

EDHIΦY V1.2 DOCUMENTATION

This document describes the structure and mutual relationships between the tables of EDHIΦY (SQLite implementation), v1.2.

PHILOSOPHER

Each row in the table corresponds to a philosopher covered in EDHIPHY.

Field	Description	Type	Default	Other
Phi_ID	Unique identifier of the philosopher in EDHIPHY. 'P:1 [No Author]' is a special author used for JSTOR articles that do not have an author	TEXT		PK
Phi_Standard_Label	Standard label of the philosopher	TEXT		
Wiki_ID	Link to the WikiData page, when available. The ID in WikiData begins with "Q"	TEXT		FK
Wiki	1 if the philosopher is in WikiData, 0 otherwise	INTEGER		
JSTOR	1 if the philosopher is an author in JSTOR, 0 otherwise	INTEGER		
Hiring	1 if the philosopher is in the Strassfeld hiring dataset, 0 otherwise	INTEGER		
PHD	1 if the philosopher's dissertation is available in the ProQuest dataset, 0 otherwise	INTEGER		

BIOGRAPHY

Standardized biographical data of the philosophers, extracted by multiple sources.

Field	Description	Type	Default	Other
-------	-------------	------	---------	-------

Field	Description	Type	Default	Other
Phi_ID	Unique philosopher identifier. Biography data are present only for those philosophers that have at least one of the properties in the table	TEXT		PK, FK
Surname	Surname obtained by Hiring data, PhD data, Wiki Data or automatic extraction from JSTOR author string, with the following priority: Hiring > PhD > Wiki > artificial JSTOR surname. If in WikiData more than one surname was recorded, the multiple surnames were merged into one	TEXT		
First_Name	First name obtained by PhD data, Hiring data or Wiki Data, with the following priority: PhD > Hiring > Wiki. If in WikiData more than one first name was recorded, the multiple names were merged into one	TEXT		
Gender	Gender recorded in WikiData	TEXT		
BirthYear	Birth year as recorded in WikiData. When more than one birth year was recorded, only the first one is selected. Can be negative when it is B.C.	INTEGER		

VARIANT

Variants of philosophers' names. This table is used to generate the Entity Ruler that is employed in the NER.

Field	Description	Type	Default	Other
Variant_ID	Unique identifier of the variant	TEXT		PK
Phi_ID	Identifier of the philosopher to which the variant is associated	TEXT		FK

Field	Description	Type	Default	Other
Variant	String of the variant	TEXT		
Variant_origin	How the variant was generated	TEXT		
Variant_length	Length of the variant in characters	INTEGER		
Ambiguity	It is 'Ambiguous' if the variant is shared by more than one Phi_ID, 'Univocal' if the variant is associated to only one philosopher. Is NULL for the 'Too_short' variants	TEXT		
Matching_Phi_IDs	List of the Phi_IDs that match the variant, separated by ' '	TEXT		
N_Multiple_Phi_IDs	Number of Phi_IDs that are associated to the variant	INTEGER		
Issue_level	Different issues that can affect a variant (levels: 'Highly_problematic', 'Maybe_problematic', 'Too_short', 'Unlikely_variant', 'Too_young_philosopher'). It is used to filter the table when the Entity Ruler is generated. Is NULL when the variant is ok	TEXT		

CLASSIFICATION

Properties attributed to the philosophers. They are of various types and can have various sources. For instance, the property 'Approach' comes from Strassfeld (2020) manual classification.

Field	Description	Type	Default	Other
Property_ID	Unique identifier of the property	TEXT		PK
Phi_ID	The same philosopher can have more than one property (or no property at all)	TEXT		FK

Field	Description	Type	Default	Other
Property	Value of the property	TEXT		
Type	Category of the property	TEXT		
Source	Source of the property	TEXT		
Manual	1 if the property comes from a manual or human-based annotation, 0 if comes from algorithms	INTEGER		

WIKI_ENTRY

Philosophers' data extracted from WikiData and furtherly elaborated.

Field	Description	Type	Default	Other
Wiki_ID	Link to the WikiData page of the philosopher. The entity ID in WikiData begins with "Q"	TEXT		PK, FK
Wiki_Label	Standard label of the WikiData entity	TEXT		
Wiki_aliases	List of the entity's aliases in WikiData, separated by ', '. Can be NULL	TEXT		
Wiki_description	Short description attached to the philosopher in WikiData. Can be NULL	TEXT		
Wiki_surname	Surname associated with the entity. When there are multiple surnames, they are separated by ' '. Can be NULL	TEXT		
Mult_surname	'Yes' if there is more than one surname associated to the entity in WikiData, 'No' otherwise	TEXT		

Field	Description	Type	Default	Other
Wiki_firstname	First name associated with the entity. When there are multiple first names, they are separated by ' '. Can be NULL	TEXT		
Mult_first	'Yes' if there is more than one first name associated to the entity in WikiData, 'No' otherwise	TEXT		
Wiki_BirthYear	Birthyear associated with the entity. When there are multiple birthyears, they are separated by ' '. When there is no birthyear available, there is the special label '[No birth year available]'. Can be NULL. WARNING: the value is NOT a number. For queries that involve mathematical operations, use the <BirthYear> field in the <Biography> table	TEXT		
Mult_years	'Yes' if there is more than one birth year associated to the entity in WikiData, 'No' otherwise	TEXT		
Wiki_gender	Gender associated with the entry. Can be NULL. Levels: 'male', 'female', 'genderqueer', 'non-binary', 'trans man', 'trans woman'	TEXT		
Mult_gender	It is always 'No' because there is no entry with more than one gender assigned. Can be NULL	TEXT		

HIRE_POSITION

This table is based on the dataset manually compiled by J. Strassfeld (see Strassfeld, 2020, 'American Divide: The making of "Continental" philosophy', Modern Intellectual History, DOI: 10.1017/S1479244318000513)

Field	Description	Type	Default	Other
-------	-------------	------	---------	-------

Field	Description	Type	Default	Other
Position_ID	Unique identifier of the job position.	TEXT		PK
Phi_ID	Unique philosopher ID. The same philosopher can hold more than one position, even in the same institution	TEXT		FK
Hire_Name	Name as recorded by Strassfeld. It has the structure Surname, FirstName	TEXT		
Institution_ID	Unique institution ID	TEXT		FK
Hired_year	Starting year of the position. It is NULL if it is before 1930	INTEGER		
Hired_before_1930	1 if the position started before 1930, 0 otherwise	INTEGER		
Hire_rank	Rank when hired	TEXT		
Left_year	Ending year of the position. It is NULL if it is after 1980	INTEGER		
Left_after_1980	1 if the position ended after 1980, 0 otherwise	INTEGER		
Highest_rank	Highest rank obtained	TEXT		

PHD DISSERTATION

Each entry describes a PhD dissertation in philosophy in America. The source is ProQuest (data were elaborated by Guido Bonino and Paolo Tripodi, see the documentation of the original file). Each candidate can have only one PhD dissertation (multiple dissertations of the same candidate were removed).

Field	Description	Type	Default	Other
PQ_ID	Unique identifier used in ProQuest	TEXT		PK

Field	Description	Type	Default	Other
Phi_ID	Unique identifier of the philosopher who got the PhD. Note that each philosopher can have only one PhD dissertation (the few multiple dissertations attached to the same candidate were removed).	TEXT		FK
Title	Title of the dissertation. '***NO TITLE PROVIDED***' if there is no recorded title	TEXT		
PhD_Candidate	Name of the candidate as recorded in ProQuest. It has the structure [Surname, FirstName]	TEXT		
PhD_Year	Year of the PhD awarding. First year = 1861, Last year = 1979	INTEGER		
Institution_ID	Unique identifier of the institution	TEXT		FK

INSTITUTION

Institutions for hiring data and PhDs.

Field	Description	Type	Default	Other
Institution_ID	Unique identifier of the institution, based on the affiliation ID used in ProQuest	TEXT		PK
Institution	Name of the university, based on the label used in ProQuest	TEXT		

JSTOR_AUTHORSHIP

Each entry is an authorship from the JSTOR dataset. The Primary Key of the table is a compound of Phi_ID and Article_ID.

Field	Description	Type	Default	Other
-------	-------------	------	---------	-------

Field	Description	Type	Default	Other
Phi_ID	Unique philosopher ID of the author of the article	TEXT		PK, FK
Author_position	Position of the author in the byline, as recorded in JSTOR	INTEGER		
original_JSTOR_author	Original author string in JSTOR, before cleaning	TEXT		
Article_ID	JSTOR unique identifier of the article. It is also the link to JSTOR webpage where the pdf can be accessed	TEXT		PK, FK

JSTOR_ARTICLE

Each entry is a bibliographic record from JSTOR. The columns are the metadata in JSTOR and are self-explanatory.

Field	Description	Type	Default	Other
Article_ID	JSTOR unique identifier of the article. It is also the link to JSTOR webpage where the pdf can be accessed	TEXT		PK
Authors_Phi_ID	List of the unique identifiers of the philosophers that are authors of the article, separated by ','	TEXT		
title	Title of the article	TEXT		
publicationYear	Publication year of the article	INTEGER		
doi	Is always NULL	TEXT		
docType	Is always 'article'	TEXT		
docSubType	Is always 'research-article'	TEXT		

Field	Description	Type	Default	Other
provider	Is always 'JSTOR'	TEXT		
collection	Is always NULL	TEXT		
datePublished	This is a date formatted as a text. Structure: YYYY-MM-DD	TEXT		
issueNumber	Issue number. Can be NULL	INTEGER		
volumeNumber	Volume number. Can be NULL	INTEGER		
url	Link to the stable URL of the article's pdf	TEXT		
creator	The original string of authors as recorded in JSTOR, before cleaning. Authors are separated by ';'. Is NULL for articles with no author recorded in JSTOR. Note that in EDHIPHY, these are attributed to the special 'author' P:1	TEXT		
publisher	Publisher of the journal	TEXT		
language	Language of the article. Multiple languages are separated by ';'.	TEXT		
pageStart	Starting page of the article. Can be NULL	INTEGER		
pageEnd	Ending page of the article. Can be NULL	INTEGER		
placeOfPublication	Is always NULL	TEXT		
keyphrase	Is always NULL	TEXT		

Field	Description	Type	Default	Other
wordCount	Number of words as recorded in JSTOR (hence before our cleaning)	INTEGER		
pageCount	Page span of the article	INTEGER		
file	Is always 'part-1.json.gz'. Refers to the json file where the full-texts are stored	TEXT		
N_authors	Number of authors of the article	INTEGER		
Journal_ID	Unique identifier of the journal	TEXT		FK

JOURNAL

Data about the 12 journals in the database.

Field	Description	Type	Default	Other
Journal_ID	Unique identifier of the journal in EDHIPHY	TEXT		PK
Journal_standard_name	Standard name of the journal in EDHIPHY	TEXT		
Standard_abbreviation	Standard ISO 4 abbreviation of the journal title	TEXT		
Journal_name_JSTOR	Journal name in JSTOR	TEXT		
Foundation_Year	Foundation year of the journal	INTEGER		

JOURNAL_VARIANT

Tables containing alternative names of journals, for journals that changed their name over time.

Field	Description	Type	Default	Other
Journal_ID	The same journal can have more than one variant (or no variant at all)	TEXT		FK
Journal_variant	Alternative name of the journal	TEXT		
Starting_date	Starting year of the journal variant. Can be NULL	INTEGER		
Ending_date	Ending year of the journal variant. Can be NULL	INTEGER		

FULL-TEXT

This table contains the cleaned full-text of the article. WARNING: THIS TABLE CANNOT BE SHARED, IT IS JSTOR PROPRIETARY DATA!

Field	Description	Type	Default	Other
id		TEXT		PK, FK
Full_text	Our cleaned version of the full-text of the article	TEXT		
Length_words	Length of the cleaned article in number of words	INTEGER		
Lenght_characters	Length of the cleaned article in characters	INTEGER		

MENTION

Each row in the table is a mention extracted from the cleaned full-text using NER supplied with the Entity Ruler and then disambiguated when needed.

A mention is a link connecting a mentioning article to a mentioned philosopher.

A philosopher can be linked to 0, 1 or more rows in this table.

WARNING: The Ambiguous mentions in the table have NULL value in the <Mentioned_Phi_ID> field (the Foreign Key) because they are linked to no <Phi_ID>.

Field	Description	Type	Default	Other
Mention_ID	Unique identifier of the single mention	TEXT		PK
Mentioning_Article_ID	Unique identifier of the article where the mention appears	TEXT		FK
Ent_text	String extracted with NER	TEXT		
Ent_label	Entity type. It is always 'PHILOSOPHER'	TEXT		
Associated_Phi_IDs	List of Phi_IDs that are associated with the extracted string, separated by ' '. WARNING: Do not use this field to establish a join with the table <Philosopher>!	TEXT		
Disambiguated_Phi_ID	List of the Phi_ID that remain after the disambiguation process, separated by ' '. It is equal or shorter than Associated_Phi_IDs. WARNING: Do not use this field to establish a join with the table <Philosopher>!	TEXT		
Disambiguation_status	Type of disambiguation applied and outcome of the disambiguation process. If 'Ambiguous', the disambiguation failed. Levels: 'Disambiguated_article', 'Disambiguated_author', 'Disambiguated_0_mentions', 'Algorithmically_Disambiguated', 'Disambiguated_author_corpus', 'Ambiguous', 'Non_ambiguous' (when no disambiguation was required)	TEXT		

Field	Description	Type	Default	Other
Mentioned_Phi_ID	When the disambiguation succeeded, it shows the unique identifier of the mentioned philosopher. If the disambiguation failed, it is NULL. Use this field to establish a join link with the table <Philosopher>	TEXT		FK

MENTION_CONTEXT

For each mention, including the ambiguous ones, it is reported the text that surrounds it, obtained from the cleaned version of the article's full-text. The surrounding text is comprised between 100 characters before the mention and 100 characters after the mention.

Field	Description	Type	Default	Other
Mentioning_Article_ID	Unique identifier of the article where the mention appears	TEXT		FK
Mention_ID	Unique identifier of the single mention	TEXT		PK, FK
Mention_context	Text surrounding the mention, comprised between 100 characters before the mention and 100 characters after the mention	TEXT		