분류자 모델 작성.....	45
데이터 소스에서 분류자 모델로 데이터 추가.....	46
분류자 모델에 참조 데이터 행 추가.....	47

분류자 모델에 레이블 추가.	47
참조 데이터 행에 레이블 할당.	47
사용되지 않은 레이블 값 식별.	48
분류자 모델에서 행 삭제.	48
분류자 모델에서 레이블 삭제.	48
분류자 모델 컴파일.	48
필터 작업 및 찾기 작업.	49
데이터 값을 사용하여 참조 데이터 행 필터링.	49
레이블 값을 사용하여 참조 데이터 행 필터링.	49
참조 데이터 행에서 값 찾기.	49
작업 복사 및 붙여넣기.	50
분류자 모델을 다른 콘텐츠 집합에 복사.	50
다른 콘텐츠 집합에서 분류자 모델 가져오기.	50

장 5: 확률 모델..... 51

확률 모델 개요.	51
확률 모델 구조.	52
라벨러 변환 예제.	52
파서 변환 예제.	53
확률 모델 옵션.	53
확률 모델 데이터 보기.	54
확률 모델 레이블 보기.	55
확률 모델 참조 데이터.	56
확률 모델 레이블 데이터.	56
오버플로우 레이블.	57
확률 모델 속성.	58
확률 모델 구성.	58
빈 확률 모델 작성.	59
데이터 개체를 기반으로 확률 모델 생성.	59
데이터 소스에서 확률 모델로 데이터 추가.	60
확률 모델에 참조 데이터 행 추가.	60
확률 모델에 레이블 추가.	61
참조 데이터 값에 레이블 할당.	61
여러 데이터 값에 레이블 할당.	61
확률 모델에서 행 삭제.	62
확률 모델에서 레이블 삭제.	62
확률 모델 컴파일.	63
확률 모델에서 데이터 행 찾기.	63
레이블 할당으로 참조 데이터 필터링.	63
사용되지 않은 레이블 값 찾기.	63
작업 복사 및 붙여넣기.	64
확률 모델을 다른 콘텐츠 집합에 복사.	64

다른 콘텐츠 집합에서 확률 모델 가져오기.....	64
클립보드에 참조 데이터 행 복사.....	64
부록 A: 참조 데이터 및 Informatica Data Engineering Integration.....	66
참조 데이터 및 Informatica Data Engineering Integration.	66
주소 유효성 검사를 위한 참조 데이터.....	66
주소 참조 데이터 파일 설치.....	67
인덱스.....	68

서문

*Informatica 참조 데이터 가이드*에서 Informatica Developer와 Informatica Analyst에서 사용할 수 있는 파일 및 참조 데이터 개체에 대해 자세히 알아보십시오. 참조 데이터를 사용하면 소스 데이터의 값을 비즈니스 요구 사항에 따라 식별, 분석, 개선 또는 제거할 수 있습니다.

Informatica 리소스

Informatica는 Informatica Network 및 기타 온라인 포털을 통해 다양한 범위의 제품 리소스를 제공합니다. 리소스를 통해 Informatica 제품 및 솔루션을 최대한 활용하고 다른 Informatica 사용자 및 주제별 전문가로부터 배울 수 있습니다.

Informatica 네트워크

Informatica Network는 Informatica 기술 자료, Informatica 글로벌 고객 지원 센터 등 여러 리소스로 연결되는 관문입니다. Informatica Network를 시작하려면 <https://network.informatica.com>을 방문하십시오.

Informatica Network 멤버인 경우 다음 옵션이 가능합니다.

- 기술 자료에서 제품 리소스를 검색할 수 있습니다.
- 제품 사용 가능 여부에 대한 정보를 봅니다.
- 지원 사례를 생성하고 검토할 수 있습니다.
- 거주 지역의 Informatica 사용자 그룹 네트워크를 검색하고 동료와 협업 관계 유지

Informatica 기술 자료

Informatica 기술 자료를 사용하여 사용 방법 문서, 모범 사례, 비디오 자습서, 자주 묻는 질문에 대한 답변 등 제품 리소스를 확인할 수 있습니다.

기술 자료를 검색하려면 <https://search.informatica.com>을 방문하십시오. 기술 자료에 대한 질문, 의견 또는 아이디어가 있는 경우 KB_Feedback@informatica.com을 통해 Informatica 기술 자료 팀에 문의해 주시기 바랍니다.

Informatica 설명서

Informatica 설명서 포털에서 확장된 설명서 라이브러리를 탐색하여 현재 및 최근 제품 릴리스를 확인할 수 있습니다. 설명서 포털을 탐색하려면 <https://docs.informatica.com>을 방문하십시오.

제품 설명서에 대한 질문, 의견 또는 아이디어가 있는 경우 infa_documentation@informatica.com에서 Informatica 설명서 팀에 문의해 주시기 바랍니다.

Informatica Product Availability Matrix

PAM(Product Availability Matrix)은 제품 릴리스에서 지원하는 운영 체제 버전, 데이터베이스 및 데이터 소스 유형과 대상을 나타냅니다.

<https://network.informatica.com/community/informatica-network/product-availability-matrices>에서 Informatica PAM을 찾을 수 있습니다.

Informatica Velocity

Informatica Velocity는 수백 가지 데이터 관리 프로젝트의 실제 경험을 토대로 Informatica 전문 서비스업에서 개발한 팁과 모범 사례 모음입니다. Informatica Velocity는 전 세계의 조직과 협력하여 성공적인 데이터 관리 솔루션을 계획, 개발, 배포 및 유지 관리하는 Informatica 컨설턴트의 포괄적인 지식을 보여줍니다.

Informatica Velocity 리소스는 <http://velocity.informatica.com>에서 확인할 수 있습니다. Informatica Velocity에 대한 질문, 주석 또는 아이디어가 있으시면 Informatica 전문 서비스업(ips@informatica.com)에 문의하십시오.

Informatica Marketplace

Informatica Marketplace는 Informatica 구현을 확대 및 개선하기 위한 솔루션을 찾을 수 있는 포럼입니다. Marketplace에서 Informatica 개발자와 파트너가 제공하는 수백 개의 솔루션을 활용하여 생산성을 향상시키고 프로젝트의 구현에 걸리는 시간을 줄일 수 있습니다. <https://marketplace.informatica.com>에서 Informatica Marketplace를 찾을 수 있습니다.

Informatica 글로벌 고객 지원 센터

전화 또는 Informatica Network를 통해 글로벌 지원 센터에 문의할 수 있습니다.

해당 지역의 Informatica 글로벌 고객 지원 전화 번호는 Informatica 웹 사이트 (<https://www.informatica.com/services-and-training/customer-success-services/contact-us.html>)를 방문하여 찾을 수 있습니다.

Informatica Network에서 온라인 지원 리소스를 찾으려면 <https://network.informatica.com>을 방문하고 eSupport 옵션을 선택하십시오.

제 1 장

참조 데이터 소개

이 장에 포함된 항목:

- [참조 데이터 개요, 9](#)
- [Informatica 참조 데이터, 10](#)
- [사용자 정의 참조 데이터, 10](#)
- [참조 테이블, 11](#)
- [참조 데이터 개체 및 버전 제어, 12](#)

참조 데이터 개요

Informatica 변환은 참조 데이터를 사용하여 데이터를 분석하고 업데이트할 수 있습니다. Developer tool과 Analyst 도구에서 참조 데이터 개체를 작성할 수 있습니다. 또한 참조 데이터 개체 및 파일을 모델 리포지토리 및 파일 시스템으로 가져올 수 있습니다.

다음과 같은 유형의 참조 데이터를 작성하고 편집할 수 있습니다.

참조 테이블

참조 테이블에는 표준 버전과 대체 버전의 데이터 값 집합이 포함됩니다. Developer tool에서 참조 테이블을 변환에 추가하여 소스 데이터 값이 정확하고 올바른 형식인지 확인할 수 있습니다.

대부분의 참조 테이블에는 두 개 이상의 열이 포함됩니다. 한 개의 열에는 표준 또는 기본 설정 버전의 값이 포함되고 다른 열에는 대체 버전이 포함됩니다. 참조 테이블을 변환에 추가하면 변환에서 입력 포트 데이터를 검색하여 참조 테이블에도 나타나는 값을 찾습니다. 작업하는 데이터 프로젝트에 유용한 모든 데이터를 사용하여 테이블을 작성할 수 있습니다.

콘텐츠 집합

콘텐츠 집합은 리포지토리 또는 파일에 참조 데이터 값을 지정하는 모델 리포지토리 개체입니다. 콘텐츠를 집합을 변환에 추가하면 변환에서 입력 데이터를 검색하여 콘텐츠 집합의 데이터 패턴과 일치하는 값을 찾습니다.

Informatica에서 다음과 같은 유형의 참조 데이터를 다운로드하고 설치할 수 있습니다.

주소 참조 데이터 파일

특정 국가에서 배송 가능한 주소 데이터를 포함하는 참조 데이터 파일입니다. 주소 유효성 검사기 변환이 참조 데이터를 읽습니다. 주소 참조 데이터 파일은 작성하거나 편집할 수 없습니다.

주소 참조 데이터는 정의된 기간에 대해 최신이고 분기별 등의 정기적인 간격으로 데이터를 새로 고쳐야 합니다.

ID 인구집단 파일

개인, 가정 및 회사 ID에 대한 정보를 포함하는 참조 데이터 파일입니다. 일치 변환 및 비교 변환이 채우기 파일을 기반으로 입력 데이터에서 잠재적인 ID를 찾습니다. ID 인구집단 파일은 작성하거나 편집할 수 없습니다.

Informatica 참조 데이터

Informatica에서 주소 참조 데이터와 ID 채우기 데이터를 구입하고 다운로드할 수 있습니다.

참조 데이터는 Informatica 응용 프로그램과 별개로 다운로드되고 설치됩니다. 시스템에 설치할 참조 데이터에 대한 정보는 관리자에게 문의하십시오.

확인하려는 주소가 포함된 국가에 대한 주소 데이터 파일을 구독 기반으로 구매해야 합니다. 구독 기간 동안 언제든지 Informatica에서 최신 주소 데이터를 다운로드할 수 있습니다.

사용자 정의 참조 데이터

데이터 개체의 값을 사용하여 참조 데이터 개체를 작성할 수 있습니다.

예를 들어 프로젝트 또는 조직과 관련된 값을 포함하는 데이터 개체 또는 프로필 열을 선택할 수 있습니다. 열 값을 기반으로 사용자 지정 참조 데이터 개체를 작성합니다.

데이터 열을 기반으로 참조 데이터 개체를 작성하여 다음을 확인할 수 있습니다.

- 열의 데이터 행이 동일한 유형의 정보를 포함하는지 여부.
- 소스 값이 올바른지 여부. 참조 개체에 올바른 값 목록이 포함될 수도 있고 잘못된 값 목록이 포함될 수도 있습니다.

다음 표에는 참조 데이터를 포함할 수 있는 프로젝트 데이터 열의 일반적인 예가 나와 있습니다.

정보	참조 데이터 예
SKU(Stock Keeping Unit) 코드	SKU 열을 사용하여 조직에 대한 올바른 SKU 코드를 포함하는 참조 테이블을 작성할 수 있습니다. 참조 테이블을 통해 데이터 집합에서 올바른거나 잘못된 SKU 코드를 찾을 수 있습니다.
직원 코드	직원 코드 또는 직원 ID 열을 기반으로 올바른 직원 코드를 포함하는 참조 테이블을 작성할 수 있습니다. 참조 테이블을 통해 직원 데이터에서 오류를 찾을 수 있습니다.
고객 계정 번호	고객 계정 열에 대해 프로필을 실행하여 계정 번호 패턴을 확인할 수 있습니다. 프로필을 사용하여 잘못된 데이터 패턴의 토큰 집합을 작성할 수 있습니다. 토큰 집합을 기반으로 올바른 계정 번호 구조를 따르지 않는 계정 번호를 찾을 수 있습니다.
고객 이름	고객 이름 열에 이름, 중간 이름, 성이 포함되는 경우 열에 있는 문자열의 예상 구조를 정의하는 확률 모델을 작성할 수 있습니다. 확률 모델을 기반으로 열에 속하지 않는 데이터 문자열을 찾을 수 있습니다.

참조 테이블

Analyst 도구 및 Developer tool에서 참조 테이블을 작성하고 업데이트할 수 있습니다.

참조 테이블은 모델 리포지토리에 메타데이터를 저장합니다. 참조 테이블은 참조 데이터 웨어하우스 또는 다른 데이터베이스에 열 데이터를 저장할 수 있습니다. 참조 데이터 웨어하우스에서 열 데이터를 저장하는 경우 Informatica 서비스는 테이블을 관리되는 참조 테이블로 식별합니다. 다른 데이터베이스에서 열 데이터를 저장하는 경우 Informatica 서비스는 테이블을 관리되지 않는 참조 테이블로 식별합니다.

다른 데이터베이스의 데이터를 참조 데이터 웨어하우스로 가져오는 경우 원시 연결 또는 ODBC 연결을 사용하여 데이터를 가져옵니다. 관리되지 않는 데이터베이스를 참조 테이블의 데이터 소스로 지정하는 경우 원시 연결을 사용하여 데이터베이스에 연결합니다.

콘텐츠 관리 서비스의 참조 데이터 웨어하우스 옵션

콘텐츠 관리 서비스는 참조 데이터 웨어하우스에 대한 연결을 저장하는 응용 프로그램 서비스입니다. IBM DB2 데이터베이스, Microsoft SQL Server 데이터베이스 또는 Oracle 데이터베이스를 참조 데이터 웨어하우스로 지정할 수 있습니다. 필요에 따라 참조 데이터 웨어하우스의 참조 테이블에 대한 스키마를 콘텐츠 관리 서비스의 속성으로 지정할 수 있습니다.

참조 데이터 웨어하우스 및 스키마에 대한 다음 규칙 및 지침을 고려하십시오.

- 관리되는 참조 테이블을 생성하기 전에 콘텐츠 관리 서비스가 참조 데이터에 사용할 데이터베이스 및 스키마를 설정합니다.
- 콘텐츠 관리 서비스에 참조 테이블 스키마를 지정하지 않으면 서비스는 데이터베이스 연결에 지정된 스키마를 사용합니다. 콘텐츠 관리 서비스 또는 데이터베이스 연결에 스키마를 지정하지 않으면 서비스는 기본 데이터베이스 스키마를 사용합니다.

참조 테이블 구조

대부분의 참조 테이블에는 두 개 이상의 열이 포함됩니다. 한 개의 열에는 올바른 버전 또는 필요한 버전의 데이터 값이 포함됩니다. 나머지 다른 열에는 소스 데이터에 나타날 수 있는 대체 버전을 비롯한 여러 가지 버전의 값이 포함됩니다.

올바른 값 또는 필요한 값을 포함하는 열을 유효한 열이라고 합니다. 변환이 매핑의 참조 테이블을 읽을 때 유효하지 않은 열에서 값을 찾습니다. 변환이 유효하지 않은 값을 찾을 경우 유효한 열에서 해당 값을 반환합니다. 여러 개의 올바른 값이 아닌 하나의 공통 값을 반환할 수도 있습니다.

유효한 열에 ZIP 코드와 같이 공식적으로 올바른 데이터가 포함될 수 있습니다. 조직에 고유한 SKU(Stock Keeping Unit) 번호와 같은 프로젝트 관련 데이터가 이 열에 포함될 수 있습니다. 또한 검색하려는 알려진 데이터 오류가 포함된 값과 같이 잘못된 데이터를 사용하여 유효한 열을 작성할 수도 있습니다.

예를 들어 소매 조직의 올바른 SKU 번호 목록이 포함된 참조 테이블을 작성합니다. 참조 테이블을 라벨러 변환에 추가하고 이 변환과의 매핑을 작성합니다. 제품 데이터베이스 테이블에 대해 매핑을 실행합니다. 매핑이 실행되면 라벨러가 올바른 SKU 번호를 포함하지 않는 제품 레코드를 나타내는 열을 작성합니다.

참조 테이블과 파서 변환

패턴 기반 구문 분석 작업에서 테이블 데이터를 사용하려면 단일 열로 참조 테이블을 작성합니다. 패턴 기반 구문 분석을 수행하도록 파서 변환을 구성하고 참조 데이터를 변환 구성으로 가져옵니다.

참조 데이터 웨어하우스 권한

콘텐츠 관리 서비스는 권한을 사용하여 참조 테이블에 대한 사용자 작업을 제한합니다. Administrator 도구에서 보안 옵션을 사용하여 서비스 권한을 검토하거나 업데이트합니다.

참조 테이블을 작업하려면 콘텐츠 관리 서비스에서 다음과 같은 권한이 있어야 합니다.

- 참조 테이블 작성
- 참조 테이블 데이터 편집
- 참조 테이블 메타데이터 편집

관리되지 않는 참조 테이블에서 데이터를 편집하려는 경우 편집을 허용하도록 참조 테이블 개체를 구성했는지도 확인합니다.

참고: 데이터베이스 응용 프로그램에서 관리되지 않는 참조 테이블에 대한 메타데이터를 편집하는 경우 Analyst 도구를 사용하여 모델 리포지토리를 테이블과 동기화합니다. Developer tool에서 관리되지 않는 참조 테이블을 사용하려면 먼저 모델 리포지토리와 테이블을 동기화해야 합니다.

매개 변수 및 참조 테이블

매개 변수를 사용하여 모델 리포지토리에서 참조 테이블을 식별할 수 있습니다. 참조 테이블을 식별하는 Developer tool에서 매개 변수를 작성할 수 있습니다. 또는 매개 변수 파일에 참조 테이블 위치를 추가할 수 있습니다.

Developer tool에서 매개 변수를 작성할 때 매개 변수를 매핑의 변환에 추가합니다. 매개 변수 파일에 참조 테이블 위치를 추가하는 경우 명령 프롬프트에서 매핑을 실행할 때 파일을 지정합니다. 각 경우에서, 데이터 통합 서비스는 매핑을 실행할 때 매개 변수가 식별하는 참조 테이블을 읽습니다.

참조 테이블을 식별하는 매개 변수를 다음 변환에 추가할 수 있습니다.

- 대/소문자 변환기 변환
- 라벨러 변환
- 토큰 구문 분석 모드에서의 파서 변환
- 표준화 변환

참고: 명령 프롬프트에서 매핑을 실행하려면 `infacmd ms runMapping` 명령을 사용합니다.

참조 데이터 개체 및 버전 제어

참조 데이터 개체를 저장하는 모델 리포지토리가 버전 제어 응용 프로그램과 통합되는 경우 버전 제어를 개체에 적용할 수 있습니다. 버전 제어를 참조 테이블 및 콘텐츠 집합에 적용할 수 있습니다.

버전 제어를 지원하는 모델 리포지토리에서 참조 데이터 개체를 체크 인 및 체크 아웃할 수 있습니다. 체크 아웃을 실행 취소하고 개체의 이전 버전을 검색하고 개체를 이전 버전으로 복원할 수 있습니다. 참조 데이터 개체가 버전 제어를 통해 관리되지 않는 경우 모델 리포지토리가 사용자가 편집하는 참조 데이터 개체를 찾습니다. 다른 사용자는 사용자가 작업하는 잠긴 개체를 편집할 수 없습니다. 개체를 닫는 경우 모델 리포지토리가 잠금을 해제하고 다른 사용자가 개체를 편집할 수 있습니다.

참고: 버전 제어는 모델 리포지토리가 관리되지 않는 참조 테이블 개체에 대해 저장하는 메타데이터에 적용됩니다. 버전 제어는 관리되지 않는 참조 테이블의 데이터에 적용되지 않습니다. 이전 버전의 관리되지 않는 참조 테이블에서 참조 데이터를 보거나 복원할 수 없습니다.

제 2 장

Analyst 도구의 참조 테이블

이 장에 포함된 항목:

- [Analyst 도구 참조 테이블 개요, 13](#)
- [참조 테이블 속성, 13](#)
- [참조 테이블 편집기에서 참조 테이블 작성, 15](#)
- [프로필 데이터를 사용하여 참조 테이블 작성, 15](#)
- [플랫 파일을 사용하여 참조 테이블 작성, 18](#)
- [데이터베이스 테이블을 사용하여 참조 테이블 작성, 19](#)
- [버전이 지정된 모델 리포지토리에 참조 테이블로 작업, 20](#)
- [참조 테이블 업데이트, 21](#)
- [감사 내역 이벤트, 24](#)
- [참조 테이블 관련 규칙 및 지침, 25](#)

Analyst 도구 참조 테이블 개요

Analyst 도구의 디자인 작업 공간에서 참조 테이블을 작성합니다.

플랫 파일, 모델 리포지토리의 데이터 소스 및 다른 데이터베이스의 테이블을 사용하여 참조 테이블을 작성할 수 있습니다.

프로필 열을 사용하거나 프로필 열의 데이터 중 일부를 사용하여 참조 테이블을 작성할 수 있습니다. 또한 프로 필에서 선택한 열 패턴을 사용하여 참조 테이블을 작성할 수도 있습니다.

참조 테이블을 작성하거나 업데이트할 경우 테이블과 테이블에 포함된 데이터 열에 대한 속성을 구성합니다.

참조 테이블 속성

Analyst 도구에서 참조 테이블 속성을 보고 업데이트할 수 있습니다. 참조 테이블에는 일반 속성과 열 속성이 표시됩니다. 일반 속성에는 참조 테이블 이름, 작성 날짜, 데이터베이스 연결 이름 및 유효한 열 이름이 포함됩니다. 열 속성에는 열 이름, 전체 자릿수 값, 소수 자릿수 값이 포함됩니다.

속성을 읽기 전용 모드로 볼 수 있습니다. 속성을 업데이트하려면 참조 테이블을 편집하거나 체크 아웃합니다.

참조 테이블 일반 속성

일반 속성에는 참조 테이블 개체에 대한 정보가 포함됩니다.

다음 표에는 일반 속성이 설명되어 있습니다.

속성	설명
이름	참조 테이블 이름입니다.
설명	사용자가 참조 테이블에 대해 입력한 모든 설명입니다.
위치	모델 리포지토리의 참조 테이블 개체의 위치입니다.
유효한 열	참조 테이블의 유효한 열의 이름입니다.
작성 날짜	참조 테이블 이름에 대한 작성 날짜 및 시간입니다.
작성자	참조 테이블을 작성한 사용자의 로그인 이름입니다.
마지막으로 수정한 날짜	참조 테이블에 대한 가장 최근 업데이트의 날짜 및 시간입니다.
최종 수정자	가장 최근 업데이트를 수행한 사용자의 로그인 이름입니다.
연결 이름	참조 데이터 값을 저장하는 데이터베이스에 대한 연결 이름입니다.
유형	참조 테이블 유형입니다. 참조 테이블은 관리되거나 관리되지 않을 수 있습니다.

참조 테이블 열 속성

열 속성에는 열 메타데이터에 대한 정보가 포함됩니다.

다음 테이블에는 열 속성이 설명되어 있습니다.

속성	설명
이름	열 이름입니다.
데이터 유형	각 열에 있는 데이터의 데이터 유형입니다. 다음 데이터 유형 중 하나를 선택할 수 있습니다. <ul style="list-style-type: none">- bigint- 날짜/시간- 10진수- 정수- 문자열 참고: 데이터 유형에 전체 자릿수 및 소수 자릿수 값이 지정되어 있지 않으면 숫자 데이터 유형을 사용하는 Oracle 테이블에서 관리되는 참조 테이블을 생성할 수 없습니다.
전체 자릿수	각 열에 대한 전체 자릿수입니다. 전체 자릿수는 열에서 허용되는 최대 자릿수 또는 최대 문자 수입니다. 구성하는 전체 자릿수 값은 데이터 유형에 따라 달라집니다.
소수 자릿수	각 열에 대한 소수 자릿수입니다. 소수 자릿수는 열에서 소수점 오른쪽에 허용되는 최대 자릿수입니다. 10진수 열에 적용됩니다. 구성하는 소수 자릿수 값은 데이터 유형에 따라 달라집니다.

속성	설명
설명	각 열에 대한 선택적 설명입니다.
Null 가능	열이 null 값을 포함할 수 있는지 여부를 나타냅니다.
키	데이터베이스 소스에서 가져오는 테이블의 기본 키 열을 식별합니다. 관리되지 않는 참조 테이블을 업데이트하는 경우, 소스 데이터베이스의 기본 키 열에서 생성된 열에는 중복 값을 추가할 수 없습니다. 관리되는 참조 테이블의 경우에는 모든 열에 중복 값을 추가할 수 있습니다.

참조 테이블 편집기에서 참조 테이블 작성

참조 테이블 편집기에서 테이블 구조를 정의하고 데이터를 참조 테이블에 추가합니다.

1. **새로 만들기 > 참조 테이블**을 클릭합니다.
새 **참조 테이블** 마법사가 열립니다.
2. **참조 테이블 편집기 사용** 옵션을 선택하고 **다음**을 클릭합니다.
3. **새 열 추가** 옵션을 사용하여 열을 테이블에 추가합니다.
4. 각 열에 대한 속성을 구성합니다.
이 속성에는 열 이름, 데이터 유형, 전체 자릿수 및 소수 자릿수가 포함됩니다.
열에 변환이 참조 데이터 검색에서 반환할 수 있는 데이터가 포함된 경우 올바른 옵션을 선택합니다.
5. 필요에 따라 낮은 수준의 설명을 참조 테이블의 메타데이터로 포함하기 위한 열을 추가합니다.
6. 필요에 따라 테이블에 대한 감사 참고를 입력합니다.
감사 참고는 감사 내역 로그에 나타납니다.
7. **다음**을 클릭합니다.
8. 참조 테이블의 이름을 입력하고 모델 리포지토리의 참조 테이블 개체의 위치를 선택합니다.
9. **마침**을 클릭합니다.

프로필 데이터를 사용하여 참조 테이블 작성

프로필 데이터를 사용하여 프로필의 소스 데이터와 관련되는 참조 테이블을 작성할 수 있습니다. 참조 테이블을 사용하여 소스 데이터에서 여러 가지 유형의 정보를 찾을 수 있습니다.

다음과 같은 방식으로 프로필을 사용하여 참조 테이블을 작성하거나 업데이트할 수 있습니다.

- 프로필에서 열을 선택하고 참조 테이블에 추가합니다.
- 프로필 열을 찾아 열 데이터의 일부를 참조 테이블에 추가합니다.
- 프로필에서 열을 선택하고 해당 열의 패턴 값을 참조 테이블에 추가합니다.

프로필 열 데이터에서 참조 테이블 작성

프로필 데이터 열의 하나 이상의 값에서 참조 테이블을 작성할 수 있습니다. 프로필의 열을 선택하고 참조 테이블에 추가할 열 값을 선택합니다.

1. Analyst 도구에서 **라이브러리** 작업 공간을 엽니다.
2. **프로필** 자산 범주를 선택합니다.
라이브러리는 모델 리포지토리의 프로필 목록을 표시합니다.
3. 참조 테이블에 추가할 열이 포함된 프로필을 엽니다.
이 프로필 개요에는 프로필 열 이름이 나열됩니다.
4. 열 데이터를 검토합니다.
열 데이터를 보려면 열 이름을 클릭합니다.
5. 상세 프로필 보기에서 참조 테이블에 추가할 데이터 값을 선택합니다. 값을 하나씩 선택하거나 모두 선택할 수 있습니다.
6. 열 이름을 마우스 오른쪽 단추로 클릭하고 **참조 테이블에 추가**를 선택합니다.
다음 이미지는 상세 프로필 보기의 데이터 열을 보여 줍니다.

COLUMN2

Values

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

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

Drilldown
Add to Reference Table ← ①
Create Value Frequency Rule
Create Data domain

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

숫자 1은 이미지의 **참조 테이블에 추가** 옵션을 식별합니다.

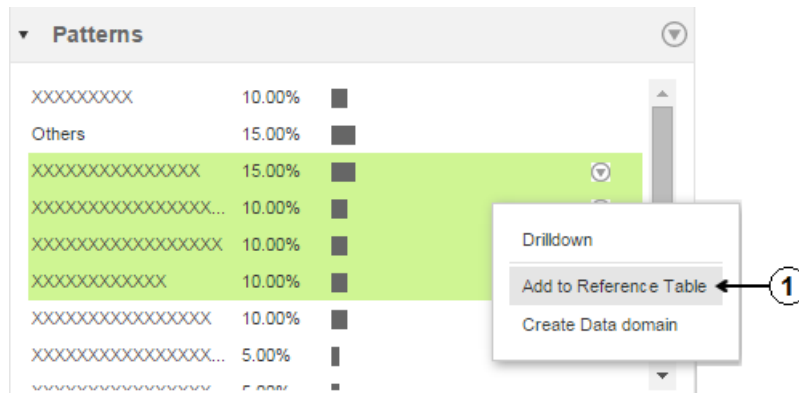
7. **참조 테이블에 추가** 마법사가 열립니다.
참조 테이블 작성 옵션을 선택합니다.
참고: 데이터를 현재 참조 테이블에 추가하기 위한 옵션을 선택할 수도 있습니다.
8. **다음**을 클릭합니다.
기본적으로 열 이름이 참조 테이블 이름으로 표시됩니다. 필요에 따라 이름을 업데이트합니다.
9. 필요에 따라 설명 및 기본값을 입력합니다.
Analyst 도구는 값이 없는 테이블 레코드에 기본값을 사용합니다.
10. **다음**을 클릭합니다.
11. 열 속성을 확인합니다.
필요에 따라 낮은 수준의 설명 메타데이터에 대한 열을 작성하도록 선택합니다.
12. **다음**을 클릭합니다.

13. 참조 테이블 이름 및 설명을 검토합니다.
필요에 따라 감사 참고를 입력합니다.
14. 참조 테이블 개체의 모델 리포지토리 위치를 선택합니다.
15. **마침**을 클릭합니다.

값 패턴에서 참조 테이블 작성

프로필 열의 열 패턴을 사용하여 참조 테이블을 작성할 수 있습니다. 이 패턴은 하나 이상의 열 필드의 데이터 값 구성을 나타냅니다. 프로필에서 열을 선택하고 작성한 참조 테이블에 추가할 패턴을 선택합니다.

1. **Analyst** 도구에서 **라이브러리** 작업 공간을 엽니다.
2. **프로필** 자산 범주를 선택합니다.
라이브러리는 모델 리포지토리의 프로필 목록을 표시합니다.
3. 참조 테이블에 추가할 값 패턴이 포함된 프로필을 엽니다.
이 프로필 개요에는 프로필 열 이름이 나열됩니다.
4. 참조 테이블에 추가할 패턴 데이터를 정의하는 열을 선택합니다.
5. 열 데이터 패턴을 검토합니다.
열 데이터를 보려면 열 이름을 클릭합니다.
6. 상세 프로필 보기에서 추가할 열 패턴을 선택합니다.
7. 선택한 패턴을 마우스 오른쪽 단추로 클릭하고 **참조 테이블에 추가**를 선택합니다.
다음 이미지는 상세 프로필 보기의 열에 대한 데이터 패턴을 보여 줍니다.



숫자 1은 이미지의 **참조 테이블에 추가** 옵션을 식별합니다.

8. **참조 테이블에 추가** 마법사가 열립니다.
참조 테이블 작성 옵션을 선택합니다.
참고: 데이터를 현재 참조 테이블에 추가하기 위한 옵션을 선택할 수도 있습니다.
9. **다음**을 클릭합니다.
기본적으로 열 이름이 참조 테이블 이름으로 표시됩니다. 필요에 따라 이름을 업데이트합니다.
10. 필요에 따라 설명 및 기본값을 입력합니다.
Analyst 도구는 값이 없는 테이블 레코드에 기본값을 사용합니다.
11. **다음**을 클릭합니다.
12. 열 속성을 확인합니다.

필요에 따라 낮은 수준의 설명 메타데이터에 대한 열을 작성하도록 선택합니다.

13. **다음**을 클릭합니다.
14. 참조 테이블 이름 및 설명을 검토합니다.
필요에 따라 감사 참고를 입력합니다.
15. 참조 테이블 개체의 모델 리포지토리 위치를 선택합니다.
16. **마침**을 클릭합니다.

플랫 파일을 사용하여 참조 테이블 작성

CSV 파일에서 참조 데이터를 가져올 수 있습니다. **새 참조 테이블** 마법사를 사용하여 파일 데이터를 가져옵니다.

참조 테이블을 작성하는 데 사용하는 각 플랫 파일에 대해 속성을 구성해야 합니다.

Analyst 도구 플랫 파일 속성

플랫 파일을 참조 테이블로 가져오는 경우 파일의 각 열에 대한 속성을 구성해야 합니다. 구성하는 옵션에 따라 Analyst 도구가 파일에서 데이터를 읽는 방식이 결정됩니다.

다음 표에는 참조 테이블에 사용할 파일 데이터를 가져올 때 구성할 수 있는 속성이 설명되어 있습니다.

속성	설명
구분자	데이터의 열을 구분하는 데 사용되는 문자입니다. 다른 구분자를 입력하려면 기타 필드를 사용합니다. 구분자는 인쇄 가능한 문자여야 하고 따옴표 문자와 이스케이프 문자를 선택한 경우 이러한 문자와 달라야 합니다. 인쇄되지 않는 다중 바이트 문자를 구분자로 선택할 수 없습니다.
텍스트 한정자	텍스트 문자열의 경계를 정의하는 따옴표 문자입니다. 따옴표 사용 안 함, 작은따옴표 또는 큰따옴표를 선택합니다. 따옴표 문자를 선택하는 경우 마법사가 따옴표 쌍 안의 구분자를 무시합니다.
열 이름	첫 줄에서 열 이름을 가져옵니다. 열 이름이 첫 번째 행에 표시되는 경우 이 옵션을 선택합니다. 마법사가 미리보기에서 첫 번째 행의 데이터를 열 이름으로 사용합니다. 기본적으로 활성화되지 않습니다.
값	행에서 값 가져오기를 시작하는 옵션입니다. 마법사가 파일을 가져올 때 읽기 시작하는 미리보기의 행 번호를 나타냅니다.

플랫 파일을 사용하여 참조 테이블 작성

플랫 파일을 사용하여 참조 테이블 데이터를 작성하는 경우 테이블에서 파일의 열 구조를 사용하고 파일 데이터를 가져옵니다.

1. **새로 만들기 > 참조 테이블**을 클릭합니다.
새 참조 테이블 마법사가 나타납니다.
2. **플랫 파일 가져오기** 옵션을 선택합니다.

3. 다음을 클릭합니다.
4. **파일 선택**을 클릭하여 플랫폼 파일을 선택합니다.
5. 플랫폼 파일의 데이터와 일치하는 코드 페이지를 선택합니다.
6. **업로드**를 클릭하여 파일 데이터를 업로드합니다.
7. 다음을 클릭합니다.
8. 플랫폼 파일 속성을 구성합니다.
이 속성은 파일이 사용하는 구분자를 식별하고 파일의 첫 번째 행에 열 이름이 포함되어 있는지 식별합니다.
9. 구성된 속성을 미리 보려면 **미리보기** 창을 새로 고칩니다.
10. 다음을 클릭합니다.
11. 각 열에 대한 속성을 구성합니다.
이 속성에는 열 이름, 데이터 유형, 전체 자릿수 및 소수 자릿수가 포함됩니다.
열에 변환이 참조 데이터 검색에서 반환할 수 있는 데이터가 포함된 경우 올바른 옵션을 선택합니다.
12. 필요에 따라 낮은 수준의 설명을 참조 테이블의 메타데이터로 포함하기 위한 열을 추가합니다.
13. 필요에 따라 테이블에 대한 감사 참고를 입력합니다.
감사 참고는 감사 내역 로그에 나타납니다.
14. 다음을 클릭합니다.
15. 참조 테이블의 이름을 입력하고 모델 리포지토리의 참조 테이블 개체의 위치를 선택합니다.
16. 필요에 따라 테이블의 설명을 입력합니다.
17. **마침**을 클릭합니다.

데이터베이스 테이블을 사용하여 참조 테이블 작성

데이터베이스 테이블에서 참조 테이블을 작성하는 경우 모델 리포지토리에 메타데이터 개체를 작성합니다. 필요한 경우 테이블 데이터를 참조 데이터 웨어하우스로 가져옵니다.

관리되는 참조 테이블을 작성하는 경우 열 데이터를 참조 데이터 웨어하우스로 가져옵니다. 관리되지 않는 참조 테이블을 작성하는 경우 열 데이터를 저장하는 데이터베이스 테이블을 식별합니다. 관리되는 참조 테이블은 ODBC 연결 또는 원시 연결에서 작성할 수 있습니다. 관리되지 않는 참조 테이블은 원시 연결에서 작성할 수 있습니다.

참조 테이블을 작성하기 전에 참조 데이터가 포함된 데이터베이스에 대한 연결이 Informatica 도메인에 있는지 확인해야 합니다. 도메인에 데이터베이스에 대한 연결이 없을 경우 Analyst 도구에서 연결을 정의합니다.

데이터베이스 연결을 정의하려면 **관리 > 연결**을 클릭합니다.

데이터베이스 테이블을 사용하여 참조 테이블 작성

참조 테이블을 작성하려면 데이터베이스에 연결하고 참조 데이터가 포함된 테이블을 선택합니다.

1. **새로 만들기 > 참조 테이블**을 선택합니다.

새 참조 테이블 마법사가 나타납니다.

2. **관계형 테이블에 연결** 옵션을 선택합니다.

참조 데이터 웨어하우스에 데이터를 저장하지 않는 참조 테이블을 작성하려면 **관리되지 않는 테이블**을 선택합니다.

관리되지 않는 참조 테이블을 사용자가 편집할 수 있도록 하려면 **편집 가능** 옵션을 선택합니다.

다음을 클릭합니다.

3. 연결 목록에서 데이터베이스 연결을 선택합니다.

다음을 클릭합니다.

4. **테이블** 패널에서 테이블을 선택합니다.

5. **속성** 패널에서 테이블 속성을 검토합니다.

필요한 경우 **데이터 미리보기**를 클릭하여 테이블 데이터를 봅니다.

다음을 클릭합니다.

6. **열 특성** 패널에서 올바른 열을 선택합니다.

관리되는 참조 테이블을 작성하는 경우 **열 특성** 패널에서 다음과 같은 작업을 수행할 수 있습니다.

- 참조 테이블 열 이름을 편집합니다.
- 행 수준 설명에 대한 메타데이터 열을 추가합니다.

7. 필요에 따라 낮은 수준의 설명을 참조 테이블의 메타데이터로 포함하기 위한 열을 추가합니다.

8. 필요에 따라 테이블에 대한 감사 참고를 입력합니다.

감사 참고는 감사 내역 로그에 나타납니다.

9. **다음을** 클릭합니다.

10. 참조 테이블의 이름을 입력하고 모델 리포지토리의 참조 테이블 개체의 위치를 선택합니다.

11. 필요에 따라 참조 테이블의 설명을 입력합니다.

12. **마침**을 클릭합니다.

버전이 지정된 모델 리포지토리에서 참조 테이블로 작업

참조 테이블을 읽기 전용 모드로 엽니다. 참조 테이블에 대해 작업하려면 편집 모드로 전환하거나 모델 리포지토리에서 참조 테이블을 체크 아웃해야 합니다.

1. Informatica 도구 모음에서 **열기**를 클릭합니다.

자산 라이브러리가 열립니다.

2. **참조 테이블** 자산 범주를 선택하고 참조 테이블 이름을 선택합니다.

참조 테이블이 읽기 전용 모드로 열립니다.

3. 참조 테이블의 현재 버전을 편집하려면 **편집**을 클릭합니다.

버전이 지정된 모델 리포지토리에서 참조 테이블을 편집하려면 참조 테이블을 체크 아웃합니다.

4. 참조 테이블에 대한 작업을 완료한 경우 **마침**을 클릭합니다. Analyst 도구가 참조 테이블에 대한 변경 내용을 저장합니다.

버전이 지정된 모델 리포지토리에서 참조 테이블을 체크 아웃한 경우 개체를 체크 인합니다. 버전이 지정된 모델 리포지토리는 사용자가 개체를 체크 인할 때까지 참조 테이블 버전을 업데이트하지 않습니다.

참조 테이블 업데이트

참조 테이블에 포함된 비즈니스 데이터는 시간이 지나면서 변경될 수 있습니다. 참조 테이블의 데이터와 메타데이터를 검토하고 업데이트하여 테이블에 정확한 정보가 포함되도록 합니다. **Analyst** 도구에서 참조 테이블을 업데이트합니다. 관리되는 참조 테이블과 관리되지 않는 참조 테이블의 데이터와 메타데이터를 업데이트할 수 있습니다.

참조 테이블 데이터와 메타데이터에 대해 다음과 같은 작업을 수행할 수 있습니다.

열 관리

열을 추가 및 삭제하고 열 속성을 편집할 수 있습니다.

행 관리

데이터 행을 참조 테이블에 추가할 수 있습니다.

참조 데이터 값 편집

참조 데이터 값을 편집할 수 있습니다.

데이터 값 바꾸기

찾기 및 교체 옵션을 사용하여 더 이상 정확하지 않거나 조직과 관련되지 않은 데이터 값을 바꿉니다. 열에서 값을 찾아 다른 값으로 바꿀 수 있습니다. 열의 모든 값을 단일 값으로 바꿀 수 있습니다.

참조 테이블 내보내기

참조 테이블을 쉼표로 구분된 값(CSV) 파일, 사전 파일 또는 Excel 파일로 내보냅니다.

관리되지 않는 테이블에 대한 편집 활성화 또는 비활성화

관리되지 않는 참조 테이블을 업데이트하여 테이블 데이터 및 메타데이터에 대한 편집을 활성화 또는 비활성화합니다.

참조 테이블 데이터 새로 고침

참조 테이블 데이터를 **Analyst** 도구에 다시 로드하여 데이터의 최신 변경 사항을 표시합니다.

열 관리

열을 참조 테이블에 추가하고 열 속성을 업데이트할 수 있습니다. 관리되지 않는 참조 테이블의 편집 가능 상태를 업데이트할 수도 있습니다.

1. **열기**를 클릭합니다.
자산 라이브러리가 열립니다.
2. **참조 테이블** 자산 범주를 선택하고 참조 테이블 이름을 선택합니다.
참조 테이블이 읽기 전용 모드로 열립니다.
3. 참조 테이블의 현재 버전을 편집하려면 **편집**을 클릭합니다.
버전이 지정된 모델 리포지토리에서 참조 테이블을 편집하려면 참조 테이블을 체크 아웃합니다.
4. **작업** 메뉴를 열고 **열 속성 변경**을 선택합니다.
열 속성 변경 대화 상자가 열립니다. 대화 상자 옵션을 사용하여 다음과 같은 작업을 수행합니다.
 - 열을 추가합니다.
 - 테이블에서 유효한 열을 변경합니다.
 - 열 이름을 변경합니다.
 - 열에 대한 설명 텍스트를 업데이트합니다.

- 관리되지 않는 참조 테이블의 편집 가능 상태를 업데이트합니다.
 - 테이블에 대한 감사 참고를 업데이트합니다.
5. 작업을 완료하면 **확인**을 클릭합니다.

행 관리

참조 테이블에서 행을 추가, 편집 또는 삭제할 수 있습니다.

1. **열기**를 클릭합니다.
자산 라이브러리가 열립니다.
2. **참조 테이블** 자산 범주를 선택하고 참조 테이블 이름을 선택합니다.
참조 테이블이 읽기 전용 모드로 열립니다.
3. 참조 테이블의 현재 버전을 편집하려면 **편집**을 클릭합니다.
버전이 지정된 모델 리포지토리에 참조 테이블을 편집하려면 참조 테이블을 체크 아웃합니다.
4. 데이터 행을 편집합니다. 다음과 같은 방식으로 데이터 행을 편집할 수 있습니다.
 - 행을 추가하려면 **작업 > 행 추가**를 선택합니다.
행 추가 대화 상자에서 유효한 열과 하나 이상의 다른 열에 값을 입력합니다. 필요에 따라 감사 참고를 입력합니다.
확인을 클릭하여 행을 추가합니다.
 - 단일 데이터 값을 업데이트하려면 값을 클릭하고 데이터를 업데이트합니다.
데이터를 업데이트한 후 행 수준 옵션을 사용하여 데이터를 허용하거나 거부합니다. 데이터를 데이터 행에 직접 입력하는 경우 감사 참고를 입력할 수 없습니다.
 - 행에서 데이터 값을 업데이트하려면 **작업 > 행 편집**을 선택합니다.
행 편집 대화 상자에서 하나 이상의 열에 값을 입력합니다. 필요에 따라 감사 참고를 입력합니다.
적용을 클릭하여 선택한 열에서 데이터를 업데이트합니다.
 - 여러 행에서 값을 업데이트하려면 편집할 행을 선택하고 **작업 > 행 편집**을 선택합니다.
여러 행 편집 대화 상자에서 하나 이상의 열에 값을 입력합니다. 필요에 따라 감사 참고를 입력합니다.
확인을 클릭하여 선택한 열에서 데이터를 업데이트합니다.
 - 행을 삭제하려면 삭제할 행을 선택하고 **작업 > 삭제**를 클릭합니다.
행 삭제 대화 상자에서 필요에 따라 감사 참고를 입력합니다.
확인을 클릭하여 행을 삭제합니다.

참고: Developer tool을 사용하여 큰 참조 테이블의 행 데이터를 편집할 수 있습니다. 예를 들어 참조 테이블의 행이 500개보다 많을 경우 Developer tool에서 테이블을 편집합니다.

값 찾기 및 바꾸기

참조 테이블의 데이터 값을 찾아 바꿀 수 있습니다. 테이블에서 업데이트해야 하는 데이터 값 인스턴스가 하나 이상일 경우 찾기 및 교체 옵션을 사용합니다.

1. **열기**를 클릭합니다.
자산 라이브러리가 열립니다.
2. **참조 테이블** 자산 범주를 선택하고 참조 테이블 이름을 선택합니다.
참조 테이블이 읽기 전용 모드로 열립니다.

3. 참조 테이블의 현재 버전을 편집하려면 **편집**을 클릭합니다.
버전이 지정된 모델 리포지토리에서 참조 테이블을 편집하려면 참조 테이블을 체크 아웃합니다.
4. **작업 > 찾기 및 교체**를 클릭합니다.
찾기 및 교체 도구 모음이 나타납니다.
5. 도구 모음에서 검색 조건을 입력합니다.
 - **찾기** 필드에 데이터 값을 입력합니다.
 - 검색할 열을 선택합니다. 기본적으로 이 작업에서는 모든 열을 검색합니다.
 - **교체할 내용** 필드에 데이터 값을 입력합니다.
6. 다음 옵션을 사용하여 값을 하나씩 바꾸거나 모든 값을 바꿉니다.
 - **다음 및 이전** 옵션을 사용하여 값을 하나씩 찾습니다.
 - 값을 바꾸려면 **바꾸기**를 선택합니다.
 - 값의 모든 인스턴스를 표시하려면 **모두 강조 표시**를 선택합니다.
 - 값의 모든 인스턴스를 바꾸려면 **모두 바꾸기**를 선택합니다.

참조 테이블 데이터 내보내기

참조 테이블의 데이터를 쉼표로 구분된 파일, 사전 파일 또는 Microsoft Excel 파일로 내보냅니다. 데이터를 읽기 전용 모드로 내보낼 수 있습니다.

1. **열기**를 클릭합니다.
자산 라이브러리가 열립니다.
2. **참조 테이블** 자산 범주를 선택하고 참조 테이블 이름을 선택합니다.
참조 테이블이 읽기 전용 모드로 열립니다.
3. **작업 > 데이터 내보내기**를 클릭합니다.
파일로 데이터 내보내기 대화 상자가 열립니다.
다음 표에는 해당 대화 상자 옵션이 설명되어 있습니다.

옵션	설명
파일 이름	데이터를 포함할 파일의 이름입니다. 내보내기 작업은 파일을 작성합니다.
파일 형식	데이터를 포함할 파일의 형식입니다. 다음 형식 중 하나를 선택합니다. <ul style="list-style-type: none"> • csv. 쉼표로 구분된 파일입니다. 기본 형식입니다. • xls. Microsoft Excel 파일입니다. • dic. Informatica 사전 파일입니다.
필드 이름을 첫 번째 행으로 내보내기	열 이름 옵션입니다. 파일의 첫 번째 행에 열 이름을 포함하도록 지정하려면 이 옵션을 선택합니다.
코드 페이지	참조 데이터의 코드 페이지입니다. 기본 코드 페이지는 UTF-8입니다.

4. **확인**을 클릭하여 파일을 내보냅니다.

관리되지 않는 참조 테이블에서 편집 활성화 및 비활성화

관리되지 않는 참조 테이블에서 데이터 값 및 열에 대한 업데이트를 활성화하거나 비활성화할 수 있습니다.

참조 테이블의 편집 가능 상태를 변경하기 전에 테이블을 저장합니다.

1. **열기**를 클릭합니다.

자산 라이브러리가 열립니다.

2. **참조 테이블** 자산 범주를 선택하고 참조 테이블 이름을 선택합니다.

참조 테이블이 읽기 전용 모드로 열립니다.

3. 참조 테이블의 현재 버전을 편집하려면 **편집**을 클릭합니다.

버전이 지정된 모델 리포지토리에서 참조 테이블을 편집하려면 참조 테이블을 체크 아웃합니다.

4. **작업** 메뉴를 열고 **열 속성 변경**을 선택합니다.

열 속성 변경 대화 상자가 열립니다.

5. **편집 가능** 옵션을 선택하거나 지웁니다.

참고: 다음 조건은 사용자 업데이트를 허용하는 관리되지 않는 참조 테이블에 적용됩니다.

- 참조 테이블은 문자열 및 숫자와 같은 단순 데이터 유형을 사용해야 합니다.
- 참조 테이블 메타데이터에 대한 제약 조건을 정의하거나 모든 열의 기본값을 지정하지 마십시오.

참조 테이블 값 새로 고침

Analyst 도구가 참조 테이블에 대해 표시하는 값을 새로 고쳐야 할 수 있습니다.

참조 테이블 값을 다시 로드하려면 **작업 > 새로 고침**을 클릭합니다. **Analyst** 도구가 데이터베이스에서 데이터 값의 현재 버전을 검색합니다.

감사 내역 이벤트

사용자가 참조 테이블에서 변경한 내용의 감사 내역을 확인할 수 있습니다. 참조 테이블에서 감사 내역 보기를 사용하여 감사 내역 이벤트를 확인합니다. **Analyst** 도구가 표시하는 감사 내역 이벤트를 필터링할 수 있습니다.

다음 테이블에는 지정할 수 있는 필터 옵션이 설명되어 있습니다.

옵션	설명
날짜	표시할 작업에 대한 시작 및 종료 날짜입니다. 달력 옵션을 사용하여 날짜를 설정합니다.
유형	감사 내역 이벤트의 유형입니다. 다음 이벤트 유형을 볼 수 있습니다. <ul style="list-style-type: none">- 데이터. 참조 테이블의 데이터 값과 관련된 이벤트입니다. 이벤트에는 행 추가, 행 삭제 및 행 업데이트 작업이 포함됩니다.- 메타데이터. 참조 테이블 메타데이터와 관련된 이벤트입니다. 이벤트에는 참조 테이블을 작성하고 열을 추가하거나 삭제하고 참조 테이블을 체크 인하는 작업이 포함됩니다. 참고: 데이터 및 메타데이터 이벤트를 동시에 볼 수 없습니다.

옵션	설명
사용자	참조 테이블을 편집한 사용자입니다. 사용자의 전체 이름 및 로그인 이름이 필터에 표시됩니다.
상태	감사 내역 로그 이벤트의 상태입니다. 상태는 참조 테이블 편집기에서 수행한 작업에 해당합니다. 예를 들어 상태는 사용자가 참조 테이블을 작성했거나 행을 추가했음을 나타낼 수 있습니다.

감사 내역 로그 이벤트에는 사용자가 삽입, 업데이트 또는 삭제한 열 값과 감사 내역 설명도 포함됩니다.

감사 내역 이벤트 보기

사용자가 참조 테이블에 대해 수행한 업데이트에 대해 알아보려면 감사 내역 이벤트를 확인합니다. 읽기 전용 모드로 감사 내역 이벤트를 확인할 수 있습니다.

1. 열기를 클릭합니다.
자산 라이브러리가 열립니다.
2. 참조 테이블 자산 범주를 선택하고 참조 테이블 이름을 선택합니다.
참조 테이블이 읽기 전용 모드로 열립니다.
3. 감사 내역을 클릭합니다.
4. 필터 옵션을 구성합니다.
업데이트 날짜, 업데이트 유형, 업데이트 상태, 업데이트를 수행한 사용자의 이름으로 필터링할 수 있습니다.
5. 표시를 클릭합니다.
로그 이벤트가 지정한 필터 옵션에 대해 표시됩니다.

참조 테이블 관련 규칙 및 지침

다음 규칙과 지침에 따라 Analyst 도구에서 참조 테이블로 작업합니다.

- Oracle, IBM DB2 또는 Microsoft SQL Server 데이터베이스에서 참조 테이블을 가져올 경우 테이블, 뷰, 스키마, 동의어 또는 열 이름에 대/소문자가 혼합되어 있으면 Analyst 도구에서 미리보기를 표시할 수 없습니다.
대/소문자를 구분하는 데이터베이스에 있는 테이블의 데이터를 미리 보려면 데이터베이스 연결의 대/소문자가 혼합된 식별자 지원 특성을 true로 설정합니다.
- 한 형식의 유추된 열 패턴을 사용하여 참조 테이블을 작성할 경우 Analyst 도구가 참조 테이블을 다른 형식의 열 패턴으로 채웁니다.
예를 들어 열 패턴 X(5)에 대한 참조 테이블을 작성할 경우 Analyst 도구가 참조 테이블에서 열 패턴에 대해 다음과 같은 형식을 표시합니다. XXXXX.
- Oracle 데이터베이스를 가져올 경우 테이블에 있는 VARCHAR2 열의 길이를 확인합니다. Analyst 도구는 길이가 1000보다 큰 VARCHAR2 열이 포함된 Oracle 데이터베이스 테이블을 가져올 수 없습니다.
- 참조 테이블을 읽으려면 테이블 데이터 값을 저장하는 데이터베이스와의 연결에서 실행 권한이 있어야 합니다. 예를 들어 참조 데이터 웨어하우스에서 데이터 값을 저장하는 경우 참조 데이터 웨어하우스와의 연결에서

실행 권한이 있어야 합니다. 읽기 또는 쓰기 모드로 참조 테이블에 액세스하려면 실행 권한이 필요합니다. 데이터베이스 연결 사용 권한은 데이터베이스의 모든 참조 데이터에 적용됩니다.

- 참조 테이블을 지정하는 변환과 함께 매핑을 실행할 때 매핑은 모델 리포지토리에서 최신 버전의 참조 테이블을 사용합니다. 변환을 구성할 때 기록 버전의 참조 테이블을 선택할 수 없습니다.

다른 사용자가 동시 **Developer tool** 세션에서 참조 테이블을 이전 버전으로 복원하는 경우 참조 테이블 버전이 더 이상 세션 전체에 걸쳐 동일하지 않습니다. 참조 테이블을 사용하는 매핑을 구성 및 실행하는 경우 현재 세션이 현재 참조 테이블 버전을 식별하지 않기 때문에 매핑이 실패할 수 있습니다. 매핑이 현재 참조 테이블을 사용하도록 하려면 매핑을 실행하기 전에 모델 리포지토리를 새로 고칩니다.

- 편집을 허용하도록 관리되지 않는 참조 테이블을 구성하는 경우 참조 테이블에서 문자열 및 숫자와 같은 단순한 데이터 유형을 사용하는지 확인합니다. 또한 참조 테이블에 참조 테이블 메타데이터에 대한 제약 조건이 정의되지 않았는지 또는 열에 기본값을 사용하는지 확인합니다.

제 3 장

Developer 도구의 참조 데이터

이 장에 포함된 항목:

- [Developer tool 참조 데이터 개요, 27](#)
- [참조 데이터와 변환, 27](#)
- [버전이 지정된 모델 리포지토리에서 참조 데이터 개체로 작업, 28](#)
- [참조 테이블, 29](#)
- [콘텐츠 집합, 33](#)

Developer tool 참조 데이터 개요

Developer tool에서 참조 데이터 개체에 대한 구성 속성을 작성하고 업데이트하고 확인할 수 있습니다.

Developer tool을 사용하여 다음과 같은 유형의 개체를 작성하고 업데이트합니다.

참조 테이블

참조 테이블에는 표준 버전과 대체 버전의 데이터 값 집합이 포함됩니다. Developer tool에서 참조 테이블을 변환에 추가하여 소스 데이터 값이 정확하고 올바른 형식인지 확인할 수 있습니다.

콘텐츠 집합

콘텐츠 집합은 리포지토리 또는 파일에 참조 데이터 값을 지정하는 모델 리포지토리 개체입니다. 콘텐츠 집합에는 데이터 품질 변환에서 검색 작업을 수행하는 데 사용할 수 있는 다양한 유형의 참조 데이터가 포함됩니다.

또한 Developer tool에서 주소 참조 데이터 파일 및 ID 채우기 파일로 작업할 수도 있습니다. 주소 유효성 검사기 변환을 구성하는 경우 주소 참조 데이터 파일을 선택합니다. ID 일치 분석을 위한 일치 변환을 구성하는 경우 ID 채우기 파일을 선택합니다.

참조 데이터와 변환

여러 가지 변환에서 참조 데이터를 읽고 데이터 품질 작업을 수행합니다.

다음 변환에서 참조 데이터를 읽을 수 있습니다.

- 주소 유효성 검사기. 주소 참조 데이터를 읽고 주소의 정확성을 확인합니다.
- 대/소문자 변환기. 참조 데이터 테이블을 읽고 대/소문자를 변경해야 하는 문자열을 식별합니다.

- 분류자. 콘텐츠 집합 데이터를 읽고 문자열의 정보 유형을 식별합니다.
- 비교. 중복 분석 중에 ID 채우기 데이터를 읽습니다.
- 라벨러. 콘텐츠 집합 데이터를 읽어 레이블 문자열을 식별합니다.
- 일치. 중복 분석 중에 ID 채우기 데이터를 읽습니다.
- 파서. 콘텐츠 집합 데이터를 읽고 포함된 정보를 기반으로 문자열을 구문 분석합니다.
- 표준화. 참조 데이터 테이블을 읽고 문자열을 일반 형식으로 표준화합니다.

Data Quality Content 설치 프로그램 파일 집합에는 사용자가 가져올 수 있는 Informatica 참조 데이터 개체가 있습니다.

버전이 지정된 모델 리포지토리에서 참조 데이터 개체로 작업

버전이 지정된 모델 리포지토리에서 참조 테이블 또는 콘텐츠 집합으로 작업하는 경우 해당 리포지토리가 버전 제어를 개체에 적용할 수 있습니다. 버전 제어를 개체에 적용하려면 사용자가 개체를 모델 리포지토리에 체크 인합니다.

참조 테이블 또는 콘텐츠 집합이 버전 제어를 통해 관리되지 않는 경우 버전 제어 시스템 외부에서 개체를 열고 업데이트할 수 있습니다. 개체를 여는 경우 모델 리포지토리가 다른 사용자가 작업할 수 없도록 개체를 잠급니다.

참조 테이블 또는 콘텐츠 집합이 버전 제어를 통해 관리되는 경우 개체를 읽기 전용 모드로 엽니다. 개체에 대해 작업하려면 모델 리포지토리에서 개체를 체크 아웃합니다. 또는 개체를 체크 아웃한 다음 엽니다. 개체를 체크 인하여 최신 변경 내용이 포함된 개체의 버전을 작성합니다.

참조 데이터 개체 체크 아웃

사용자가 모델 리포지토리에 체크 인한 콘텐츠 집합 또는 참조 테이블에 대해 작업하려면 리포지토리에서 개체를 체크 아웃합니다.

1. Object Explorer에서 참조 테이블 또는 콘텐츠 집합을 찾습니다.
2. 개체 이름을 마우스 오른쪽 단추로 클릭하고 **열기**를 클릭합니다.
개체가 읽기 전용 모드로 열립니다.
3. 개체 이름을 마우스 오른쪽 단추로 클릭하고 **체크 아웃**을 클릭합니다.
개체를 편집할 수 있습니다.

참조 데이터 개체 체크 인

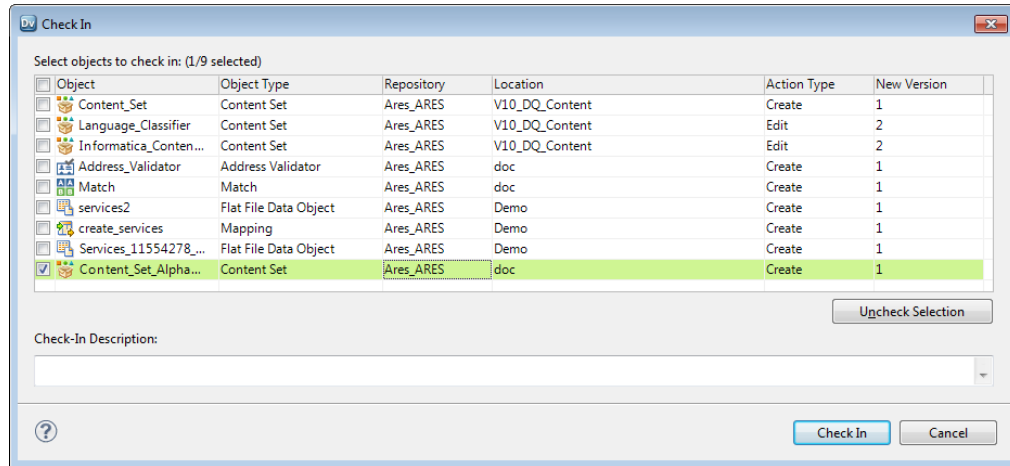
모델 리포지토리에서 체크 아웃한 콘텐츠 집합 또는 참조 테이블에 대한 작업을 완료하는 경우 개체를 체크 인합니다.

현재 체크 아웃한 개체 목록을 보려면 참조 테이블 편집기 아래의 **체크 아웃된 개체** 탭을 엽니다.

1. 참조 테이블 또는 콘텐츠 집합에 대해 변경한 내용을 저장합니다.
2. Object Explorer에서 참조 테이블 또는 콘텐츠 집합을 찾습니다.
3. 개체 이름을 마우스 오른쪽 단추로 클릭하고 **체크 인**을 클릭합니다.

체크 인 대화 상자가 열립니다.

다음 이미지는 해당 대화 상자를 보여 줍니다.



- 리포지토리에 체크 인할 하나 이상의 개체를 선택합니다.

참고: 현재 세션에서 열려 있지 않은 개체를 체크 인할 수 있습니다. 체크 아웃된 상태의 모든 개체를 체크 인할 수 있습니다.

- 필요에 따라 작업의 설명을 입력합니다.
- 체크 인을 클릭합니다.

체크 인 작업은 개체 버전 번호를 업데이트합니다. 개체를 처음 체크 인하는 경우 모델 리포지토리가 개체의 버전 1을 작성합니다.

참조 테이블

Developer tool에서 참조 테이블을 변환에 추가합니다. 입력 데이터에서 참조 테이블 값을 찾고 참조 테이블에서 해당되는 올바른 값을 출력으로 쓰도록 변환을 구성합니다.

Developer tool에서 참조 테이블을 작성하려면 다음 방법 중 하나를 사용합니다.

- 빈 참조 테이블을 작성하고 데이터 값을 입력합니다.
- 플랫 파일의 데이터를 사용하여 참조 테이블을 작성합니다.
- 데이터베이스 테이블, 동의어 또는 보기의 데이터를 사용하여 참조 테이블을 작성합니다.

참조 테이블 데이터 속성

Developer tool에서 참조 테이블 데이터 및 메타데이터에 대한 속성을 볼 수 있습니다. 모델 리포지토리의 참조 테이블을 열면 Developer tool에 속성이 표시됩니다.

참조 테이블에는 일반 속성과 열 속성이 표시됩니다. Developer tool에서 참조 테이블 속성을 볼 수 있습니다. Analyst 도구에서 참조 테이블 속성을 보고 편집할 수 있습니다.

다음 표에는 참조 테이블의 일반 속성이 설명되어 있습니다.

속성	설명
이름	참조 테이블의 이름입니다.
설명	참조 테이블에 대한 선택적 설명입니다.

다음 표에는 참조 테이블의 열 속성이 설명되어 있습니다.

속성	설명
유효함	올바른 참조 데이터가 포함된 열을 식별합니다.
이름	각 열의 이름입니다.
데이터 유형	각 열에 있는 데이터의 데이터 유형입니다.
전체 자릿수	각 열의 전체 자릿수입니다.
소수 자릿수	각 열의 소수 자릿수입니다.
설명	열의 콘텐츠에 대한 설명입니다. 참조 테이블을 작성할 때 필요한 경우 설명을 추가할 수 있습니다.
낮은 수준의 설명에 대한 열을 포함합니다.	참조 테이블에 열 데이터 설명에 대한 열이 있는지를 나타냅니다.
기본값	열에 있는 필드의 기본값입니다. 참조 테이블을 작성할 때 필요한 경우 기본값을 추가할 수 있습니다.
연결 이름	참조 테이블 데이터 값을 포함하는 데이터베이스에 대한 연결의 이름입니다.

참조 테이블 개체 작성

빈 참조 테이블을 작성하고 수동으로 값을 추가하려는 경우 이 옵션을 선택합니다.

1. Developer 도구 메뉴에서 **파일 > 새로 만들기 > 참조 테이블**을 선택합니다.
2. 새 테이블 마법사에서 **빈 항목인 참조 테이블**을 선택합니다.
3. 테이블의 이름을 입력합니다.
4. 테이블 메타데이터를 저장할 프로젝트를 선택합니다.

위치 필드에서 **찾아보기**를 클릭합니다. **위치 선택** 대화 상자가 열리고 리포지토리의 프로젝트가 표시됩니다. 필요한 프로젝트를 선택합니다.

다음을 클릭합니다.

5. 두 개 이상의 열을 테이블에 추가합니다. **새로 만들기** 옵션을 클릭하고 열을 작성합니다.

다음 표에는 각 열의 속성이 설명되어 있습니다.

속성	기본값
이름	열

속성	기본값
데이터 유형	문자열
정밀도	10
배율	0
설명	비어 있음. 선택적 속성.

- 올바른 값이 있는 열을 선택합니다. 작성하는 열의 순서를 바꿀 수 있습니다.
- 다음 표에는 선택적 속성이 설명되어 있습니다.

속성	기본값
행 수준 설명에 대한 열이 포함됩니다.	지워짐
감사 참고	비어 있음
기본값	비어 있음

마침을 클릭합니다.

참조 테이블이 Developer 도구 작업 영역에서 열립니다.

플랫 파일을 사용하여 참조 테이블 작성

플랫 파일에 저장된 데이터를 사용하여 참조 테이블을 작성할 수 있습니다.

- Developer 도구 메뉴에서 **파일 > 새로 만들기 > 참조 테이블**을 선택합니다.
- 새 테이블 마법사에서 **플랫 파일의 참조 테이블**을 선택합니다.
- 테이블의 데이터 소스로 사용할 파일을 찾습니다.
- 테이블의 이름을 입력합니다.
- 테이블 메타데이터를 저장할 프로젝트를 선택합니다.
위치 필드에서 **찾아보기**를 클릭합니다. **위치 선택** 대화 상자가 열리고 리포지토리의 프로젝트가 표시됩니다. 필요한 프로젝트를 선택합니다.

다음을 클릭합니다.

- 코드 페이지로 **UTF-8**을 설정합니다.
- 플랫 파일에서 사용하는 구분자를 지정합니다.
- 플랫 파일에 열 이름이 있을 경우 파일의 첫 번째 행에서 열 이름을 가져오는 옵션을 선택합니다.
- 다음 표에는 선택적 테이블 속성이 설명되어 있습니다.

속성	기본값
텍스트 한정자	따옴표 없음
가져오기 시작 행	행 1
행 구분자	\012 LF (\n)

속성	기본값
연속 구분자를 하나로 처리	지워짐
이스케이프 문자	비어 있음
데이터에 이스케이프 문자 유지	지워짐
미리 볼 최대 행 수	500

다음을 클릭합니다.

10. 올바른 값이 있는 열을 선택합니다.
11. 다음 표에는 선택적 속성이 설명되어 있습니다.

속성	기본값
행 수준 설명에 대한 열이 포함됩니다.	지워짐
감사 참고	비어 있음
기본값	비어 있음
미리 볼 최대 행 수	500

마침을 클릭합니다.

참조 테이블이 Developer 도구 작업 영역에서 열립니다.

관계형 소스에서 참조 테이블 작성

관계형 테이블, 동의어 또는 보기에서 참조 테이블을 작성할 수 있습니다.

관리되는 참조 테이블을 작성하는 경우 열 데이터를 참조 데이터 웨어하우스로 가져옵니다. 관리되지 않는 참조 테이블을 작성하는 경우 열 데이터를 저장하는 데이터베이스 테이블을 식별합니다. 관리되는 참조 테이블은 ODBC 연결 또는 원시 연결에서 작성할 수 있습니다. 관리되지 않는 참조 테이블은 원시 연결에서 작성할 수 있습니다.

참조 테이블을 작성하기 전에 참조 데이터가 포함된 데이터베이스에 대한 연결이 Informatica 도메인에 있는지 확인해야 합니다.

연결 탐색기에서 데이터베이스 연결을 구성할 수 있습니다. Developer tool에서 연결 탐색기가 표시되지 않으면 Developer tool 메뉴에서 **창 > 보기 표시 > 연결 탐색기**를 선택합니다.

관계형 소스에서 참조 테이블 작성

참조 테이블을 작성하려면 데이터베이스에 연결하고 참조 데이터가 포함된 테이블을 선택합니다.

1. Developer tool 메뉴에서 **파일 > 새로 만들기 > 참조 테이블**을 선택합니다.
2. 테이블 작성 마법사에서 **관계형 소스의 참조 테이블**을 선택합니다.

다음을 클릭합니다.

3. 데이터베이스 연결을 선택합니다.

연결 필드에서 **찾아보기**를 클릭합니다. **연결 선택** 대화 상자가 열리고 사용 가능한 데이터베이스 연결이 표시됩니다.

연결을 선택하면 **확인**을 클릭합니다.

4. 데이터베이스 리소스를 선택합니다.

리소스 필드에서 **찾아보기**를 클릭합니다. **리소스 선택** 대화 상자가 열리고 데이터베이스 연결에 대한 리소스가 표시됩니다. 데이터베이스를 탐색하고 데이터베이스 테이블, 동의어 또는 보기를 선택합니다.

필요한 경우 리소스의 항목 정보를 미리 볼 수 있습니다.

5. 테이블의 이름을 입력합니다.

6. 참조 테이블 개체의 위치를 선택합니다.

위치 필드에서 **찾아보기**를 클릭합니다. **위치 선택** 대화 상자가 열리고 리포지토리의 프로젝트가 표시됩니다.

위치를 선택한 후 **다음**을 클릭합니다.

7. 참조 데이터 웨어하우스에 데이터를 저장하지 않는 참조 테이블을 작성하려면 **관리되지 않는 테이블**을 선택합니다.

관리되지 않는 참조 테이블을 사용자가 편집할 수 있도록 하려면 **편집 가능** 옵션을 선택합니다.

다음을 클릭합니다.

8. 올바른 값이 있는 열을 선택합니다.

9. 다음 테이블에는 지정할 수 있는 속성이 설명되어 있습니다.

속성	기본값
행 수준 설명에 대한 열이 포함됩니다.	지워짐
설명	지워짐
기본값	비어 있음
감사 참고	비어 있음
미리 볼 최대 행 수	500

10. **마침**을 클릭합니다.

콘텐츠 집합

콘텐츠 집합은 다른 참조 데이터 개체에 대한 데이터 또는 메타데이터를 저장하는 모델 리포지토리 개체입니다. 콘텐츠 집합에는 문자 집합, 패턴 집합, 토큰 집합, 정규식, 확률 모델 및 분류자 모델이 포함될 수 있습니다. 콘텐츠 집합을 사용하여 단일 프로젝트, 정보 유형 또는 비즈니스 용도와 관련된 참조 데이터 개체를 정의하고 구성할 수 있습니다.

Developer tool에는 모델 리포지토리에 나타나지 않는 시스템 정의의 문자 집합과 토큰 집합이 있습니다. 시스템 정의의 개체를 보고 사용하려면 라벨러 변환, 파서 변환 또는 표준화 변환에서 전략을 구성합니다.

문자 집합

문자 집합에는 특정 문자 및 문자 범위를 식별하는 식이 포함됩니다. 문자 레이블 지정 모드를 사용하는 라벨러 변환에서 문자 집합을 사용할 수 있습니다.

문자 범위는 순차적 문자 코드 범위를 지정합니다. 예를 들어 문자 범위 "[A-C]"는 대문자 "A", "B" 및 "C"와 일치합니다. 이 문자 범위는 소문자 "a", "b" 또는 "c"와 일치하지 않습니다.

문자 집합을 사용하여 레이블 지정 작업의 일부로 특정 문자 또는 문자 범위를 식별할 수 있습니다. 예를 들어 전화 번호가 포함된 열에서 모든 숫자에 레이블을 지정할 수 있습니다. 숫자에 레이블을 지정한 후 파서 변환에서 패턴을 식별하여 문제가 있는 패턴을 별도의 출력 포트에 쓸 수 있습니다.

문자 집합 속성

문자 집합에 대한 문자 레이블 지정 작업을 결정하는 속성을 구성합니다.

다음 표에는 사용자 정의 문자 집합에 대한 속성이 설명되어 있습니다.

속성	설명
레이블	라벨러 변환이 문자 집합과 일치하는 데이터에 적용하는 레이블을 정의합니다.
표준 모드	시작 범위 및 끝 범위의 필드가 포함된 간단한 편집 보기를 활성화합니다.
시작 범위	문자 범위의 첫 번째 문자를 지정합니다.
끝 범위	문자 범위의 마지막 문자를 지정합니다. 단일 문자로 구성된 범위의 경우 이 필드를 비워 둡니다.
고급 모드	범위 문자 및 구분자 문자를 사용하여 문자 범위를 수동으로 입력할 수 있는 고급 편집 보기를 활성화합니다.
범위 문자	문자 범위를 나타내는 기호를 임시로 변경합니다. 문자 집합을 닫으면 범위 문자는 기본 문자로 되돌아갑니다.
구분자 문자	문자 범위를 구분하는 기호를 임시로 변경합니다. 문자 집합을 닫으면 구분자 문자는 기본 문자로 되돌아갑니다.

분류자 모델

분류자 모델은 입력 문자열을 분석하고 문자열에 포함될 가능성이 가장 높은 정보의 유형을 결정합니다. 분류자 변환에서 분류자 모델을 사용합니다.

분류자 모델에는 참조 데이터 행 및 레이블 값이 포함됩니다. 행은 분류자 변환에 연결할 수 있는 포트의 입력 데이터를 나타냅니다. 레이블 값은 데이터 행에 포함되어 있는 정보의 유형을 설명합니다. 분류자 모델을 구성할 때 모델의 각 참조 데이터에 레이블을 할당합니다.

분류자 모델에서 참조 데이터 행을 레이블에 연결하려면 모델을 컴파일합니다. 컴파일 프로세스에서는 데이터 행과 레이블 값 간에 일련의 논리적 연관이 생성됩니다. 모델을 읽는 매핑을 실행하면 데이터 통합 서비스가 모델 논리를 분류기 변환 입력 데이터에 적용합니다. 데이터 통합 서비스는 각 입력 데이터 필드의 정보를 가장 정확히 설명하는 레이블을 반환합니다.

Developer tool에서 분류자 모델을 작성할 수 있습니다. 모델 리포지토리는 분류자 모델 개체를 저장합니다.

Developer tool은 데이터 행, 레이블 및 컴파일 데이터를 Informatica 디렉터리 구조의 파일에 기록합니다.

패턴 집합

패턴 집합에는 토큰 레이블 지정 작업의 출력에서 데이터 패턴을 식별하는 식이 포함됩니다. 패턴 집합을 사용하여 토큰화된 데이터 출력 포트를 분석하고 일치하는 문자열을 하나 이상의 출력 포트에 쓸 수 있습니다. 패턴 구문 분석 모드를 사용하는 파서 변환에서 패턴 집합을 사용합니다.

예를 들어 이름과 이니셜을 식별하는 패턴 집합을 사용하도록 파서 변환을 구성할 수 있습니다. 이 변환에서는 패턴 집합을 사용하여 토큰 레이블 지정 모드에서 라벨러 변환의 출력을 분석합니다. 출력의 이름과 이니셜을 별도의 포트에 쓰도록 파서 변환을 구성할 수 있습니다.

패턴 집합 속성

패턴 집합의 패턴을 결정하는 속성을 구성합니다.

다음 표에는 사용자 정의 패턴 집합에 대한 속성이 설명되어 있습니다.

속성	설명
패턴	패턴 파서가 검색할 패턴을 정의합니다. 하나의 패턴 집합에 대해 여러 개의 패턴을 입력할 수 있습니다. 와일드카드, 문자 및 문자열의 조합을 사용하여 구성된 패턴을 입력할 수 있습니다.

확률 모델

확률 모델은 입력 데이터 값을 분석하고 값에 포함될 가능성이 가장 높은 정보의 유형을 결정합니다. 확률 모델은 라벨러 변환 및 파서 변환에서 사용합니다.

확률 모델에는 참조 데이터 값과 레이블 값이 포함됩니다. 참조 데이터 값은 변환에 연결하는 입력 포트의 데이터를 나타냅니다. 레이블 값은 참조 데이터 값에 포함되는 정보의 유형을 설명합니다. 사용자는 모델의 각 참조 데이터 값에 레이블을 할당합니다.

확률 모델에서 참조 데이터 값을 레이블에 연결하려면 모델을 컴파일합니다. 컴파일 프로세스에서는 데이터 값과 레이블 간의 일련의 논리적 연관이 생성됩니다. 모델을 읽는 매핑을 실행하면 데이터 통합 서비스가 모델 논리를 변환 입력 데이터에 적용합니다. 데이터 통합 서비스는 입력 데이터 값을 가장 정확히 설명하는 레이블을 반환합니다.

Developer tool에서 확률 모델을 작성할 수 있습니다. 모델 리포지토리는 확률 모델 개체를 저장합니다.

Developer tool은 데이터 값, 레이블 및 컴파일 데이터를 Informatica 디렉터리 구조의 파일에 씁니다.

정규식

콘텐츠 집합의 컨텍스트에서 정규식은 구문 분석 및 레이블 지정 작업에서 사용할 수 있는 식입니다. 정규식을 사용하여 입력 데이터에 있는 하나 이상의 문자열을 식별할 수 있습니다. 토큰 구문 분석 모드를 사용하는 파서 변환에서 정규식을 사용할 수 있습니다. 또한 토큰 레이블 지정 모드를 사용하는 라벨러 변환에서도 정규식을 사용할 수 있습니다.

파서 변환에서는 정규식을 사용하여 입력 데이터에서 패턴을 일치시키고 모든 일치하는 문자열을 하나 이상의 출력으로 구문 분석합니다. 예를 들어 정규식을 사용하여 입력 데이터에서 모든 전자 메일 주소를 식별하고 각각의 전자 메일 주소 구성 요소를 서로 다른 출력으로 구문 분석할 수 있습니다.

라벨러 변환에서는 정규식을 사용하여 입력 패턴을 일치시키고 단일 레이블을 작성합니다. 출력이 여러 개인 정규식이 여러 레이블을 생성하지는 않습니다.

정규식 속성

정규식이 출력 문자열을 식별하고 쓰는 방식을 결정하는 속성을 구성합니다.

다음 표에는 사용자 정의 정규식에 대한 속성이 설명되어 있습니다.

속성	설명
출력 수	정규식이 쓰는 출력 포트의 수를 정의합니다.
정규식	파서 변환이 문자열을 일치시키는 데 사용하는 패턴을 정의합니다.
식 테스트	정규식을 테스트하기 위해 입력하는 데이터를 포함합니다. 이 필드에 데이터를 입력하면 필드에서 정규식과 일치하는 문자열이 강조 표시됩니다.
다음 식	정규식과 일치하는 다음 문자열로 이동하고 문자열의 글꼴을 굵게 변경합니다.
이전 식	정규식과 일치하는 이전 문자열로 이동하고 문자열의 글꼴을 굵게 변경합니다.

토큰 집합

토큰 집합에는 특정 토큰을 식별하는 식이 포함됩니다. 토큰 레이블 지정 모드를 사용하는 라벨러 변환에서 토큰 집합을 사용할 수 있습니다. 또한 토큰 구문 분석 모드를 사용하는 파서 변환에서도 토큰 집합을 사용할 수 있습니다.

토큰 집합을 사용하여 레이블 지정 및 구문 분석 작업의 일부로 특정 토큰을 식별합니다. 예를 들어 토큰 집합을 사용하여 "AccountName@DomainName" 형식을 사용하는 모든 전자 메일 주소에 레이블을 지정할 수 있습니다. 토큰에 레이블을 지정하면 파서 변환을 사용하여 전자 메일 주소를 지정하는 출력 포트에 쓸 수 있습니다.

토큰 집합 속성

토큰 집합에 대한 레이블 지정 작업을 결정하는 속성을 구성합니다.

다음 표에는 사용자 정의 문자 집합에 대한 속성이 설명되어 있습니다.

속성	토큰 집합 모드	설명
이름	해당 없음	토큰 집합 이름을 정의합니다.
설명	해당 없음	토큰 집합을 설명합니다.
토큰 집합 옵션	해당 없음	토큰 집합이 정규식 모드 또는 문자 모드를 사용하는지 정의합니다.
레이블	정규식	라벨러 변환이 토큰 집합과 일치하는 데이터에 적용하는 레이블을 정의합니다.
정규식	정규식	라벨러 변환이 문자열을 일치시키는 데 사용하는 패턴을 정의합니다.

속성	토큰 집합 모드	설명
식 테스트	정규식	정규식을 테스트하기 위해 입력하는 데이터를 포함합니다. 이 필드에 데이터를 입력하면 필드에서 정규식과 일치하는 문자열이 강조 표시됩니다.
다음 식	정규식	정규식과 일치하는 다음 문자열로 이동하고 문자열의 글꼴을 굵게 변경합니다.
이전 식	정규식	정규식과 일치하는 이전 문자열로 이동하고 문자열의 글꼴을 굵게 변경합니다.
레이블	문자	라벨러 변환이 문자 집합과 일치하는 데이터에 적용하는 레이블을 정의합니다.
표준 모드	문자	시작 범위 및 끝 범위의 필드가 포함된 간단한 편집 보기를 활성화합니다.
시작 범위	문자	문자 범위의 첫 번째 문자를 지정합니다.
끝 범위	문자	문자 범위의 마지막 문자를 지정합니다. 단일 문자 범위의 경우 이 필드를 비워 둡니다.
고급 모드	문자	범위 문자 및 구분자 문자를 사용하여 문자 범위를 수동으로 입력할 수 있는 고급 편집 보기를 활성화합니다.
범위 문자	문자	문자 범위를 나타내는 기호를 임시로 변경합니다. 문자 집합을 닫으면 범위 문자는 기본 문자로 돌아갑니다.
구분자 문자	문자	문자 범위를 구분하는 기호를 임시로 변경합니다. 문자 집합을 닫으면 구분자 문자는 기본 문자로 돌아갑니다.

확률 모델 및 분류자 모델에 대한 규칙 및 지침

모델 리포지토리에서 각 확률 모델 및 분류자 모델은 Informatica 디렉터리 구조의 파일을 식별합니다. 파일에는 사용자가 Developer tool에서 모델에 추가한 데이터 값 및 레이블이 포함됩니다. 또한 데이터 값과 레이블 간의 연관을 정의하는 컴파일 논리도 포함됩니다.

확률 모델 또는 분류자 모델 작업 시 다음 규칙 및 지침을 고려합니다.

- 모델을 포함하는 매핑을 실행하면 데이터 통합 서비스가 컴파일된 모델 논리를 변환 입력 데이터에 적용합니다. 매핑이 실행될 때 데이터 통합 서비스는 모델의 데이터 값 또는 레이블을 읽지 않습니다.

- 필요한 경우 확률 모델 또는 분류자 모델에서 데이터 값 및 레이블을 제거할 수 있습니다. 예를 들어 중요한 데이터 또는 독점적 데이터를 모델에서 제거하도록 결정할 수 있습니다. **Developer tool**에서 개별 데이터 값 및 레이블을 제거할 수 있습니다. 모델 리포지토리에서 모델을 내보낼 때 모든 데이터 값 및 레이블을 제거할 수 있습니다.

참고: 모델에서 모든 데이터 값 및 레이블을 제거할 경우 모델을 컴파일할 수 없습니다.

- 하나 이상의 데이터 값 또는 레이블을 모델에서 제거하면 컴파일된 모델 논리가 모델 파일에서 현재 데이터를 나타내지 않습니다. 모델 논리와 데이터 값 및 레이블을 동기화하려면 모델을 다시 컴파일합니다. 현재 모델 논리를 유지하려는 경우 모델을 컴파일하지 마십시오.
- 분류자 모델 또는 확률 모델의 데이터를 보호하려면 **Informatica** 디렉터리 구조의 모델 파일을 백업합니다. 모델에서 모든 데이터 값 및 레이블을 제거하기 전에 파일을 백업합니다.

- 콘텐츠 관리 서비스 호스트 시스템에서 모델 파일을 찾습니다.

확률 모델 파일의 기본 위치 및 파일 이름 확장명은 다음과 같습니다.

`<Informatica_Installation_Directory>/tomcat/bin/ner/<파일 이름>.ner`

분류자 모델 파일의 기본 위치 및 파일 이름 확장명은 다음과 같습니다.

`<Informatica_Installation_Directory>/tomcat/bin/classifier/<파일 이름>.classifier`

- **Informatica** 설치를 업그레이드할 경우 매핑에서 확률 모델 및 분류자 모델을 사용하기 전에 모델을 컴파일해야 할 수 있습니다. 모델에 데이터가 포함되지 않는 경우 **Informatica** 디렉터리 구조의 현재 파일을 데이터가 포함된 백업 파일로 바꿉니다.

분류자 모델 및 확률 모델에서 레이블 관리

확률 모델 또는 분류자 모델에서 레이블을 검토하고 업데이트하려면 **레이블 관리** 대화 상자를 사용합니다.

1. 분류자 모델이 포함된 콘텐츠 집합을 엽니다.
2. 모델 이름을 선택하고 **편집**을 클릭합니다.
3. **레이블 관리** 대화 상자를 엽니다.

대화 상자에 모델 레이블이 나열됩니다.

관련 항목:

- [“분류자 모델 레이블 관리” 페이지 44](#)
- [“확률 모델 구성” 페이지 58](#)

콘텐츠 집합 작성

콘텐츠 집합을 작성하여 단일 프로젝트, 정보 유형 또는 비즈니스 용도를 나타내는 참조 데이터 개체를 관리할 수 있습니다.

1. **개체 탐색기** 보기에서 콘텐츠 집합을 저장할 프로젝트 또는 폴더를 선택합니다.
2. **파일 > 새로 만들기 > 콘텐츠 집합**을 클릭합니다.
3. 콘텐츠 집합의 이름을 입력합니다.
4. 필요한 경우 **찾아보기**를 선택하고 콘텐츠 집합의 모델 리포지토리 위치를 변경합니다.
5. **마침**을 클릭합니다.

콘텐츠 집합에 참조 데이터 개체 작성

콘텐츠 집합에 문자 집합, 패턴 집합, 토큰 집합, 정규식, 확률 모델 및 분류자 모델을 작성할 수 있습니다.

1. 편집기에서 콘텐츠 집합을 열고 **콘텐츠** 보기를 선택합니다.
2. 참조 데이터 개체 유형을 선택합니다.
3. **추가**를 클릭합니다.
4. 참조 데이터 개체의 이름을 입력합니다.
필요한 경우 개체에 대한 설명을 입력합니다.
5. 참조 데이터 개체 속성을 구성합니다.
6. **마침**을 클릭합니다.

팁: 한 콘텐츠 집합에서 다른 콘텐츠 집합으로 참조 데이터 개체를 복사할 수 있습니다. **복사 대상 위치** 옵션과 **붙여넣기 원본 위치** 옵션을 사용하여 콘텐츠 집합에 개체의 복사본을 작성합니다. 콘텐츠 집합 개체를 여러 개 선택하려면 **Ctrl** 키를 사용합니다.

제 4 장

분류자 모델

이 장에 포함된 항목:

- [분류자 모델 개요, 40](#)
- [분류자 모델 구조, 41](#)
- [분류자 점수, 41](#)
- [분류자 변환 예, 41](#)
- [분류자 모델 옵션, 42](#)
- [분류자 모델 참조 데이터, 43](#)
- [분류자 모델 레이블 데이터, 44](#)
- [분류자 모델 구성, 45](#)
- [필터 작업 및 찾기 작업, 49](#)
- [작업 복사 및 붙여넣기, 50](#)

분류자 모델 개요

분류자 모델은 콘텐츠 집합의 참조 데이터 개체입니다. 분류자 모델을 사용하여 여러 값이 포함된 긴 텍스트 문자열을 분석할 수 있습니다. 분류자 모델은 각 문자열에서 가장 공통되는 유형의 정보를 식별합니다.

분류자 변환에 분류자 모델을 추가합니다. 변환에서 분류자 모델 데이터와 각 입력 행의 데이터 사이에 공통된 값을 검색합니다. 변환에서 공통 값을 사용하여 각 행이 나타내는 정보 유형을 분류합니다.

입력 데이터에 다음과 같은 특성이 있을 경우 분류자 모델을 사용합니다.

- 입력 데이터에 텍스트가 있습니다. 분류자 모델은 자연어 프로세스를 텍스트 데이터에 적용하여 텍스트의 정보 유형을 식별합니다. 자연어 프로세스는 입력 문자열에서 관련 단어를 검색합니다. 자연어 프로세스는 관련되지 않은 단어를 무시합니다.
- 입력 데이터 문자열에 여러 값이 있습니다. 예를 들어 각 필드에 전자 메일 메시지 내용이 포함된 데이터 열을 작성할 수 있습니다.

분류자 변환은 문자열 데이터 유형을 읽습니다. 변환은 입력 문자열의 길이에 제한을 두지 않습니다.

Developer 도구에서 분류자 모델을 컴파일합니다. 모델을 컴파일할 때 모델 내 유사한 데이터 값 사이의 관계를 작성합니다. 분류자 변환은 컴파일된 데이터를 사용하여 입력 데이터에서 정보를 검색합니다.

분류자 모델 구조

분류자 모델에는 참조 데이터 값 및 레이블 값이 포함됩니다. 참조 데이터 값은 분류하려는 데이터를 나타냅니다. 레이블 값은 분류자 변환이 데이터에서 식별할 수 있는 정보의 유형을 지정합니다.

분류자 모델에는 컴파일 데이터도 포함됩니다. 분류자 변환은 컴파일 데이터를 사용하여 모델의 참조 데이터와 변환 입력 데이터 간의 유사성을 측정합니다. 분류자 모델을 컴파일할 때 컴파일 데이터를 작성하거나 업데이트합니다. 분류자 변환이 입력 데이터와 모델 데이터를 비교하는 경우 변환에서 입력 데이터의 각 행을 설명하는 레이블 값을 반환합니다.

Developer tool은 참조 데이터 값, 레이블 값 및 컴파일 데이터를 **Informatica** 디렉터리 구조의 파일에 씁니다. 모델 리포지토리의 분류자 모델 개체는 파일 이름을 저장합니다. 분류자 모델을 저장할 때 현재 참조 데이터 값과 레이블 값을 파일에 씁니다. 모델을 컴파일할 때 파일의 컴파일 데이터를 업데이트합니다. **Developer tool**의 모델 속성에서 파일 이름을 읽을 수 있습니다.

분류자 점수

분류자 변환은 입력 데이터의 각 행을 분류자 모델의 모든 참조 데이터 행과 비교합니다. 변환에서는 각 비교에 대해 점수를 계산합니다. 이 점수는 입력 행과 참조 데이터 행 사이의 유사성 정도를 나타냅니다.

분류자 변환이 포함된 매핑을 실행하면 매핑에서 점수가 가장 높은 참조 데이터 행을 나타내는 레이블을 반환합니다. 점수 범위는 0에서 1 사이입니다. 점수가 높으면 입력 데이터와 모델 데이터가 매우 일치함을 나타냅니다.

분류자 점수를 검토하여 레이블 출력이 각 입력 데이터 행을 정확하게 설명하는지 확인할 수 있습니다. 또한 점수를 검토하여 분류자 모델이 입력 데이터에 적절한지도 확인할 수 있습니다. 변환 출력에서 낮은 점수의 백분율이 높을 경우 분류자 모델이 부적절할 수 있습니다. 비교를 개선하려면 모델을 다시 컴파일합니다. 컴파일된 모델이 점수를 높이지 않을 경우 변환에서 모델을 바꿉니다.

분류자 변환 예

분류자 모델과 분류자 변환을 사용하여 전자 메일 메시지에 포함된 텍스트를 기반으로 메시지를 분류할 수 있습니다.

예를 들어, 사용자가 소프트웨어 제조업체의 고객 지원 센터에서 근무하는 데이터 스튜어드라고 가정합니다. 사용자는 고객이 지원 센터로 보내는 전자 메일 메시지를 검토합니다. 이 회사는 고객이 여러 나라에 분포하므로 지원 센터에서는 다양한 언어의 전자 메일을 받습니다. 고객에게 가장 적합하게 회신할 수 있는 부서로 각 전자 메일을 보내도록 언어별로 전자 메일을 분류하려고 합니다.

전자 메일을 분류하려면 다음 단계를 수행합니다.

1. 전자 메일 메시지를 단일 파일 또는 데이터베이스 테이블에 씁니다.
2. 파일 또는 데이터베이스 테이블을 읽는 모델 리포지토리에서 데이터 개체를 작성합니다.
3. 메시지에 사용되는 각 언어에 대해 모델 리포지토리에서 데이터 개체를 작성합니다.
4. 각 언어의 샘플 텍스트가 들어 있는 분류자 모델을 작성합니다.

참고: 전자 메일 메시지 데이터의 샘플 데이터를 모델의 소스 데이터로 사용할 수 있습니다.

5. 재사용 가능한 분류자 변환에 분류자 모델을 추가합니다.
6. 매핑을 구성하여 메시지 데이터에 분류자 변환을 적용합니다.

매핑을 구성하려면 다음 단계를 수행합니다.

- 매핑에 분류자 변환과 데이터 개체를 추가합니다.
- 분류자 변환 입력 포트를 소스 데이터 개체에 연결합니다.
- 분류자 변환 출력 포트를 대상 데이터 개체에 연결합니다.

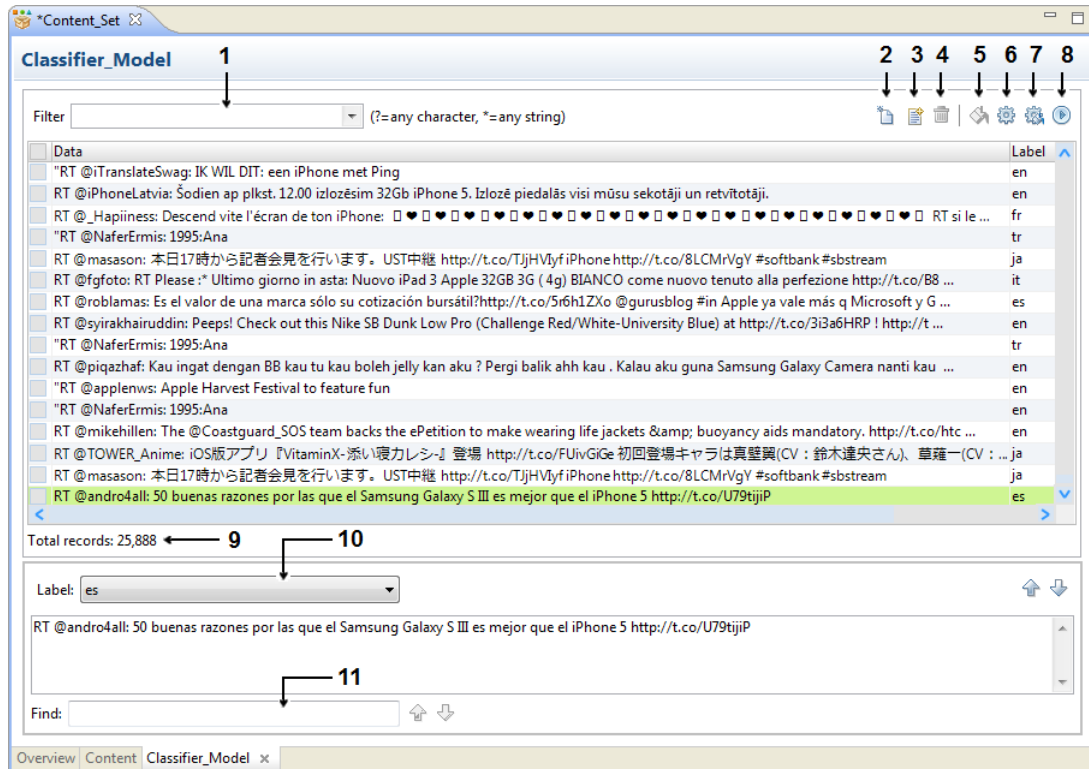
매핑을 실행하면 분류자 변환이 전자 메일 메시지를 분석하고 전자 메일 텍스트를 올바른 데이터 대상에 씁니다. 데이터 대상을 각 부서의 팀원과 공유할 수 있습니다.

분류자 모델 옵션

Developer tool은 상단 창과 하단 창이 있는 편집기에 분류자 모델 데이터를 표시합니다. 상단 창에는 데이터에 할당하는 모든 레이블과 참조 데이터의 각 행이 표시됩니다. 하단 창에는 선택하는 행의 콘텐츠가 표시됩니다.

참조 데이터 행을 검토하고 레이블을 사용하지 않는 행을 식별하려는 경우에는 상단 창을 사용합니다. 행의 콘텐츠를 검토하고 행에 레이블을 할당하려는 경우에는 하단 창을 사용합니다. 상단 창의 각 행에는 약 100자의 데이터가 표시됩니다. 하단 창에는 사용자가 선택하는 행의 데이터가 모두 표시됩니다.

다음 이미지는 분류자 모델 편집기를 보여 줍니다.



편집기에는 다음 옵션이 포함됩니다.

1. 필터 필드
사용자가 지정한 레이블 또는 데이터 값을 기반으로 참조 데이터 행의 목록을 필터링합니다.
2. 행 추가
빈 참조 데이터 행을 삽입합니다.

3. 데이터 추가
모델 리포지토리의 데이터 개체에서 데이터를 가져옵니다.
4. 삭제
사용자가 선택하는 참조 데이터 행을 삭제합니다. 행을 선택하려면 확인란을 사용합니다.
5. 레이블 할당
사용자가 선택하는 하나 이상의 참조 데이터 행에 레이블을 할당합니다. 행을 선택하려면 확인란을 사용합니다.
6. 속성 편집
분류자 모델 속성을 표시합니다.
7. 레이블 관리
레이블 관리 대화 상자를 엽니다. 분류자 모델에서 레이블 값을 추가하거나 삭제하려면 대화 상자를 사용합니다.
8. 컴파일
분류자 모델을 컴파일합니다.
9. 총 레코드 수
분류자 모델에 있는 참조 데이터 행의 수를 나타냅니다.
10. 레이블 필드
현재 참조 데이터 행에 적용할 수 있는 레이블 값을 표시합니다.
11. 찾기 필드
현재 참조 데이터 행에서 지정하는 데이터 값을 찾습니다.

분류자 모델 참조 데이터

분류자 모델에는 텍스트 문장, 단락 또는 페이지를 포함할 수 있는 참조 데이터 열이 있습니다. 참조 데이터는 분류자 모델이 매핑에서 읽을 수 있는 다양한 유형의 텍스트 입력을 나타냅니다. 모델을 작성할 경우 매핑을 실행할 때 찾을 것으로 예상되는 텍스트 유형이 참조 데이터에 포함되었는지 확인합니다.

매핑 소스 데이터를 사용하여 분류자 모델을 작성할 수 있습니다. 소스 데이터 샘플을 선택하고 데이터 샘플을 모델로 복사합니다.

분류자 모델 참조 데이터로 작업할 때 다음과 같은 규칙과 지침을 고려하십시오.

- 참조 데이터 필드에는 길이 제한이 없습니다. 각각의 데이터 필드에 텍스트 페이지를 입력할 수 있습니다.
- 데이터 개체에서 참조 데이터를 가져옵니다.
- 참조 데이터 값을 편집할 수는 없습니다. 하지만 데이터 행을 삭제할 수는 있습니다.
- 분류자 모델을 컴파일하는 경우 컴파일 프로세스가 참조 데이터의 모든 숫자 값을 무시합니다.

분류자 모델 레이블 데이터

분류자 모델에는 참조 데이터 행의 정보 유형을 요약하는 하나 이상의 설명 레이블이 포함되어 있습니다. 각 참조 데이터 행에 레이블을 할당합니다.

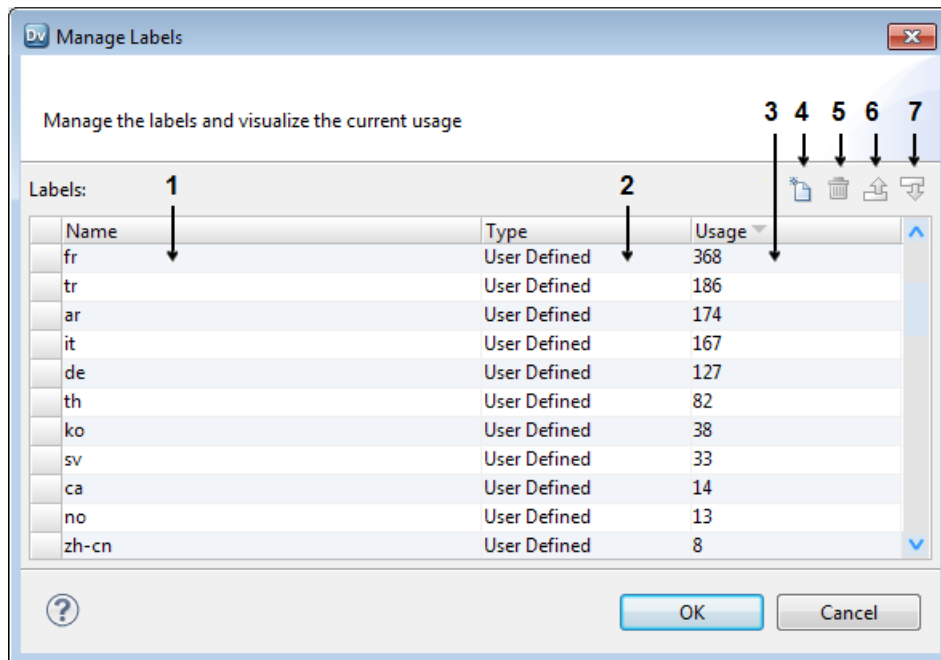
데이터 소스에서 분류자 모델로 데이터를 추가할 때 열을 레이블 데이터 열로 지정할 수 있습니다. 모델에서 레이블을 작성할 수도 있습니다.

레이블은 해당 레이블이 설명하는 참조 데이터 값과 독립적입니다. 레이블을 사용하는 참조 데이터 행을 삭제해도 모델에서 레이블이 삭제되지 않습니다. 레이블을 삭제해도 레이블과 연결된 참조 데이터 값이 삭제되지는 않습니다.

분류자 모델 레이블 관리

분류자 모델에서 레이블 값을 검토하고 업데이트하려면 **레이블 관리** 대화 상자를 사용합니다. 레이블 값을 정렬 및 업데이트할 수도 있습니다.

다음 이미지는 **레이블 관리** 대화 상자를 보여 줍니다.



레이블 관리 대화 상자에는 다음 요소가 포함되어 있습니다.

1. 이름 열.
분류자 변환이 입력 데이터 행에 적용할 수 있는 레이블 값이 포함되어 있습니다. 레이블을 이름별로 정렬할 수 있습니다.
2. 유형 열.
레이블 값의 소스를 식별합니다. 분류자 모델은 모든 레이블을 사용자 정의 값으로 식별합니다.
3. 사용 열.
각 레이블을 사용하는 참조 데이터 행의 수를 나타냅니다. 행 수를 기준으로 레이블을 정렬할 수 있습니다.
4. 추가 단추.
분류자 모델에 레이블을 추가합니다. 행의 이름 열에 레이블 값을 입력합니다.

참고: 레이블 값을 업데이트하려면 값을 두 번 클릭하고 원하는 값을 입력하십시오.

5. 삭제 단추.
분류자 모델에서 레이블을 삭제합니다.
6. 위쪽 화살표.
대화 상자에서 레이블을 한 행 위로 이동합니다.
7. 아래쪽 화살표.
대화 상자에서 레이블을 한 행 아래로 이동합니다.

분류자 모델 구성

분류자 모델의 구성 단계는 분류하려는 데이터로 시작합니다. 모델에 추가하는 참조 데이터의 콘텐츠는 분류자 변환에 연결하는 데이터를 반영해야 합니다. 이 변환에서는 입력 데이터의 데이터 값 및 패턴을 분류자 모델의 데이터 값 및 패턴과 비교합니다.

분류자 변환에서 사용할 수 있는 분류자 모델을 작성하려면 다음 태스크를 수행합니다.

1. 모델에 추가할 참조 데이터 값과 레이블 값을 식별합니다.
분류하려는 데이터의 조각을 사용할 수 있습니다. 데이터 조각을 읽는 모델 리포지토리에서 데이터 개체를 작성합니다.
2. 콘텐츠 집합을 작성하고 콘텐츠 집합에 분류자 모델을 추가합니다.
3. 모델에 참조 데이터 값을 추가합니다.
4. 모델에 레이블 값을 추가합니다.
데이터 개체에서 모델 리포지토리로 데이터를 가져올 수 있습니다. 단일 레이블 또는 참조 데이터의 단일 행을 입력할 수도 있습니다.
5. 참조 데이터의 각 행에 레이블을 할당합니다.
단일 작업에서 여러 행에 레이블을 할당할 수 있습니다.
6. 모델을 컴파일합니다.
분류자 모델을 컴파일한 후에는 분류자 변환에서 모델을 사용할 수 있습니다.

분류자 모델 작성

데이터 개체를 분류자 모델 데이터의 소스로 사용합니다.

분류자 모델은 분류자 변환의 입력 데이터를 모델 참조 데이터의 소스로 사용할 때 최적의 성능을 발휘합니다.

1. 개체 탐색기에서 콘텐츠 집합을 열거나 작성합니다.
2. **콘텐츠** 보기를 선택합니다.
3. **분류자 모델**을 선택하고 **추가**를 클릭합니다.
분류자 모델 마법사가 나타납니다.
4. 분류자 모델의 이름을 입력합니다.
필요한 경우 모델에 대한 텍스트 설명을 입력합니다.
5. 모델 리포지토리를 탐색하여 가져올 데이터가 포함된 데이터 개체를 선택합니다.
소셜 미디어 데이터 개체를 선택하지 않습니다.

다음을 클릭합니다.

6. 데이터 개체에서 열을 검토하고 모델에 추가할 하나 이상의 열을 선택합니다. 동일한 작업에서 참조 데이터 열과 레이블 열을 추가할 수 있습니다.

- 데이터 열을 참조 데이터로 가져오려면 열 이름을 선택하고 **데이터**를 클릭합니다.

여러 데이터 열을 선택할 수 있습니다. Developer tool이 사용자가 선택하는 열의 콘텐츠를 단일 열로 병합합니다.

- 데이터 열을 레이블 값으로 가져오려면 열 이름을 선택하고 **레이블**을 클릭합니다.

참조 데이터와 레이블 값을 가져올 때 Developer tool이 각 행의 레이블을 동일한 행의 참조 데이터 문자열에 할당합니다. 열을 선택하기 전에 데이터를 미리 볼 수 있습니다. 모델을 작성한 후 레이블 할당을 변경할 수 있습니다.

다음을 클릭합니다.

7. 데이터 소스에서 가져올 행 수를 선택합니다.

Developer tool은 기본적으로 데이터 소스의 모든 행을 가져옵니다. 숫자를 입력하는 경우 모델이 데이터 집합의 시작부터 행을 계산합니다.

8. **마침**을 클릭하고 모델을 저장합니다.

모델을 작성한 후 레이블 할당을 확인하고 모델을 컴파일합니다.

데이터 소스에서 분류자 모델로 데이터 추가

단일 작업에서 여러 개의 참조 데이터 값 또는 레이블 값 행을 분류자 모델로 가져올 수 있습니다.

1. 분류자 모델이 포함된 콘텐츠 집합을 엽니다.

2. 모델 이름을 선택하고 **편집**을 클릭합니다.

3. **데이터 추가**를 클릭합니다.

분류자 모델 마법사가 나타납니다.

4. 모델 리포지토리를 탐색하여 가져올 데이터가 포함된 데이터 개체를 선택합니다.

소셜 미디어 데이터 개체를 선택하지 않습니다.

다음을 클릭합니다.

5. 데이터 개체에서 열을 검토하고 모델에 추가할 하나 이상의 열을 선택합니다. 동일한 작업에서 참조 데이터 열과 레이블 열을 추가할 수 있습니다.

- 데이터 열을 참조 데이터로 가져오려면 열 이름을 선택하고 **데이터**를 클릭합니다.

여러 데이터 열을 선택할 수 있습니다. Developer tool이 사용자가 선택하는 열의 콘텐츠를 단일 열로 병합합니다.

- 데이터 열을 레이블 값으로 가져오려면 열 이름을 선택하고 **레이블**을 클릭합니다.

참조 데이터와 레이블 값을 가져올 때 Developer tool이 각 행의 레이블을 동일한 행의 참조 데이터 문자열에 할당합니다. 열을 선택하기 전에 데이터를 미리 볼 수 있습니다. 모델을 작성한 후 레이블 할당을 변경할 수 있습니다.

다음을 클릭합니다.

6. 데이터 소스에서 가져올 행 수를 선택합니다.

Developer tool은 기본적으로 데이터 소스의 모든 행을 가져옵니다. 숫자를 입력하는 경우 모델이 데이터 집합의 시작부터 행을 계산합니다.

7. **마침**을 클릭하고 모델을 저장합니다.

분류자 모델에 참조 데이터 행 추가

분류자 모델에 하나의 참조 데이터 행을 추가할 수 있습니다.

1. 분류자 모델이 포함된 콘텐츠 집합을 엽니다.
2. 모델 이름을 선택하고 **편집**을 클릭합니다.
3. **행 추가**를 클릭합니다.
Developer tool이 참조 데이터의 현재 행 아래에 행을 추가합니다.
4. 행에 참조 데이터 값을 입력합니다.
Windows 바로 가기를 사용하여 행에 데이터를 붙여 넣을 수 있습니다.

분류자 모델에 레이블 추가

분류자 모델에 단일 레이블을 추가할 수 있습니다.

1. 분류자 모델이 포함된 콘텐츠 집합을 엽니다.
2. 모델 이름을 선택하고 **편집**을 클릭합니다.
3. **레이블 관리** 대화 상자를 엽니다.
대화 상자에 모델 레이블이 나열됩니다.
4. **새로 만들기**를 클릭합니다.
Developer tool이 레이블 목록의 맨 아래에 행을 추가합니다.
5. 이름 열에서 기본값을 두 번 클릭하고 레이블 이름을 입력합니다.
6. **확인**을 클릭합니다.

레이블을 작성한 후 하나 이상의 참조 데이터 행에 레이블을 할당할 수 있습니다. **레이블 관리** 대화 상자의 사용 열은 레이블을 사용하는 행의 수를 나타냅니다.

참조 데이터 행에 레이블 할당

단일 작업에서 하나 이상의 참조 데이터 행에 레이블을 할당할 수 있습니다.

1. 모델이 포함된 콘텐츠 집합을 엽니다.
2. 모델 이름을 선택하고 **편집**을 클릭합니다.
3. 하나 이상의 참조 데이터 행을 선택합니다. 행을 선택하려면 확인란 옵션을 사용합니다.
참고: 지정한 데이터 값이 포함된 행을 모두 표시하는 필터 옵션을 사용할 수 있습니다. 값이 포함된 행을 모두 선택하려면 모두 선택 확인란 옵션을 사용하십시오.
4. **레이블 할당**을 클릭합니다.
Developer tool이 분류자 모델의 레이블 목록을 표시합니다.
5. 레이블 값을 선택하고 **할당**을 클릭합니다.
Developer tool이 사용자가 선택한 참조 데이터 행을 레이블 값으로 업데이트합니다.
필요한 경우 모델을 컴파일하여 레이블 이름을 분류자 모델 논리에 추가합니다.

사용되지 않은 레이블 값 식별

분류자 모델에서 사용되지 않은 상태로 유지되는 레이블 값을 찾으려면 **레이블 관리** 대화 상자를 사용합니다. **레이블 관리** 대화 상자는 분류자 모델의 레이블 값에 대한 사용 데이터를 표시합니다. 사용 데이터를 사용하면 레이블 값을 사용하는 참조 데이터 행의 수를 확인하고 사용되지 않은 레이블 값을 찾을 수 있습니다.

1. 분류자 모델이 포함된 콘텐츠 집합을 엽니다.
2. 모델 이름을 선택하고 **편집**을 클릭합니다.
3. **레이블 관리** 대화 상자를 엽니다. 대화 상자에 분류자 모델 레이블이 나열됩니다.
4. 각 레이블의 사용 열 데이터를 검토합니다.

사용 열에는 레이블을 사용하는 참조 데이터 행의 수가 나열됩니다. 레이블 값이 사용되지 않은 경우 사용 열의 값은 0입니다.

분류자 모델에서 행 삭제

단일 작업으로 분류자 모델에서 하나 이상의 참조 데이터 행을 삭제할 수 있습니다.

1. 모델이 포함된 콘텐츠 집합을 엽니다.
2. 모델 이름을 선택하고 **편집**을 클릭합니다.
3. 하나 이상의 참조 데이터 행을 선택합니다. 행을 선택하려면 확인란 옵션을 사용합니다.
4. **삭제**를 클릭합니다.

Developer tool이 분류자 모델에서 선택한 행을 제거됩니다.

작업을 실행 취소하려면 키보드에서 **Ctrl + Z** 키를 누릅니다.

분류자 모델에서 레이블 삭제

분류자 모델에서 레이블을 삭제하려면 **레이블 관리** 대화 상자를 사용합니다.

1. 모델이 포함된 콘텐츠 집합을 엽니다.
2. 모델 이름을 선택하고 **편집**을 클릭합니다.
3. **레이블 관리** 대화 상자를 엽니다.
4. **삭제**를 클릭합니다.
5. **예**를 클릭하여 작업을 확인합니다.

Developer tool이 모델에서 레이블을 삭제합니다. Developer tool은 레이블을 사용하는 어떤 참조 데이터 행도 삭제하지 않습니다.

6. **확인**을 클릭하여 대화 상자를 닫습니다.

작업을 실행 취소하려면 키보드에서 **Ctrl + Z** 키를 누릅니다.

분류자 모델 컴파일

분류자 모델에서 레이블 값 또는 참조 데이터 값을 편집할 때마다 모델을 컴파일해야 합니다. 모델을 컴파일할 때 모델의 컴파일 데이터를 업데이트합니다.

- ▶ 컴파일 데이터를 업데이트하려면 Developer 도구에서 모델을 열고 **컴파일**을 클릭합니다.

필터 작업 및 찾기 작업

지정한 조건과 일치하는 참조 데이터 행을 표시하거나 숨기려면 필터 작업을 사용합니다. 필터를 적용할 때 분류자 모델에서 표시하는 데이터 행에 대해 추가적인 작업을 수행할 수 있습니다. 예를 들어 모든 데이터 행에 레이블 값을 적용할 수 있습니다.

필터 옵션을 사용하여 다음 태스크를 수행할 수 있습니다.

- 입력한 값이 포함되어 있는 참조 데이터 행을 찾습니다.
- 선택한 레이블을 사용하는 참조 데이터 행을 찾습니다.
- 레이블을 사용하지 않는 참조 데이터 행을 찾습니다.

참조 데이터 행 내에서 데이터 값을 검색할 수도 있습니다.

데이터 값을 사용하여 참조 데이터 행 필터링

하나 이상의 참조 데이터 행에 원하는 데이터 값이 포함되어 있는지 확인하려면 필터를 사용합니다.

1. 분류자 모델이 포함된 콘텐츠 집합을 엽니다.
2. 모델 이름을 선택하고 **편집**을 클릭합니다.
3. 필터 필드에 값을 입력합니다.

입력하는 값에 와일드카드 문자를 포함할 수 있습니다.

Developer tool이 필터 텍스트가 포함된 참조 데이터 행을 표시합니다.

레이블 값을 사용하여 참조 데이터 행 필터링

선택한 레이블을 사용하는 참조 데이터 행을 표시하거나 숨기려면 필터를 사용합니다.

1. 모델이 포함된 콘텐츠 집합을 엽니다.
2. 모델 이름을 선택하고 **편집**을 클릭합니다.
3. 필터 메뉴에서 레이블 값을 선택합니다.

Developer tool이 레이블 값을 사용하는 참조 데이터 행을 표시합니다.

참고: 레이블을 사용하지 않는 참조 데이터 행을 찾으려면 필터 메뉴에서 **레이블 없음** 옵션을 선택합니다.

참조 데이터 행에서 값 찾기

선택한 행에서 데이터 값을 검색하려면 찾기 필드를 사용합니다.

1. 모델이 포함된 콘텐츠 집합을 엽니다.
2. 모델 이름을 선택하고 **편집**을 클릭합니다.
3. 참조 데이터 행을 선택합니다.
4. 찾기 필드에 값을 입력합니다.

참조 데이터 행에서 값의 인스턴스가 강조 표시됩니다.

5. 행에서 값의 추가 인스턴스를 찾으려면 위쪽 또는 아래쪽 화살표를 사용합니다.

작업 복사 및 붙여넣기

한 콘텐츠 집합에서 모델 리포지토리의 다른 콘텐츠 집합으로 분류자 모델을 복사할 수 있습니다. 다른 Developer 도구 사용자와 리소스를 공유할 수 있도록 분류자 모델을 복사합니다.

모델을 다른 콘텐츠 집합으로 복사하거나 모델을 현재 콘텐츠 집합으로 가져올 수 있습니다. 한 번의 작업에서 리포지토리의 여러 콘텐츠 집합에서 여러 모델을 가져올 수 있습니다.

모델을 복사하면 콘텐츠 관리 서비스가 서비스 시스템에 모델 데이터 파일의 복사본을 작성합니다. 각 모델은 서로 다른 데이터 파일을 사용합니다.

분류자 모델을 다른 콘텐츠 집합에 복사

한 콘텐츠 집합에서 모델 리포지토리의 다른 콘텐츠 집합으로 분류자 모델을 복사할 수 있습니다. 분류자 모델을 복사할 때 모델 개체와 소스 및 대상 콘텐츠 집합을 지정합니다.

1. 분류자 모델이 포함된 콘텐츠 집합을 엽니다.
2. 분류자 모델을 선택하고 **복사 대상**을 클릭합니다.
3. 모델 리포지토리를 탐색하고 콘텐츠 집합을 선택합니다.
분류자 모델을 현재 프로젝트 또는 다른 프로젝트의 콘텐츠 집합에 복사할 수 있습니다.
4. **확인**을 클릭합니다.
Developer도구가 분류자 모델을 선택한 콘텐츠 집합에 복사합니다.

다른 콘텐츠 집합에서 분류자 모델 가져오기

한 콘텐츠 집합에서 모델 리포지토리의 다른 콘텐츠 집합으로 분류자 모델을 가져올 수 있습니다. 분류자 모델을 가져올 때 하나 이상의 모델 개체와 소스 및 대상 콘텐츠 집합을 지정합니다.

1. 분류자 모델이 포함된 콘텐츠 집합을 엽니다.
2. 분류자 모델을 선택하고 **붙여넣기 원본 위치**를 클릭합니다.
3. 모델 리포지토리를 탐색하고 분류자 모델을 선택합니다.
분류자 모델을 현재 프로젝트 또는 다른 프로젝트에서 붙여넣을 수 있습니다.
4. **확인**을 클릭합니다.
Developer도구가 분류자 모델을 현재 콘텐츠 집합에 붙여넣습니다.

제 5 장

확률 모델

이 장에 포함된 항목:

- [확률 모델 개요, 51](#)
- [확률 모델 구조, 52](#)
- [라벨러 변환 예제, 52](#)
- [파서 변환 예제, 53](#)
- [확률 모델 옵션, 53](#)
- [확률 모델 참조 데이터, 56](#)
- [확률 모델 레이블 데이터, 56](#)
- [확률 모델 속성, 58](#)
- [확률 모델 구성, 58](#)
- [작업 복사 및 붙여넣기, 64](#)

확률 모델 개요

확률 모델은 콘텐츠 집합에 생성하는 참조 데이터 개체입니다. 확률 모델을 사용하여 여러 데이터 값을 포함하는 데이터 문자열을 분석할 수 있습니다. 확률 모델은 문자열에 포함된 각 값의 정보 유형을 식별합니다. 라벨러 변환 및 파서 변환에 확률 모델을 추가할 수 있습니다.

라벨러 변환에서 확률 모델을 사용하여 입력 문자열의 각 값에 설명 레이블을 적용할 수 있습니다. 라벨러 변환은 이러한 레이블을 단일 출력 포트에 씁니다. 파서 변환에서 확률 모델을 사용하여 입력 문자열의 각 값을 해당 값의 정보를 나타내는 포트에 쓸 수 있습니다. 파서 변환은 각 정보 유형에 대한 출력 포트를 생성합니다.

Developer tool에서 확률 모델을 설계하고 컴파일합니다. 확률 모델을 정의할 때 일련의 데이터 행을 모델에 추가하고 각 행의 각 값에 레이블을 지정합니다. 확률 모델을 컴파일하면 **Developer tool**이 데이터 값과 추가한 레이블 간에 연결을 생성합니다. 라벨러 변환과 파서 변환에서는 자연어 프로세스를 통해 확률 모델 데이터를 입력 포트 데이터와 비교합니다.

자연어 프로세스에서는 다음과 같은 기술을 기반으로 데이터 값에 포함된 정보의 유형을 식별합니다.

- 자연어 프로세스는 유사한 값을 인식하고 동일한 레이블을 해당 값에 적용합니다.
- 자연어 프로세스는 값 데이터를 문자열의 인접 값과 비교할 수 있습니다. 자연어 프로세스는 값의 시퀀스를 분석하여 각 문자열의 사용을 파악하고 해당 문자열이 나타내는 정보의 유형을 확인합니다.

확률 모델 구조

확률 모델에는 참조 데이터 값 및 레이블 값의 행이 포함됩니다. 참조 데이터 값은 변환 입력 데이터에 나타날 수 있는 서로 다른 값을 나타냅니다. 레이블 값은 입력 데이터에 포함될 것으로 예상하는 정보의 유형을 식별합니다.

확률 모델에는 컴파일 데이터도 포함됩니다. 라벨러 변환과 파서 변환은 컴파일 데이터를 사용하여 모델의 참조 데이터와 변환 입력 데이터 간의 유사성을 측정합니다. 확률 모델을 컴파일할 때 컴파일 데이터를 작성하거나 업데이트합니다.

데이터 행에는 단일 값 또는 여러 값이 포함될 수 있습니다. 각 데이터 행의 구조는 서로 다를 수 있습니다. 데이터 행의 서로 다른 값에 동일한 레이블을 할당할 수 있습니다. 또는 한 행의 여러 위치에 나타나는 동일한 값에 서로 다른 레이블을 할당할 수도 있습니다. 데이터 통합 서비스는 매핑이 실행될 때 입력 문자열 값의 상대적 위치를 고려합니다. 확률 모델을 컴파일하려면 먼저 하나 이상의 데이터 값에 각각의 레이블을 할당합니다.

Developer tool은 참조 데이터 값, 레이블 값 및 컴파일 데이터를 Informatica 디렉터리 구조의 파일에 씁니다. 모델 리포지토리의 확률 모델 개체는 파일 이름을 저장합니다. 확률 모델을 저장할 때 현재 참조 데이터 값과 레이블 값을 파일에 씁니다. 모델을 컴파일할 때 파일의 컴파일 데이터를 업데이트합니다. **Developer tool**의 모델 속성에서 파일 이름을 읽을 수 있습니다.

참고: 확률 모델의 기능을 최적화하려면 각 데이터 행에 여러 참조 데이터 값이 있는지 확인합니다. 각 행에서 값의 순서는 해당 값이 변환 입력 데이터에 나타나는 순서와 가능한 한 근접하게 일치해야 합니다. 데이터 행에 단일 참조 데이터 값이 포함되어 있을 경우에는 라벨러 변환 또는 파서 변환이 확률 분석 중 자연어 프로세스를 적용할 수 없습니다.

라벨러 변환 예제

보험 회사의 고객 데이터베이스에 여러 가지 데이터 입력 오류가 있습니다. 이 보험 회사의 데이터 스튜어더로서, 각 열에 포함된 여러 가지 유형의 데이터를 확인하기 위해 라벨러 변환과의 매핑을 구성합니다.

다음 표에는 고객 데이터베이스의 샘플 데이터가 설명되어 있습니다.

행 ID	필드 1	필드 2	필드 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 LEVENER	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

매핑을 실행하면 라벨러 변환이 입력 데이터를 확률 모델 참조 데이터와 비교합니다. 라벨러 변환이 각 입력 포트의 데이터에 대한 레이블을 선택합니다. 라벨러 변환이 레이블을 출력 포트에 씁니다. 각 출력 행에는 해당 입력 행의 데이터 구조를 정의하는 레이블 집합이 포함됩니다.

다음 표에는 라벨러 변환이 출력 포트에 추가하는 레이블이 설명되어 있습니다.

행 ID	출력 레이블
1	number organization contact
2	number contact organization
3	number organization contact
4	number contact organization
5	number organization contact
6	organization number contact
7	organization number contact
8	contact organization number

파서 변환 예제

수퍼마켓에서 제품 설명을 데이터베이스 테이블의 단일 열에 저장합니다. 제품 설명에는 다양한 유형의 정보를 나타내는 여러 가지 데이터 값이 포함됩니다. 이 수퍼마켓의 데이터 스튜어드로서, 제품 설명에 포함된 다양한 유형의 정보에 대한 열을 생성하려고 합니다.

데이터 값을 올바른 필드로 구조화하기 위해 파서 변환과의 매핑을 구성합니다.

다음 데이터 조각에는 오렌지 주스에 대한 제품 설명이 포함되어 있습니다.

Sunnydream Orange Juice Unsweetened 12 oz

다음 표에는 파서 변환이 입력 데이터를 기반으로 생성하는 출력 데이터가 설명되어 있습니다.

제품 이름	제품 유형	제품 세부 정보	제품 크기
Sunnydream	Orange Juice	Unsweetened	12 oz

확률 모델 옵션

확률 모델을 편집할 때 데이터 보기 또는 레이블 보기에서 작업할 수 있습니다.

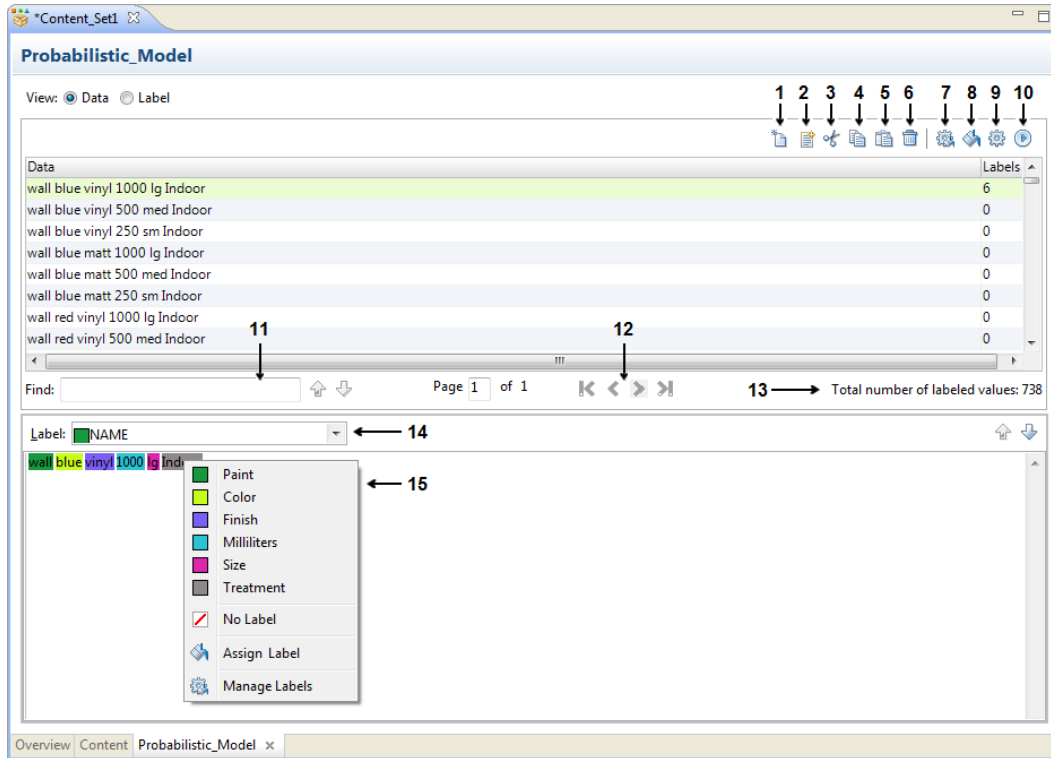
데이터 보기를 사용하여 모델에 참조 데이터 행을 추가하고 각 행의 데이터 값에 레이블을 할당할 수 있습니다. 모델에서 레이블 값의 사용에 관한 자세한 정보를 검토하려면 레이블 보기를 사용합니다. 데이터 보기 및 레이블 보기의 확률 모델에 레이블을 추가할 수 있습니다.

확률 모델 데이터 보기

데이터 보기에는 확률 모델의 참조 데이터 행과 각 행에 할당한 레이블 값의 수가 표시됩니다. 데이터 보기에는 현재 모델의 값에 할당한 총 레이블 수도 표시됩니다.

참조 데이터 행을 선택할 때 행의 값이 찾기 필드 아래의 편집기에 나타납니다. 행의 참조 데이터 값에 레이블을 할당하려면 편집기에서 값을 마우스 오른쪽 단추로 클릭하고 레이블 값을 선택합니다.

다음 이미지는 데이터 보기를 선택할 때 사용할 수 있는 확률 모델 옵션을 보여 줍니다.



데이터 보기에는 다음 옵션이 포함됩니다.

1. 행 추가
빈 데이터 행을 삽입합니다.
2. 데이터 추가
모델 리포지토리의 데이터 개체에서 데이터를 가져옵니다.
3. 잘라내기
확률 모델에서 데이터 행을 제거하고 클립보드에 데이터 행을 추가합니다.
4. 복사
클립보드에 데이터 행을 복사합니다.
5. 붙여넣기
클립보드에서 확률 모델로 데이터를 붙여 넣습니다.
6. 삭제
확률 모델에서 데이터 행을 삭제합니다.
7. 레이블 관리

레이블 관리 대화 상자를 엽니다. 확률 모델에서 레이블 값을 추가하거나 삭제하려면 대화 상자를 사용합니다.

8. 레이블 할당

사용자가 선택하는 하나 이상의 참조 데이터 값에 레이블을 할당합니다. 옵션을 사용하여 모델에 있는 모든 참조 데이터 인스턴스에 레이블을 할당할 수 있습니다.

9. 속성 편집

확률 모델 속성을 표시합니다.

10. 컴파일

확률 모델을 컴파일합니다.

11. 찾기 필드

입력한 참조 데이터 값이 포함되어 있는 행을 모델에서 찾습니다. 위쪽 및 아래쪽 화살표를 사용하여 값이 포함된 행으로 이동합니다.

12. 앞으로 및 뒤로 화살표

모델의 데이터 값 행 사이를 앞으로 또는 뒤로 이동합니다.

13. 레이블 지정된 값의 총 수

레이블을 사용하는 참조 데이터 값의 수를 나타냅니다.

14. 레이블 필드

선택한 참조 데이터 값에 적용할 수 있는 레이블 값을 표시합니다.

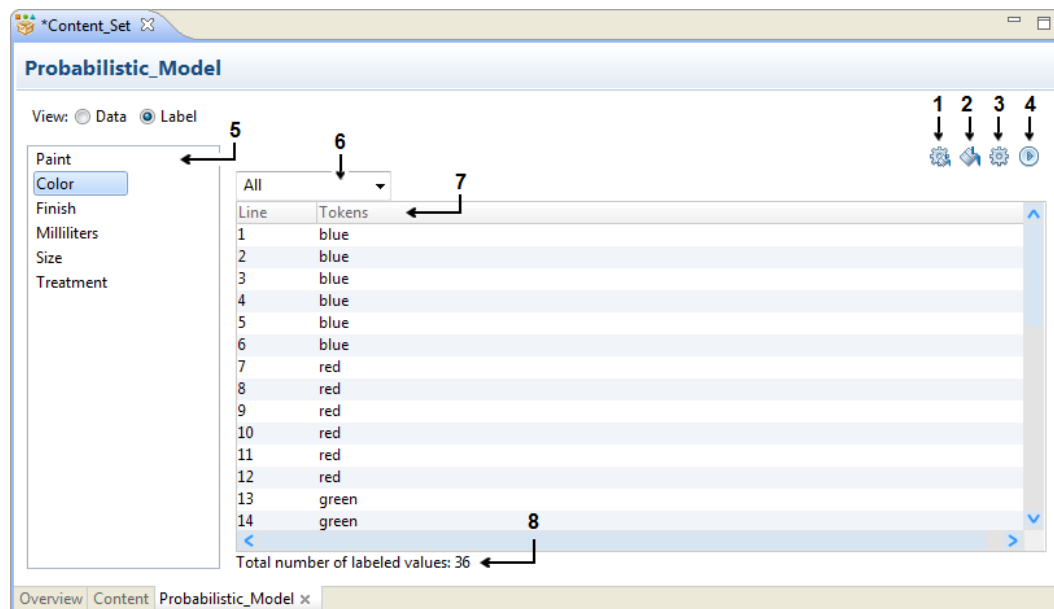
15. 레이블 메뉴

하나 이상의 참조 데이터 값에 레이블을 할당하는 데 사용할 수 있는 옵션 목록을 표시합니다. 메뉴를 열려면 참조 데이터 편집기에서 참조 데이터 값을 마우스 오른쪽 단추로 클릭합니다.

확률 모델 레이블 보기

레이블 보기에는 확률 모델에 정의한 레이블이 나열됩니다. 레이블을 선택하면 레이블 보기에 각 행의 레이블에 할당한 데이터 값이 표시됩니다.

다음 이미지는 데이터 보기를 선택할 때 사용할 수 있는 확률 모델 옵션을 보여 줍니다.



레이블 보기에는 다음 옵션이 포함됩니다.

1. 레이블 관리
레이블 관리 대화 상자를 엽니다. 확률 모델에서 레이블 값을 추가하거나 삭제하려면 대화 상자를 사용합니다.
2. 레이블 할당
사용자가 선택하는 하나 이상의 참조 데이터 값에 레이블을 할당합니다.
단일 데이터 값에 레이블을 할당하거나 단일 작업에서 여러 값에 레이블을 할당할 수 있습니다.
3. 속성 편집
확률 모델 속성을 표시합니다.
4. 컴파일
확률 모델을 컴파일합니다.
5. 레이블 값 목록
모델의 참조 데이터 값에 할당할 수 있는 레이블을 나열합니다.
6. 할당 필터
선택한 레이블을 사용하는 참조 데이터 값 목록을 필터링합니다. 필터 옵션은 데이터 값에 레이블을 할당하는 데 사용한 방법을 기반으로 참조 데이터 값을 표시하거나 숨깁니다.
필터를 적용하는 경우 레이블 보기의 레이블이 지정된 값의 총 수가 필터 조건을 충족하는 값 수를 반영합니다.
7. 참조 데이터 값 열
현재 레이블을 사용하는 참조 데이터 값을 나열합니다.
8. 레이블 지정된 값의 총 수
현재 레이블을 사용하는 참조 데이터 값의 수를 나타냅니다.

확률 모델 참조 데이터

확률 모델의 참조 데이터 값은 매핑에서 변환에 연결할 수 있는 입력 데이터의 유형을 나타냅니다.

Developer 도구에서 참조 데이터 행을 추가, 편집 및 삭제할 수 있습니다. 클립보드에서 데이터를 붙여 넣을 수 있고 데이터 소스에서 데이터를 가져올 수 있습니다. 참조 데이터 값을 추가한 후 각 행의 각 데이터 값에 레이블을 할당합니다.

확률 모델 레이블 데이터

확률 모델의 레이블 값은 참조 데이터 값에 포함될 수 있는 정보의 유형을 나타냅니다. 모델에 참조 데이터 행을 추가할 때 각 행에 있는 각각의 값에 레이블을 할당합니다. 모델에 추가하는 레이블은 레이블 보기 및 데이터 보기의 메뉴 옵션에 표시됩니다.

모델의 레이블을 참조 데이터 값에 할당할 수 있습니다. 동일한 값이 서로 다른 참조 데이터 행에서 의미가 다를 경우 각 행의 값에 서로 다른 레이블을 할당할 수 있습니다.

레이블 값의 범위는 확률 분석 중 라벨러 변환 또는 파서 변환이 읽는 입력 포트의 범위에 해당할 수 있습니다. 확률 모델에는 변환이 각 입력 포트의 데이터 값에 적용할 수 있는 레이블 값이 하나 이상 포함되어 있어야 합니다.

예를 들어, 웨어하우스가 8개의 열을 정의하는 쉼표로 구분된 파일에 인벤토리 데이터를 저장할 수 있습니다. 인벤토리 데이터를 데이터베이스 테이블로 구분 분석하는 매핑을 설계합니다. 각 데이터 열에 대한 레이블 값으로 확률 모델을 작성합니다. 매핑을 실행할 때 파서 변환이 입력 데이터의 각 값을 대상 테이블의 올바른 열에 씁니다.

다음 테이블에서는 확률 모델에서 작성할 수 있는 레이블 값과 인벤토리 데이터의 열을 보여 줍니다.

인벤토리 열 이름	레이블 이름
Product_Name	Product_Name
Quantity	Quantity
Location	Location
Barcode	Barcode
SKU	Stock_Keeping_Unit
Arrival_Date	Arrival_Date
Cost_Price	Cost_Price

참고: 입력 열 이름을 사용하거나 다른 이름을 사용할 수 있습니다. 이름이 일치해야 할 필요는 없습니다.

오버플로우 레이블

변환에서 입력 데이터 값에 레이블을 적용할 수 없는 경우 변환은 해당 데이터 값을 오버플로우 데이터로 처리합니다. 라벨러 변환은 식별할 수 없는 모든 데이터 값에 오버플로우 레이블을 적용합니다. 파서 변환은 식별할 수 없는 모든 데이터 값을 오버플로우 포트에 씁니다.

다음 테이블에서는 파서 변환이 오버플로우 포트를 사용하여 확률 모델이 인식하지 못하는 주소 데이터 요소를 구분 분석하는 방식을 보여 줍니다.

입력 데이터	Street_Name 포트	Street_Descriptor 포트	오버플로우 포트
Park Place	Park	Place	오버플로우 데이터 없음
Park Avenue	Park	Avenue	오버플로우 데이터 없음
Madison Avenue	Madison	Avenue	오버플로우 데이터 없음
Central Park	Central	Park	오버플로우 데이터 없음
Washington Square Park	Washington	Square	Park
Madison Square Garden	Madison	Square	Garden

또한 파서 변환은 입력 값의 수가 모델에 있는 레이블의 수보다 클 경우 값을 오버플로우 포트에 씁니다. 변환에서 확률 모델을 사용하기 전에 입력 데이터를 검토하고 모델에 올바른 개수의 레이블 값이 포함되었는지 확인합니다.

확률 모델 속성

확률 모델의 일반 속성과 고급 속성을 검토할 수 있습니다.

속성 편집기를 열려면 데이터 보기 또는 레이블 보기에서 **속성 편집** 옵션을 선택합니다.

일반 속성에는 확률 모델의 이름, 모델에 대한 설명 및 모델 데이터 파일의 이름이 표시됩니다. 고급 속성에는 Developer 도구가 확률 모델을 컴파일하는 데 사용하는 계산 속성이 표시됩니다.

확률 모델의 컴파일에서 기본 요소는 *n-gram*입니다. *n-gram*은 단어를 완성하기 위해 다른 글자 뒤에 오거나 앞에 오는 일련의 글자입니다. 매핑이 실행되면 라벨러 변환 또는 파서 변환이 확률 모델의 참조 데이터 열에 각 값에 대한 여러 *n-gram*을 생성합니다. 변환에서는 입력 데이터 값을 참조 데이터 값 및 *n-gram*과 비교합니다. 확률 모델의 고급 속성은 확률 모델이 *n-gram* 및 다른 모델 요소를 처리하는 방식에 따라 결정됩니다.

참고: 고급 속성의 기본값은 확률 분석 및 확률 모델 컴파일에서 기본이 되는 설정을 나타냅니다. 고급 속성을 편집할 경우 확률 분석의 정확성에 부정적인 영향을 줄 수도 있습니다. 변경 사항의 결과를 이해하지 않는 한 고급 속성을 편집하지 마십시오.

관련 항목:

- [“확률 모델 및 분류자 모델에 대한 규칙 및 지침” 페이지 37](#)

확률 모델 구성

확률 모델 구성 단계는 수행하려는 분석의 유형으로 시작됩니다. 입력 문자열의 각 값에서 정보 유형을 식별하려면 라벨러 변환의 확률 모델을 사용합니다. 입력 문자열의 데이터 값을 서로 다른 출력 포트에 구문 분석하려면 파서 변환의 확률 모델을 사용합니다.

데이터 레이블 지정 및 데이터 구문 분석에 동일한 확률 모델을 사용할 수 있습니다. 라벨러 변환의 모델을 사용하는 경우 변환에서 사용자가 선택하는 각 입력 포트에 대해 단일 출력 포트를 작성합니다. 파서 변환의 모델을 사용하는 경우 변환에서 식별하는 입력 데이터의 각 유형에 대해 출력 포트를 작성합니다.

확률 모델을 작성하려면 다음 태스크를 수행합니다.

1. 모델에 추가할 참조 데이터 값과 레이블 값을 식별합니다.
분석하려는 데이터의 조각을 사용할 수 있습니다. 데이터 조각을 읽는 모델 리포지토리에서 데이터 개체를 작성합니다.
2. 콘텐츠 집합을 작성하고 콘텐츠 집합에 확률 모델을 추가합니다.
3. 모델에 참조 데이터 값을 추가합니다.
4. 모델에 레이블 값을 추가합니다.
데이터 개체에서 모델 리포지토리로 데이터를 가져올 수 있습니다. 단일 레이블 또는 참조 데이터의 단일 행을 입력할 수도 있습니다.
데이터 구문 분석을 위해 확률 모델을 사용하려면 모델에 변환에서 작성해야 하는 각 출력 포트에 대한 레이블 값이 포함되어 있는지 확인합니다.
5. 각 행의 각 데이터 값을 참조하도록 레이블을 할당합니다.
단일 작업에서 여러 참조 데이터 값에 레이블을 할당할 수 있습니다.
6. 모델을 컴파일합니다.
확률 모델을 컴파일한 후에는 변환에서 모델을 사용할 수 있습니다.

빈 확률 모델 작성

참조 데이터 또는 레이블 데이터가 포함되지 않은 확률 모델 개체를 작성할 수 있습니다. 빈 모델을 작성하고 데이터를 추가하거나 데이터를 모델로 가져옵니다.

1. 개체 탐색기에서 콘텐츠 집합을 열거나 작성합니다.
2. 콘텐츠 보기를 선택합니다.
3. **확률 모델**을 선택하고 **추가**를 클릭합니다.
확률 모델 마법사가 열립니다.
4. **확률 모델** 옵션을 선택합니다.
다음을 클릭합니다.
5. 확률 모델의 이름을 입력합니다.
필요한 경우 모델에 대한 텍스트 설명을 입력합니다.
6. **마침**을 클릭합니다.

데이터 개체를 기반으로 확률 모델 생성

데이터 개체를 확률 모델 데이터의 소스로 사용할 수 있습니다.

확률 모델은 라벨러 또는 파서 변환의 입력 데이터를 모델 참조 데이터의 소스로 사용할 때 최적의 성능을 발휘합니다.

1. 개체 탐색기에서 콘텐츠 집합을 열거나 작성합니다.
2. **콘텐츠** 보기를 선택합니다.
3. **확률 모델**을 선택하고 **추가**를 클릭합니다.
확률 모델 마법사가 열립니다.
4. **데이터 개체의 확률 모델** 옵션을 선택합니다.
다음을 클릭합니다.
5. 확률 모델의 이름을 입력합니다.
필요한 경우 모델에 대한 텍스트 설명을 입력합니다.
6. 모델 리포지토리를 탐색하여 가져올 데이터가 포함된 데이터 개체를 선택합니다.
소셜 미디어 데이터 개체를 선택하지 않습니다.
다음을 클릭합니다.
7. 데이터 개체에서 열을 검토하고 모델에 추가할 하나 이상의 열을 선택합니다. 동일한 작업에서 참조 데이터 열과 레이블 열을 추가할 수 있습니다.
 - 데이터 열을 참조 데이터로 가져오려면 열 이름을 선택하고 **데이터**를 클릭합니다.
여러 데이터 열을 선택할 수 있습니다. **Developer tool**이 사용자가 선택하는 열의 콘텐츠를 단일 열로 병합합니다.
 - 데이터 열을 레이블 값으로 가져오려면 열 이름을 선택하고 **레이블**을 클릭합니다.참조 데이터와 레이블 값을 가져올 때 **Developer tool**이 각 행의 레이블을 동일한 행의 참조 데이터 문자열에 할당합니다. 열을 선택하기 전에 데이터를 미리 볼 수 있습니다. 모델을 작성한 후 레이블 할당을 변경할 수 있습니다.
다음을 클릭합니다.
8. 데이터 소스에서 가져올 행 수를 선택합니다.

Developer tool은 기본적으로 데이터 소스의 모든 행을 가져옵니다. 숫자를 입력하는 경우 모델이 데이터 집합의 시작부터 행을 계산합니다.

- 가져온 데이터 값에 대한 구분자를 지정합니다.

참조 데이터 값과 레이블 값에 서로 다른 구분자를 지정할 수 있습니다. 기본 구분자는 문자 공백입니다.

- 마침**을 클릭하고 모델을 저장합니다.

확률 모델을 작성한 후 레이블 할당을 확인하고 모델을 컴파일합니다.

데이터 소스에서 확률 모델로 데이터 추가

단일 작업에서 여러 개의 참조 데이터 값 및 레이블 값 행을 확률 모델로 가져올 수 있습니다.

- 확률 모델이 포함된 콘텐츠 집합을 엽니다.

- 모델 이름을 선택하고 **편집**을 클릭합니다.

- 데이터 추가**를 클릭합니다.

확률 모델 마법사가 열립니다.

- 모델 리포지토리를 탐색하여 가져올 데이터가 포함된 데이터 개체를 선택합니다.

소셜 미디어 데이터 개체를 선택하지 않습니다.

다음을 클릭합니다.

- 데이터 개체에서 열을 검토하고 모델에 추가할 하나 이상의 열을 선택합니다. 동일한 작업에서 참조 데이터 열과 레이블 열을 추가할 수 있습니다.

- 데이터 열을 참조 데이터로 가져오려면 열 이름을 선택하고 **데이터**를 클릭합니다.

여러 데이터 열을 선택할 수 있습니다. **Developer tool**이 사용자가 선택하는 열의 콘텐츠를 단일 열로 병합합니다.

- 데이터 열을 레이블 값으로 가져오려면 열 이름을 선택하고 **레이블**을 클릭합니다.

참조 데이터와 레이블 값을 가져올 때 **Developer tool**이 각 행의 레이블을 동일한 행의 참조 데이터 문자열에 할당합니다. 열을 선택하기 전에 데이터를 미리 볼 수 있습니다. 모델을 작성한 후 레이블 할당을 변경할 수 있습니다.

다음을 클릭합니다.

- 데이터 소스에서 가져올 행 수를 선택합니다.

Developer tool은 기본적으로 데이터 소스의 모든 행을 가져옵니다. 숫자를 입력하는 경우 모델이 데이터 집합의 시작부터 행을 계산합니다.

- 가져온 데이터 값에 대한 구분자를 지정합니다.

참조 데이터 값과 레이블 값에 서로 다른 구분자를 지정할 수 있습니다. 기본 구분자는 문자 공백입니다.

- 마침**을 클릭하고 모델을 저장합니다.

확률 모델에 참조 데이터 행 추가

확률 모델에 빈 행을 추가하려면 데이터 보기를 사용합니다.

- 모델이 포함된 콘텐츠 집합을 엽니다.

모델 이름을 선택하고 **편집**을 클릭합니다.

- 데이터 보기를 선택합니다.

- 모델에 빈 행을 추가하려면 **새로 만들기**를 클릭합니다.

- 추가한 행을 선택하고 하나 이상의 참조 데이터 값을 행에 입력합니다.

5. 확률 모델을 저장합니다.

모델을 저장한 후 행의 각 값에 레이블을 할당합니다. 필요한 경우 모델을 컴파일합니다.

확률 모델에 레이블 추가

확률 모델에 단일 레이블을 추가할 수 있습니다. 모델 데이터 값이 나타내는 모든 유형의 정보에 레이블을 추가합니다. 파서 변환에서 확률 모델을 사용하는 경우 변환에서 생성할 것으로 예상되는 각 출력 포트에 대해 레이블을 추가합니다.

1. 모델이 포함된 콘텐츠 집합을 엽니다.
2. 모델 이름을 선택하고 **편집**을 클릭합니다.
3. 데이터 보기 또는 레이블 보기에서 **레이블 관리**를 클릭합니다.
레이블 관리 대화 상자가 나타납니다.
4. **레이블 관리** 대화 상자에서 **새로 만들기**를 클릭합니다.
레이블이 대화 상자의 첫 번째 빈 행에 나타납니다.
5. 레이블 이름을 편집합니다. 필요한 경우 레이블 색상을 업데이트합니다.
6. **확인**을 클릭하여 레이블을 모델에 추가합니다.
7. 확률 모델을 저장합니다.

레이블을 추가한 후에는 레이블을 하나 이상의 데이터 값에 할당해야 합니다.

참조 데이터 값에 레이블 할당

참조 데이터 행의 단일 데이터 값에 레이블을 할당할 수 있습니다.

데이터 값이 행 또는 서로 다른 행의 여러 위치에 나타나는 경우 동일한 데이터 값에 여러 레이블을 할당할 수 있습니다.

1. 모델이 포함된 콘텐츠 집합을 엽니다.
2. 모델 이름을 선택하고 **편집**을 클릭합니다.
3. 데이터 보기를 선택합니다.
4. 레이블이 잘못되었거나 없는 데이터 값을 찾습니다. 레이블을 사용하는 데이터 값은 색상으로 구분되어 있습니다.
5. 데이터 값이 포함된 데이터 행을 선택합니다.
행이 편집기에 표시됩니다.
6. 편집기에서 데이터 값을 마우스 오른쪽 단추로 클릭하고 컨텍스트 메뉴에서 레이블을 선택합니다.
Developer 도구가 레이블을 데이터 값에 할당합니다.
7. 확률 모델을 저장합니다.

확률 모델을 저장한 후 필요한 경우 모델을 컴파일합니다.

여러 데이터 값에 레이블 할당

단일 작업에서 여러 참조 데이터 값에 레이블을 할당할 수 있습니다.

1. 모델이 포함된 콘텐츠 집합을 엽니다.
2. 모델 이름을 선택하고 **편집**을 클릭합니다.
3. **레이블 할당**을 클릭합니다.

여러 값에 레이블 할당 대화 상자가 열립니다.

4. 찾기 필드에 하나 이상의 문자를 입력합니다.

찾기 필드에 와일드카드 문자를 입력할 수 있습니다.

5. 필요한 경우 추가 검색 조건을 선택합니다.

다음 옵션을 선택하거나 지울 수 있습니다.

- 대/소문자 구분.

검색 작업 시 대/소문자를 구분하도록 지정합니다. 이 옵션과 함께 와일드카드 문자를 사용하지 마십시오.

- 전체 문자열 일치. 검색 작업에서 참조 데이터 값의 문자와 입력하는 문자 간 전체 일치를 찾으도록 지정합니다. 이 옵션과 함께 와일드카드 문자를 사용하지 마십시오.

- 레이블 지정된 값 무시.

검색 작업에서 레이블을 사용하는 참조 데이터 값을 건너뛰도록 지정합니다.

6. 검색 조건과 일치하는 참조 데이터 값에 할당할 레이블을 선택합니다.

레이블 없음 옵션을 선택할 수도 있습니다. 입력한 문자가 포함된 참조 데이터 값에서 레이블을 제거하려면 이 옵션을 선택합니다.

7. 시작을 클릭합니다.

Developer tool이 사용자가 정의한 검색 조건과 일치하는 모든 참조 데이터 값에 레이블을 할당합니다.

참고: 단일 작업에서 레이블을 지정한 참조 데이터 값을 보려면 레이블 보기에서 **대량으로 할당됨** 필터를 사용합니다.

확률 모델에서 행 삭제

단일 작업으로 확률 모델에서 하나 이상의 참조 데이터 행을 삭제할 수 있습니다.

1. 모델이 포함된 콘텐츠 집합을 엽니다.
2. 모델 이름을 선택하고 **편집**을 클릭합니다.
3. 데이터 보기에서 하나 이상의 참조 데이터 행을 선택합니다.
4. **삭제**를 클릭합니다.

Developer tool이 분류자 모델에서 선택한 행을 제거됩니다.

작업을 실행 취소하려면 키보드에서 **Ctrl + Z** 키를 누릅니다.

확률 모델에서 레이블 삭제

모델에서 레이블 값을 삭제할 경우 해당 레이블을 사용했던 모든 참조 데이터 값이 모델에서 유지됩니다. 다른 레이블 값을 각 참조 데이터 값에 할당하십시오.

1. 모델이 포함된 콘텐츠 집합을 엽니다.
2. 모델 이름을 선택하고 **편집**을 클릭합니다.
3. 데이터 보기 또는 레이블 보기에서 **레이블 관리**를 클릭합니다.
4. **레이블 관리** 대화 상자에서 레이블 값을 선택합니다.
5. **삭제**를 클릭합니다.
6. **확인**을 클릭하여 레이블을 삭제합니다.
7. 확률 모델을 저장합니다.

참고: 레이블은 확률 모델의 구조적 요소입니다. 모델을 변환에 추가한 후 레이블을 추가하거나 제거할 경우 모델을 사용하는 작업이 무효화됩니다. 업데이트한 모델을 사용하려면 변환 작업을 삭제하고 다시 생성합니다.

확률 모델 컴파일

확률 모델에서 데이터 또는 레이블 할당을 업데이트할 때 모델을 컴파일할 수 있습니다. 현재 참조 데이터 값과 현재 레이블 값 간의 연결을 사용하여 모델 논리를 업데이트하려면 모델을 컴파일합니다.

확률 모델을 컴파일하기 전에 각 레이블 값이 최소 하나의 참조 데이터 값을 식별하는지 확인합니다.

- ▶ 모델을 컴파일하려면 Developer 도구에서 모델을 열고 **컴파일**을 클릭합니다.

확률 모델에서 데이터 행 찾기

사용자가 입력한 값이 포함된 참조 데이터 행을 찾으려면 데이터 보기를 사용합니다.

1. 확률 모델이 포함된 콘텐츠 집합을 엽니다.
2. 모델 이름을 선택하고 **편집**을 클릭합니다.
3. 데이터 보기를 선택합니다.
4. **찾기** 필드에 하나 이상의 문자를 입력합니다.
데이터 보기에 모델에서 입력한 값이 포함된 첫 번째 행이 표시됩니다.
5. 위쪽 또는 아래쪽 화살표를 사용하여 값이 포함된 다른 행으로 이동합니다.

레이블 할당으로 참조 데이터 필터링

레이블 보기를 사용하여 지정한 레이블을 사용하는 참조 데이터 값을 찾습니다. 레이블을 할당하는 데 사용한 방법을 기준으로 결과를 필터링합니다.

1. 확률 모델이 포함된 콘텐츠 집합을 엽니다.
2. 모델 이름을 선택하고 **편집**을 클릭합니다.
3. 레이블 보기에서 레이블 값을 선택합니다.
확률 모델이 레이블을 사용하는 참조 데이터 값의 목록을 표시합니다. 해당 모델은 레이블을 사용하는 데이터 값의 수도 표시합니다.
4. 필터를 레이블을 사용하는 참조 데이터 값 목록에 적용합니다.
다음 필터 중 하나를 선택합니다.
 - 모두 레이블을 사용하는 참조 데이터 값을 표시합니다. 모두가 기본 옵션입니다.
 - 사용자가 할당함. 레이블을 할당했을 때 개별적으로 선택한 참조 데이터 값을 표시합니다.
 - 대량으로 할당됨. 레이블을 대량 할당 작업의 일부로 할당한 참조 데이터 값을 표시합니다.확률 모델은 필터 조건을 충족하는 참조 데이터 값을 표시합니다.

사용되지 않은 레이블 값 찾기

참조 데이터 값에 할당하지 않은 레이블 값을 찾으려면 레이블 보기를 사용합니다. 각 레이블을 최소 하나의 참조 데이터 값에 할당해야 합니다.

1. 확률 모델이 포함된 콘텐츠 집합을 엽니다.
2. 모델 이름을 선택하고 **편집**을 클릭합니다.

- 레이블 보기에서 레이블 값을 선택합니다.

확률 모델이 레이블을 사용하는 참조 데이터 값의 목록을 표시합니다. 레이블을 사용하는 데이터 값의 총 수도 표시됩니다.

데이터 값의 총 수가 0이라면 확률 모델의 어떤 참조 데이터 값에도 레이블을 할당하지 않은 것입니다.

작업 복사 및 붙여넣기

한 콘텐츠 집합에서 모델 리포지토리의 다른 콘텐츠 집합으로 확률 모델을 복사할 수 있습니다. 다른 Developer tool 사용자와 리소스를 공유할 수 있도록 확률 모델을 복사합니다.

모델을 다른 콘텐츠 집합으로 복사하거나 모델을 현재 콘텐츠 집합으로 가져올 수 있습니다. 한 번의 작업에서 리포지토리의 여러 콘텐츠 집합에서 여러 모델을 가져올 수 있습니다.

모델을 복사하면 콘텐츠 관리 서비스가 Informatica 서비스 호스트 시스템에 모델 데이터 파일의 복사본을 작성합니다. 각 모델은 서로 다른 데이터 파일을 사용합니다.

확률 모델을 다른 콘텐츠 집합에 복사

한 콘텐츠 집합에서 모델 리포지토리의 다른 콘텐츠 집합으로 확률 모델을 복사할 수 있습니다. 확률 모델을 복사할 때 모델 개체와 소스 및 대상 콘텐츠 집합을 지정합니다.

- 확률 모델이 포함된 콘텐츠 집합을 엽니다.
- 확률 모델을 선택하고 **복사 대상**을 클릭합니다.
- 모델 리포지토리를 탐색하고 콘텐츠 집합을 선택합니다.

확률 모델을 현재 프로젝트 또는 다른 프로젝트의 콘텐츠 집합에 복사할 수 있습니다.

- 확인**을 클릭합니다.

Developer 도구가 선택한 콘텐츠 집합에 확률 모델을 복사합니다.

다른 콘텐츠 집합에서 확률 모델 가져오기

한 콘텐츠 집합에서 모델 리포지토리의 다른 콘텐츠 집합으로 확률 모델을 가져올 수 있습니다. 확률 모델을 가져올 때 하나 이상의 모델 개체와 소스 및 대상 콘텐츠 집합을 지정합니다.

- 확률 모델이 포함된 콘텐츠 집합을 엽니다.
- 확률 모델을 선택하고 **붙여넣기 원본 위치**를 클릭합니다.
- 모델 리포지토리를 탐색하고 확률 모델을 선택합니다.

현재 프로젝트 또는 다른 프로젝트의 콘텐츠 집합에서 확률 모델을 붙여넣을 수 있습니다.

- 확인**을 클릭합니다.

Developer 도구가 확률 모델을 현재 콘텐츠 집합에 붙여넣습니다.

클립보드에 참조 데이터 행 복사

확률 모델에서 클립보드로 하나 이상의 참조 데이터 행을 복사할 수 있습니다. 다른 확률 모델에 행을 붙여 넣을 수 있습니다.

- 확률 모델이 포함된 콘텐츠 집합을 엽니다.

2. 모델 이름을 선택하고 **편집**을 클릭합니다.
 3. 데이터 보기에서 하나 이상의 참조 데이터 행을 선택합니다.
 4. **Ctrl + C** 키를 사용하여 행을 클립보드에 복사합니다.
참조 데이터와 참조 데이터에 할당한 레이블 값이 복사됩니다.
- Ctrl + V** 키를 사용하여 텍스트 편집기 또는 다른 확률 모델의 데이터 보기에 행을 붙여 넣을 수 있습니다.

부록 A

참조 데이터 및 Informatica Data Engineering Integration

이 부록에 포함된 항목:

- [참조 데이터 및 Informatica Data Engineering Integration, 66](#)

참조 데이터 및 Informatica Data Engineering Integration

Informatica Data Engineering Integration은 Informatica 도메인과 Informatica 클라이언트 응용 프로그램을 Hadoop 클러스터와 결합합니다. Developer tool의 매핑을 클러스터로 푸시다운하고 클러스터의 노드에서 해당 매핑을 실행할 수 있습니다.

푸시다운 작업은 매핑이 읽는 모든 참조 테이블 데이터, 콘텐츠 집합 데이터 및 ID 채우기 데이터를 클러스터에 복사합니다. 푸시다운 작업에서 매핑과 함께 복사된 참조 데이터는 매핑이 실행된 후 클러스터에서 삭제됩니다.

주소 유효성 검사 참조 데이터는 푸시다운 작업에서 복사되지 않습니다. 주소 유효성 검사를 수행하는 매핑을 푸시하는 경우 매핑을 실행하는 각 노드에 주소 유효성 검사 참조 데이터 파일을 설치해야 합니다. 주소 유효성 검사 참조 데이터 파일은 주소 유효성 검사 매핑이 실행된 후 클러스터에서 삭제되지 않습니다.

주소 유효성 검사를 위한 참조 데이터

Hadoop 환경에서 주소 유효성 검사 매핑을 실행하는 경우 주소 참조 데이터 파일은 매핑이 실행되는 각 DataNode에 상주해야 합니다. Informatica Data Engineering Integration은 DataNode에 파일을 설치할 때 사용할 수 있는 셸 스크립트와 함께 설치됩니다.

셸 스크립트를 사용하여 단일 작업으로 주소 참조 데이터 파일을 DataNode에 설치합니다. 스크립트는 노드의 이름 또는 IP 주소가 포함된 파일을 읽습니다. 스크립트는 주소 참조 데이터 파일을 파일이 식별하는 각 노드에 복사합니다.

스크립트 이름은 `copyRefDataToComputeNodes.sh`입니다.

Data Engineering Integration 설치의 다음 디렉터리에서 이 스크립트를 찾을 수 있습니다.

<Informatica 설치 디렉터리>/tools/dq/av

다음 테이블에는 스크립트에 사용되는 옵션이 설명되어 있습니다.

옵션	설명
-n	Hadoop 클러스터에 있는 DataNode의 이름 또는 IP 주소 목록을 포함하는 파일입니다. 각 노드 이름 또는 IP 주소를 파일의 개별 줄에 입력합니다. 기본적으로 스크립트는 \$BASEDIR/HadoopDataNodes 디렉터리의 파일을 읽습니다. 여기서 \$BASEDIR이 셸 스크립트의 위치입니다.
-p	주소 참조 데이터 파일을 설치할지 확인하는 프롬프트입니다. 기본적으로 스크립트는 소스 디렉터리에서 DataNode의 대상 디렉터리로 파일을 복사할 것인지 확인하는 프롬프트를 표시합니다. 일정에 따라 셸 스크립트를 실행하는 경우 프롬프트를 비활성화할 수 있습니다. 기본 옵션 값은 Y입니다. 프롬프트를 비활성화하려면 값을 N으로 설정합니다.
-s	스크립트가 노드로 복사하는 주소 참조 데이터 파일의 소스 디렉터리입니다. 기본적으로 스크립트는 로컬 시스템의 /reference_data 디렉터리에서 파일을 읽습니다. 참고: 주소 참조 데이터 파일은 .MD 파일 이름 확장명을 사용합니다. 소스 디렉터리에는 주소 참조 데이터 파일이 포함되어야 하며 다른 파일은 없어야 합니다.
-t	스크립트가 주소 참조 데이터 파일을 복사하는 각 노드의 디렉터리입니다. 기본적으로 스크립트는 각 노드의 /reference_data 디렉터리에 파일을 복사합니다.
-u	스크립트를 실행하는 사용자의 사용자 이름입니다. 사용자는 암호를 사용하지 않는 보안 셸로 노드에 액세스할 수 있어야 합니다.

주소 참조 데이터 파일 설치

Hadoop 클러스터의 DataNode에 주소 참조 데이터 파일을 설치하려면 `copyRefDataToComputeNodes.sh` 셸 스크립트를 실행합니다. 또는 작업 스케줄러 응용 프로그램에서 지정된 간격으로 셸 스크립트를 실행하는 작업을 정의합니다.

스크립트를 실행하거나 작업을 정의하기 전에 스크립트에 지정된 옵션 값을 검토합니다. 기본값을 그대로 사용하거나 값을 업데이트할 수 있습니다.

명령 프롬프트에서 주소 참조 데이터 파일 설치

명령 프롬프트에서 파일을 설치하려면 다음 단계를 수행하십시오.

1. 명령 프롬프트에서 다음 디렉터를 엽니다.
<Informatica 설치 디렉터리>/tools/dq/av
2. `copyRefDataToComputeNodes.sh`를 실행합니다.
필요에 따라 스크립트 옵션에 대한 하나 이상의 값을 입력합니다. 옵션에 대한 값을 입력하지 않으면 옵션에 대한 기본값을 사용하여 스크립트가 실행됩니다.
기본적으로 파일의 설치를 확인하라는 메시지가 표시됩니다. 파일을 설치하려면 Y를 입력합니다.

예약된 작업을 사용하여 주소 참조 데이터 파일 설치

지정한 시간 간격으로 셸 스크립트를 실행하는 작업을 정의할 수 있습니다. 작업 스케줄러 응용 프로그램에 작업을 추가합니다. 파일을 설치하는 작업을 정의할 경우 설치 확인 프롬프트를 비활성화해야 합니다.

프롬프트를 비활성화하려면 셸 스크립트에 다음 옵션을 설정합니다.

-p n

인덱스

A

Analyst 도구

참조 데이터 값 찾기 및 바꾸기 [22](#)

B

패턴 집합 [35](#)

프로필 열 데이터에서 참조 테이블 작성

참조 테이블 [16](#)

행 관리

참조 테이블 [22](#)

확률 모델

규칙 및 지침 [37](#)

콘텐츠 집합 [35](#)

D

Data Engineering Integration

주소 참조 데이터 설치 [67](#)

주소 참조 데이터 설치 스크립트 [66](#)

참조 데이터 요구 사항 [66](#)

H

Hadoop 환경

주소 참조 데이터 설치 [67](#)

주소 참조 데이터 설치 스크립트 [66](#)

참조 데이터 요구 사항 [66](#)

N

문자 집합 [34](#)

버전 제어

Analyst 도구의 참조 테이블 [20](#)

Developer tool의 참조 테이블 [28](#)

참조 테이블 [12](#)

콘텐츠 집합 [12, 28](#)

분류자 모델

규칙 및 지침 [37](#)

콘텐츠 집합 [34](#)

수동으로 참조 테이블 작성

참조 테이블 [15](#)

열 관리

참조 테이블 [21](#)

열 패턴을 사용하여 참조 테이블 작성

참조 테이블 [17](#)

정규식 [35](#)

참조 테이블

Analyst 도구 개요 [13](#)

참조 테이블 (계속)

Analyst 도구에서 값 찾기 및 바꾸기 [22](#)

Analyst 도구에서 새로 고침 [24](#)

Analyst 도구의 속성 [13](#)

Developer tool 개요 [29](#)

Developer tool의 속성 [29](#)

감사 내역 테이블 보기 [25](#)

관리 및 비관리 [11](#)

관리되는 참조 테이블 [11](#)

관리되지 않는 참조 테이블 [11](#)

권한 [12](#)

버전 제어 [12, 20, 28](#)

수동으로 참조 테이블 작성 [15](#)

열 관리 [21](#)

열 패턴을 사용하여 참조 테이블 작성 [17](#)

참조 데이터 웨어하우스 [11](#)

참조 테이블 가져오기 [18](#)

참조 테이블 내보내기 [23](#)

콘텐츠 관리 서비스 [11](#)

패턴 기반 구문 분석 [11](#)

프로필 열을 사용하여 참조 테이블 작성 [16](#)

행 관리 [22](#)

참조 테이블 가져오기

참조 테이블 [18](#)

참조 테이블 내보내기

참조 테이블 [23](#)

콘텐츠 관리 서비스

참조 데이터 웨어하우스 [11](#)

참조 테이블 권한 [12](#)

콘텐츠 집합

문자 집합 [34](#)

버전 제어 [12, 20, 28](#)

분류자 모델 [34](#)

정규식 [35](#)

토큰 집합 [36](#)

패턴 집합 [35](#)

확률 모델 [35](#)

토큰 집합 [36](#)



감사 테이블 이벤트 보기

참조 테이블 [25](#)

관리되는 참조 테이블 [11](#)

관리되지 않는 참조 테이블

모델 리포지토리와 동기화 [12](#)

정의 [11](#)

편집 활성화 및 비활성화 [24](#)

권한

콘텐츠 관리 서비스 [12](#)