

Famulus normalisation

One statement for a sourced reading

Gerd Graßhoff ^{ID}

Philosophisches Institut, Forschungsinstitut des Deutschen Museums
gerd.grasshoff@hu-berlin.de

ABSTRACT

Expression normalisation is stored as one kind of row: a statement with a source, a scope, and a predicate from a closed registry. This paper reports a first reading of the name index in Euler's correspondence with Goldbach, Opera omnia IVA 4, and a correspondent sample from the Kármán Nachlass. The index occupies 42 OCR pages. Every page that carries a running header places that header 7 page-indexes after the printed number. The parse yields 993 entries, 3,893 printed-page attestations, and 6,467 statements. Of the register's 311 eulerOO-IVA identities, 130 index heads match one identity by surname and initial. Twenty-seven heads match more than one identity and stay unresolved. The Kármán sample of 50 correspondent folders has not been reviewed, so no precision is stated. The vector store is still in place, because the comparison with the golden queries has not been run.

1 A statement is the row

A mention, a life date, an office, an identity, and an expert answer are the same kind of row. The row names a subject, a predicate, an object, and a source. The predicate is taken from a registry. A claim that the registry does not name is refused. Two writers who store the same claim from the same source obtain the same identifier, so a merge is a union.

The time of a statement is four day-number bounds. An Allen relation may stand without a calendar date. Path consistency may tighten a bound, and it records which pair forced the tightening. A bound that crosses itself is reported. It is not deleted.

2 The name index of IVA 4

The volume on disk is `euler00-IVA/978-3-0348-0892-7`. Its bibliography file names the work *Correspondence of Leonhard Euler with Christian Goldbach*, edited by Franz Lemmermeyer and Martin Mattmüller, in *Leonhardi Euleri Opera omnia*, series IVA, volume 4, and dates the online edition 2016 [1]. The name index is the OCR of page indexes 1214 through 1255, 42 pages.

The running header gives the printed page. On page index 1230 the header reads "Name Index 1223". On page index 1223 it reads "1216 Name Index". The difference is 7 on every page in the range that carries such a header. A printed page (n) is therefore stored as page index ($n + 7$). Roman numerals in the index are kept as printed tokens and are not given a page index.

Each OCR text node that is an entry is stored as its own quote, so the 993 entries are 993 quotes. The head is read by a pattern: the name, a life span when the parentheses contain years, a "see" pointer, and the page list after the colon. A year that the pattern does not recognise is not filled in. The run of 17 October 2026 produced:

Pages	42
Header offset	7
Entries	993
Attestations mapped to a page index	3,893
Statements	6,467

Entities	1,035
----------	-------

The global register holds 386 name rows for eulerOO-IVA and 311 distinct identities. An index head is linked when its surname is the last token of exactly one of those identities, or when the given-name initial narrows several forms to one identity. That links 130 heads. Twenty-seven heads, among them the several Bernoullis and the several Eulers, match more than one identity. They are listed and not collapsed.

One entry, read from page index 1223, is Johann Jacob van Delen, with the years 1743 and 1786 in parentheses and the printed pages 570, 574, 1129, and 1132. Those become page indexes 577, 581, 1136, and 1139. The biography clause is the OCR node itself.

■ 3 A Kármán sample

`letters_resolved` in the Kármán Nachlass holds 15,976 rows in 1,066 folders. Correspondent folders are those whose identifier begins with a numeric pair, continues with a name, and contains a year, excluding calculation files, inventories, and bibliographies. There are 401 such folders. The sample is the 25 folders with the most letters and 25 others drawn with seed 17. It contains 1,290 letters, of which 407 have an empty `date_when`. The largest folder in the sample is `11-40_guggenheim_harry_f_1926_1958_i`, with 64 letters. The owner has not reviewed the sample. No precision is stated.

■ 4 What is not claimed

The vector store used by the query service is still the recall path for paraphrase. The golden queries were not compared with the statement routes, so that store was not removed.

The calendar service answers a Julian day number, the weekday of that number, and the Gregorian Easter date. Easter is accepted when the date falls on a Sunday and lies between 22 March and 25 April. Territorial adoption tables, feast days beyond Easter, and the historical hour systems are not in this build.

The astronomical service on port 5142 is a separate expert. A sentence that names a longitude points at a calculation tool, or it says that the stated position cannot be retrieved.

■ Bibliography

- [1] L. Euler, *Correspondence of Leonhard Euler with Christian Goldbach*, vol. 4. in Leonhardi Euleri Opera omnia, Series quarta A, vol. 4. Basel: Bernoulli-Euler-Gesellschaft, 2016.