SE305 COURSE PROJECT FINAL REPORT

Jeremy Liu, Qianyang Peng, Jingyu Cui

Contents

Abstract

With the developing ofacknowledgement.

1 ER Models

Our system is

2 Table Designs

We finished out table design according to our ER model. Our final design contains 9 tables. Reference information is not stored in our tables, while all other information are stored. The design of our tables is elaborated below:

2.1 entity

Table schema:

```
CREATE TABLE IF NOT EXISTS `wikidata`.`
    entity` (
  `serial_id` BIGINT(32) NOT NULL
     AUTO_INCREMENT,
  `entity_id` VARCHAR(32) NOT NULL,
  `entity_language` VARCHAR(16) NOT NULL
     ,
  `entity_type` VARCHAR(16) NULL DEFAULT
      NULL,
  `entity_text` VARBINARY(255) NULL
     DEFAULT NULL,
  PRIMARY KEY (`serial_id`),
  INDEX `EID` (`entity_id` ASC),
  INDEX `ELANG` (`entity_language` ASC),
  INDEX `ETYPE` (`entity_type` ASC))
ENGINE = InnoDB
```

The data strored in corresponding fields:

- **serial_id**: Auto incremental serial ID. As language information is introduced here and entity ID can not be used as Primary Key, so we use this serial_id here as a Primary Key.
- **entity_id**: ID of the corresponding entity. Eg. Q5, P110, Q123423, etc.

- **entity_language**: Language of the label of the corresponding entity. Eg. Zh-cn, en, etc.
- **entity_type**: The entity type identifier. ”item” for data items, and ”property” for properties.
- **entity_text**: Contains the labels in different languages. Eg. square kilometre, Kilometro quadrato, Vierkante kilometer, etc.

Indexed tables: besides indexing the Primary Key serial_id, other keys including entity_id, entity_language and entity_type are also indexed.

2.2 description

Table schema:

```
CREATE TABLE IF NOT EXISTS `wikidata`.`
    description` (
  `serial_id` BIGINT(32) NOT NULL
     AUTO_INCREMENT,
  `entity_id` VARCHAR(32) NOT NULL,
  `desc_language` VARCHAR(8) NULL,
  `desc_text` VARBINARY(255) NULL,
  PRIMARY KEY (`serial_id`),
  INDEX `EID` (`entity_id` ASC),
  INDEX `DLANG` (`desc_language` ASC))
ENGINE = InnoDB
```

The data strored in corresponding fields:

- **serial_id**: Same as in table ’entity’.
- **entity_id**: Same as in table ’entity’.
- **desc_language**: Similar to entity_language in table ’entity’.
- **desc_text**: Similar to entity_text in table ’entity’.

Indexed tables: besides indexing the Primary Key serial_id, other keys including entity_id, desc_language and are also indexed.

2.3 mainsnak

Table schema:

```
CREATE TABLE IF NOT EXISTS `wikidata`.`
    mainsnak` (
  `snak_id` VARCHAR(64) NOT NULL,
  `entity_id` VARCHAR(32) NOT NULL,
  `property_id` VARCHAR(32) NOT NULL,
  `serial` INT(4) NOT NULL,
  `claimtype` VARCHAR(32) NULL DEFAULT
     NULL,
  `snaktype` VARCHAR(32) NULL DEFAULT
     NULL,
  `datatype` VARCHAR(32) NULL DEFAULT
     NULL,
  `rank` VARCHAR(32) NULL DEFAULT NULL,
  PRIMARY KEY (`snak_id`),
  INDEX `EID` (`entity_id` ASC),
  INDEX `PID` (`property_id` ASC),
  INDEX `CTYPE` (`claimtype` ASC),
  INDEX `STYPE` (`snaktype` ASC),
  INDEX `DTYPE` (`datatype` ASC))
ENGINE = InnoDB
```

The data strored in corresponding fields:

- **snak_id**: An arbitrary identifier for the claim, used as Primary Key of this table.
- **entity_id**: ID of the entity this claim belongs to.
- **property_id**: The property this claim is describing.
- **serial**: This field is X means this is the Xth claim of this property.
- **claimtype**: the type of the claim - currently either statement or claim.
- **snaktype**: The type of the snak. Currently, this is one of value, somevalue or novalue.
- **datatype**: The datatype field indicates how the value of the Snak can be interpreted. This fields indicate the concrete table begining with ”datavalue_”
- **rank**: The rank expresses whether this value will be used in queries, and shown be visible per default on a client system. The value is either preferred, normal or deprecated.

2.4 datavalue_string

Table schema:

```
CREATE TABLE IF NOT EXISTS `wikidata`.`
    datavalue_string` (
  `snak_id` VARCHAR(64) NOT NULL,
  `value` VARBINARY(255) NULL DEFAULT
     NULL,
  PRIMARY KEY (`snak_id`))
ENGINE = InnoDB
```

The data strored in corresponding fields:

- **snak_id**: Primary key. Refering to the table mainsnak.
- **value**: The content of the string.

2.5 datavalue_time

Table schema:

```
CREATE TABLE IF NOT EXISTS `wikidata`.`
    datavalue_time` (
  `snak_id` VARCHAR(64) NOT NULL,
  `time` VARCHAR(64) NULL DEFAULT NULL,
  `timezone` VARCHAR(32) NULL DEFAULT
      NULL,
  `before` VARCHAR(32) NULL DEFAULT NULL
      ,
  `after` VARCHAR(32) NULL DEFAULT NULL,
  `precision` INT(8) NULL DEFAULT NULL,
  `calendarmodel` VARCHAR(255) NULL
      DEFAULT NULL,
  PRIMARY KEY (`snak_id`))
ENGINE = InnoDB
```

The data strored in corresponding fields:

- **snak_id**: Primary key. Refering to the table mainsnak.
- **time**: The content of the string.
- **timezone**: Signed integer. Now unused.
- **before**: Begin of an uncertainty range, given in the unit defined by the precision field. This cannot be used to represent a duration.
- **after**: End of an uncertainty range, given in the unit defined by the precision field. This cannot be used to represent a duration.
- **precision**: To what unit is the given date/time significant.
- **calendarmodel**: A URI of a calendar model, such as gregorian or julian. Typically given as the URI of a data item on the repository.

2.6 datavalue_globecoordinate

Table schema:

```
CREATE TABLE IF NOT EXISTS `wikidata`.`
    datavalue_globecoordinate` (
  `snak_id` VARCHAR(64) NOT NULL,
  `latitude` FLOAT NULL DEFAULT NULL,
  `longitude` FLOAT NULL DEFAULT NULL,
  `altitude` FLOAT NULL DEFAULT NULL,
  `precision` FLOAT NULL DEFAULT NULL,
  `globe` VARCHAR(255) NULL DEFAULT NULL
      ,
  PRIMARY KEY (`snak_id`))
ENGINE = InnoDB
```

The data strored in corresponding fields:

- **snak_id**: Primary key. Refering to the table mainsnak.
- **latitude**: The latitude part of the coordinate in degrees, as a float literal (or an equivalent string).
- **longitude**: The longitude part of the coordinate in degrees, as a float literal (or an equivalent string).
- **altitude**: Deprecated and no longer used. Will be dropped in the future.
- **precision**: the coordinate's precision, in (fractions of) degrees, given as a float literal (or an equivalent string).
- **precision**: To what unit is the given date/time significant.
- **globe**: the URI of a reference globe.

2.7 datavalue_quantity

Table schema:

```
CREATE TABLE IF NOT EXISTS `wikidata`.`
    datavalue_quantity` (
  `snak_id` VARCHAR(64) NOT NULL,
  `amount` VARCHAR(64) NULL DEFAULT NULL
      ,
  `upperBound` VARCHAR(64) NULL DEFAULT
      NULL,
  `lowerBound` VARCHAR(64) NULL DEFAULT
      NULL,
  `unit` VARCHAR(64) NULL DEFAULT NULL,
  PRIMARY KEY (`snak_id`))
ENGINE = InnoDB
```

The data strored in corresponding fields:

- **snak_id**: Primary key. Refering to the table mainsnak.
- **amount**: The nominal value of the quantity, as an arbitrary precision decimal string. The string always starts with a character indicating the sign of the value, either "+" or "-".

- **upperBound**: Optionally, the upper bound of the quantity's uncertainty interval, using the same notation as the amount field.
- **lowerBound**: Optionally, the lower bound of the quantity's uncertainty interval, using the same notation as the amount field.
- **unit**: the URI of a unit (or "1" to indicate a unit-less quantity).

2.8 datavalue_wikibase

Table schema:

```
CREATE TABLE IF NOT EXISTS `wikidata`.`
    datavalue_wikibase` (
  `snak_id` VARCHAR(64) NOT NULL,
  `id` VARCHAR(32) NULL DEFAULT NULL,
  PRIMARY KEY (`snak_id`))
ENGINE = InnoDB
```

The data strored in corresponding fields:

- **snak_id**: Primary key. Refering to the table mainsnak.
- **id**: defines the id of the entity.

2.9 qualifier

Table schema:

```
CREATE TABLE IF NOT EXISTS `wikidata`.`
    qualifier` (
  `serial_id` BIGINT(32) NOT NULL
     AUTO_INCREMENT,
  `hash` VARCHAR(64) NULL DEFAULT NULL,
  `snaktype` VARCHAR(32) NULL DEFAULT
     NULL,
  `property_id` VARCHAR(32) NULL DEFAULT
      NULL,
  `datatype` VARCHAR(32) NULL DEFAULT
     NULL,
  INDEX `STYPE` (`snaktype` ASC),
  INDEX `PID` (`property_id` ASC),
  INDEX `DTYPE` (`datatype` ASC),
  PRIMARY KEY (`serial_id`))
ENGINE = InnoDB
```

The data strored in corresponding fields:

- **serial_id**: Primary key.
- **hash**: Hash value, may be duplicated.
- **snaktype**: Similar to the table mainsnak.
- **property_id**: Similar to the table mainsnak.
- **datatype**: Similar to the table mainsnak.

3

This project i

4 ACKNOWLEDGEMENT

Our projec