

Table 1. Core schema implementation in a relational database.

Table Name	Description	Field Name	Data Type	Definition
ETDs	Core information about the ETD	id	INT ¹ (pk)	Unique identifier for the ETD in the file system
		title	VARCHAR ² (255)	Title of the ETD
		author	VARCHAR(255)	Author of the ETD
		committeeChair	VARCHAR(255)	Committee Chair
		committeeMember	VARCHAR(255)	Committee Member
		year	INT	Year of publication
		abstract	TEXT ³	Abstract of the ETD
		abstractgeneral	TEXT	General abstract of the ETD
		institution	VARCHAR(255)	Institution, e.g., degree granting body like official name of a university
		degree	VARCHAR(255)	Degree awarded (e.g., Masters, PhD)
		landing	VARCHAR(255)	URI or URL for the landing page in the repository that provides access to this ETD
		department	VARCHAR(255)	Department or program of the institution responsible for the degree
		language	VARCHAR(50)	Language of the ETD
		type	VARCHAR(50)	The category of the content of the resource (e.g., Thesis, Dissertation)
		publisher	VARCHAR(255)	Publisher of the ETD
		country	VARCHAR(100)	Country of the institution
		owner_and_statement	TEXT	Copyright and ownership information
		degree_level	VARCHAR(50)	Level of the degree, e.g., Master's, Doctoral
		grantor	VARCHAR(255)	Granting institution/agency
		license	VARCHAR(255)	License information
Rights	Copyright and access details for the ETD	id	INT (pk)	Unique identifier for the rights record
		etdId	INT	Foreign key referencing the ETD (from ETDs.id)
		rights_text	TEXT	Full text of the copyright statement
		rights_type	VARCHAR(50)	Type of rights (e.g., Open Access)
		rights_id	VARCHAR(100)	External identifier for the rights
		rights_date	TIMESTAMP ⁴	Date the rights information was applied

¹ INT: Used for storing integer values, such as in id and year fields. INT (pk) denotes a primary key, uniquely identifying each record.

² VARCHAR: Stores variable-length text strings, with the number indicating the maximum characters allowed (e.g., VARCHAR(255) for titles).

³ TEXT: The TEXT data type is used for longer pieces of text where the length might exceed the typical limits of VARCHAR.

⁴ TIMESTAMP: Records date and time values, useful for tracking events like rights information updates.

Table Name	Description	Field Name	Data Type	Definition
ETD_File	Information of the ETD file	id	INT (pk)	Unique identifier for the ETD file
		etdId	INT	Foreign key referencing the ETD (from ETDs.id)
		description	TEXT	Description of the file
		scanned	BIT ⁵	A flag value that indicates whether the ETD is scanned or not
		localRelPath	VARCHAR(255)	Relative path to the file within the system
		checksum	VARCHAR(255)	Checksum for verifying file integrity
		mimeType	VARCHAR(100)	MIME type of the file (e.g., Video, Audio)
		date_downloaded	TIMESTAMP	Date the file was downloaded
References	References cited within the ETD	id	INT (pk)	Unique identifier for the reference
		etdId	INT	Foreign key referencing the ETD (from ETDs.id)
		reference_text	TEXT	Full text of the reference citation
		author	VARCHAR(255)	Author of the cited publication
		title	VARCHAR(255)	Title of the cited publication
		year	INT	Year of the cited publication
		venue	VARCHAR(255)	Venue of the cited publication
Text	Information on the text content of each ETD chapter	id	INT (pk)	Unique identifier for the text segment
		etdId	INT	Foreign key referencing the ETD (from ETDs.id)
		chapterid	INT	Chapter number
		localRelPath	VARCHAR(255)	Relative path to the text file within the system
		text	TEXT	Content of the ETD chapter
Subjects	Information about the topic of the content of the ETD	id	INT (pk)	Unique identifier for the subject
		etdid	INT	Foreign key referencing the ETD (from ETDs.id)
		subject	VARCHAR(255)	Topic of the content of ETD
		source	VARCHAR(255)	Subject were determined by author/ProQuest

⁵ BIT: The BIT data type stores binary values, representing 0 (false) or 1 (true).

Table 2. Extended Schema Information implementation in a relational database.

Table Name	Description	Field Name	Data Type	Definition
Classification_entries	Holds the entries for various classification systems	id	INT	Identifier for one (subject category) entry in a Classification
		classificationId	INT	Identifier of the Classification in which this entry fits, e.g., the ID for ProQuest
		name	VARCHAR(255)	Name of this entry in the Classification; e.g., Library Science
		description	TEXT	Description of this entry, e.g., Library and Information Science and Studies
		depth	INT	Depth in the Classification taxonomy, e.g., 3, with 1 used for the set of top level entries
Classifications	Contains details about various classification systems	id	INT	Identifier for one entry among the possible Classification systems
		name	VARCHAR(255)	Name of this Classification system, e.g., ProQuest
		description	TEXT	Description of the Classification system, e.g., ProQuest Subject Categories of 2022-2023 Academic Year
Classifiers	Stores classifier models details	id	INT	Identifier for one Classifier model
		name	VARCHAR(255)	Name of this Classifier
		description	TEXT	Description of this Classifier, e.g., SVM with the following implementation details and parameters
		uri	VARCHAR(255)	File system path for the Classifier model
		classificationId	INT	Identifier in the Classifications table of the Classification system for which the model was trained
Collection_topics	Contains topics that describe a collection using a particular Topic_model	id	INT	Identifier for one Topic among the set of topics that describe a Collection using a particular Topic_model
		collectionId	INT	Identifier of the Collection
		topicModelId	INT	Identifier of the Topic_model
		term_list	TEXT	List of terms describing this Topic
		probability_list	TEXT	Probabilities for the terms in the Term_list

Table Name	Description	Field Name	Data Type	Definition
Collections	Holds details of ETD collections for processing	id	INT	Identifier for one of the Collections of ETDs for which some processing is defined
		name	VARCHAR(255)	Name for this Collection, e.g., Scanned ETDs from USA
		description	TEXT	Description of this Collection, e.g., All ETDs resulting from scanning of each page of an ETD
		etds_list	TEXT	JSON list of the IDs of the ETDs in this Collection
ETD_classes	Stores class details for ETDs according to a particular Classification	id	INT	Identifier for the list of classes describing an ETD according to a particular Classification
		classificationId	INT	Identifier for the Classification used
		etdId	INT	Identifier for an ETD in the file system
		classifierId	INT	Identifier of the classifier
		class_list	TEXT	List of Classes describing this ETD in this Classification
		probability_list	TEXT	Probabilities for the Classes in the Class_list
ETD_topics	Holds topics assigned to ETDs in a Collection from a Topic Model	id	INT	Identifier for the set of Topics for an ETD in a Collection for a Topic Model
		etdId	INT	Identifier for an ETD in the file system
		collectionId	INT	Identifier of the Collection of ETDs whose analysis led to a Collection Topic Model
		topicModelId	INT	Identifier of the Topic Model
		topic_list	TEXT	List of topics assigned to an ETD in a Collection from among the Collection_topics
		probability_list	TEXT	Probabilities for the topics in the list of topics for this ETD

Table Name	Description	Field Name	Data Type	Definition
ETD-ETD_neighbors	Specifies the nearest neighboring ETDs in a Collection	id	INT	Identifier for one entry for an ETD that specifies the nearest neighboring ETDs in a Collection
		etdId	INT	Identifier for this ETD
		collectionId	INT	Identifier for a Collection that includes this ETD
		neighbor_list	TEXT	List of ETD ids in a Collection that are the nearest neighbors to an ETD
		probability_list	TEXT	Probabilities for the neighbors in the Neighbor_list
Object_classes	Describes classes for a text Object of an ETD according to a Classification	id	INT	Identifier for the list of classes describing a text Object of an ETD according to a particular Classification
		classificationId	INT	Identifier for the Classification used
		classifierId	INT	Identifier of the classifier used
		objectId	INT	Identifier for a text Object of an ETD
		class_list	TEXT	List of classes describing this ETD in this Classification
		probability_list	TEXT	Probabilities for the classes in the class_list
Object_metadata	Stores metadata records for an Object	id	INT	Identifier for a Metadata record for an Object
		objectId	INT	Identifier for the Object
		metadata_type	VARCHAR(255)	Type of Metadata scheme used for this Object
		metadata_record	TEXT	Record of the Metadata for this Object for this Metadata_type
Object_summaries	Holds summaries for Objects	id	INT	Identifier for one Summary for an Object
		objectId	INT	Identifier for the Object being summarized
		summarizerId	INT	Identifier for the Summarizer applied to the Object
		summary	TEXT	Text of the resulting Summary

Table Name	Description	Field Name	Data Type	Definition
Object_topics	Stores topics for a text Object derived from an ETD.	id	INT	Identifier for the set of Topics for a text Object derived from an ETD, for a Topic Model
		etdId	INT	Identifier for an ETD in the file system
		collectionId	INT	Identifier of the Collection of ETDs whose analysis led to a Collection Topic Model
		topicModelId	INT	Identifier of the Collection Topic Model
		topic_list	TEXT	List of Topics assigned to an Object in a Collection from among the Collection_topics
		probability_list	TEXT	Probabilities for the Topics in the list of topics for this Object
Object-object_neighbors	Holds summaries for Objects.	id	INT	Identifier for one entry for an Object that specifies the nearest neighboring Objects in a Collection
		objectId	INT	Identifier for an Object of an ETD of a Collection
		collectionId	INT	Identifier for the Collection that includes this Object's ETD
		neighbor_list	TEXT	List of Object_IDs in a Collection that are the nearest neighbors to an Object
		probability_list	TEXT	Probabilities for the neighbors in the Neighbor_list
Objects	Holds details about objects derived from ETDs.	id	INT	Identifier for an Object derived (e.g., through Object Detection) from an ETD
		etdId	INT	Identifier for this ETD
		object_type	VARCHAR(100)	Type of Object, e.g., Figure
		object_format	VARCHAR(100)	Representation, e.g., Text, Image
		object_text	TEXT	Text content of the Object, only if Object_format=Text
		object_path	VARCHAR(255)	File system path for the Object, only if Object_format=Image

Table Name	Description	Field Name	Data Type	Definition
Topic_models	Stores details about topic modeling implementations	id	INT	Identifier for one implementation of Topic Modeling
		name	VARCHAR(255)	Name, e.g., LDA
		description	TEXT	Textual explanation, e.g., Latent Dirichlet Allocation
		path	VARCHAR(255)	File system path for the Topic Model
		type	VARCHAR(255)	Approach, e.g., Statistical, Deep Learning
		parameters	TEXT	Set of parameters for this type, e.g., hyperparameter attribute-value pairs
User_classes	Stores user interests according to a particular Classification	id	INT	Identifier for the set of Classes of interest to a User according to a particular Classification
		users_Id	INT	Identifier for this User
		classificationId	INT	Identifier of the Classification used
		class_list	TEXT	List of Classes describing this User in this Classification
User_queries	Stores user queries and their results	id	INT	ID for the Query from a User and the Results
		users_Id	INT	Identifier for this User
		query_results_ETDs	TEXT	Ordered list of ETDs returned to this User for this Query
		query_results_Objects	TEXT	Ordered list of Objects returned to this User for this Query
		query_text	TEXT	Text of a query issued by this User
		time_stamp	TIMESTAMP	Date and time that query was issued

Table Name	Description	Field Name	Data Type	Definition
User_queries_clicks	Records clicks on user queries results	users_Id	INT	Identifier for this User
		user_queries_Id	INT	Identifier for a record for a Query issued by this User
		query_results_ETDs_clicked	TEXT	Ordered subset list from Query_results_ETDs that had clicks
		query_results_Objects_clicked	TEXT	Ordered subset list from Query_results_Objects that had clicks
User_topics	Stores topics of interest to a user	id	INT	Identifier for the set of Topics of interest to a User
		users_Id	INT	Identifier for this User
		collectionId	INT	Identifier of the Collection of ETDs whose analysis led to a Collection Topic Model
		topicModelId	INT	Identifier of the Collection Topic Model
		topic_list	TEXT	List of topics of interest to a User from among the Collection_topics
User-user_neighbors	Stores details about nearest neighboring users	id	INT	Identifier for one entry for a User that specifies the nearest neighboring Users
		users_Id	INT	Identifier for this User
		neighbor_list	TEXT	List of User_IDs that are the nearest neighbors to this User
		probability_list	TEXT	Probabilities for the neighbors in the Neighbor_list
Users	Stores user details	id	INT	Identifier serving as key for this User
		firstName	VARCHAR(255)	Names preceding the surname
		lastName	VARCHAR(255)	Surname
		emailAddress	VARCHAR(255)	Full electronic mail address
		affiliation	VARCHAR(255)	Institution affiliated with, e.g., Old Dominion University
		roles_list	TEXT	List of Roles for this User, e.g., Curator, Experimenter, User
		passwordHash	VARCHAR(255)	Hash of the password to check during return login

Table Name	Description	Field Name	Data Type	Definition
Summarizers	Contains details about various summarization implementations	id	INT	Identifier for one implementation of Summarization
		name	VARCHAR(255)	Name, e.g., TextRank
		description	TEXT	Textual explanation, e.g., Applying PageRank-like methods to sentences to create a summary
		type	VARCHAR(255)	Approach, e.g., Extractive, Abstractive
		parameters	TEXT	Set of parameters for this type, e.g., hyperparameter attribute-value pairs
		path	VARCHAR(255)	File system path for the Summarizer model