Codebook

publn_auth (character)
Patent authority from PATSTAT table tls211_pat_publn

publn_nr (character varying)
Publication number from PATSTAT table tls211_pat_publn

publn_nr_original (character varying)
"Original" publication number from PATSTAT table tls211_pat_publn

publn_kind (character)
Publication kind code from PATSTAT table tls211_pat_publn

publn_first_grant (smallint)
Indicator for granted patents from PATSTAT table tls211_pat_publn

appln_id (integer)
Application ID from PATSTAT table tls201_appln

appln_filing_date (date)
Application filing date from PATSTAT table tls201_appln

publn_date (date)
Publication date from PATSTAT table tls211_pat_publn

earliest_filing_date (date)
Earliest filing date from PATSTAT table tls201_appln

docdb_family_id (integer)
Family ID from PATSTAT table tls201_appln

process_count_2 (integer)
Number of process claims in patent (process keyword within first *two* words)

process_count_2_ind (integer)
Number of *independent* process claims in patent (process keyword within first *two* words)

process_count_5 (integer)
Number of process claims in patent (process keyword within first *five* words)

process_count_5_ind (integer)
Number of *independent* process claims in patent (process keyword within first *two* words)

process_use_count_pred (integer)
Number of process and use claims in patent (according to text mining)

process_use_count_pred_ind (integer)
Number of *independent* process and use claims in patent (according to text mining)

use_count (integer)
Number of use claims in patent (use keyword in first word)

use_count_ind (integer)
Number of *independent* use claims in patent (use keyword in first word)

product_by_process_count_2 (integer)
Number of product-by-process claims in patent (process keywords not allowed within first *two* words)

product_by_process_count_2_ind (integer)
Number of *independent* product-by-process claims in patent (process keywords not allowed within first *two* words)

total_count (integer)
Total number of claims in a patent

total_count_ind (integer)
Total number of *independent* claims in a patent

process_ratio_2 (numeric)
The share of a patent's process claims (process keyword within first *two* words)

process_ratio_2_ind (numeric)
The share of a patent's *independent* process claims (process keyword within first *two* words)

process_ratio_5 (numeric)
The share of a patent's process claims (process keyword within first *five* words)

process_ratio_5_ind (numeric)
The share of a patent's *independent* process claims (process keyword within first *five* words)

process_ratio_pred (numeric)
The share of a patent's process claims (according to text mining)

process_ratio_pred_ind (numeric)
The share of a patent's *independent* process claims (according to text mining)

use_ratio (numeric)
The share of a patent's use claims

use_ratio_ind (numeric)
The share of a patent's *independent* use claims

product_by_process_ratio_2 (numeric)
The share of a patent's product-by-process claims (process keywords not allowed within first *two* words)

product_by_process_ratio_2_ind (numeric)
The share of a patent's *independent* product-by-process claims (process keywords not allowed within first *two* words)

title_process (boolean)
The patent's title contained "process" keywords when field is set to true

title_use (boolean)
The patent's title contained "use" keywords when field is set to true

abstract_process (boolean)
The patent's abstract contained "process" keywords when field is set to true

abstract_use (boolean)
The patent's abstract contained "use" keywords when field is set to true

abstract_process_use_pred (boolean)
The patent's abstract classified as "process" or "use" according to text mining when field is set to true