Appendix C: Variable Codebook

Full-text Claims Datasets (patent_claims_fulltext and pgpubs_claims_fulltext)

Variable Name	Variable Description	Notes	Dataset
appl_id	Application identification number	U.S. patent application number issued to an applicant at filing.	PGPub only
claim_no	Claim number		Both
claim_txt	Claim text	This variable includes the full text of each claim, dependent or indepdent	Both
dependencies	Referenced claims to which the claim is dependent	This variable includes the all claim references within the text of the observed claim	Both
ind_flg	Indicator of independent claim	See identifying algorithm in Appendix 12.1.3 for more information	Both
pat_no	Patent number	U.S. patent number issued to an applicant at grant.	Patents only

Claim-level Statistics Datasets (patent_claims_stats and pgpubs_claims_stats)

Variable Name	Variable Description	Notes	Dataset
appl_id	Application identification number	U.S. patent application number issued to an applicant at filing.	PGPub Only
char_ct	Count of characters in each claim		Both
claim_no	Claim number		Both
cns_ct	Count of "consisting" in each claim		Both
cnx_flg	Indicator of a canceled claim		PGPub Only
ind_flg	Indicator of independent claim	See identifying algorithm in Appendix 12.1.3 for more information	Both
or_ct	Count of "or"s in each claim		Both
pat_no	Patent number	U.S. patent number issued to an applicant at grant.	Patents Only
pub_no	Pre-grant publication number		PGPub Only
sf_ct	Count of "selected from" in each claim		Both
word_ct	Count of words in each claim		Both

Document-level Statistics Datasets (patent_document_stats and pgpubs_document_stats)

Variable Name	Variable Description	Notes	Dataset
appl_id	Application identification number	U.S. patent application number issued to an applicant at filing.	Both
pat_clm_ct	Number of independent claims at grant		Patents Only
pat_dep_clm_ct	Number of dependent claims at grant		Patents Only
pat_dep_wrd_avg	Average count of words among dependent claims at grant		Patents Only

pat_dep_wrd_ct**	Number of words among dependent claims at grant		Patents Only
pat_dep_wrd_min**	Minimum count of words among dependent claims at grant		Patents Only
pat_no	Patent number	U.S. patent number issued to an applicant at grant.	Patents Only
pat_wrd_avg	Average count of words among independent claims at grant		Patents Only
pat_wrd_ct	Number of words among independent claims at grant		Patents Only
pat_wrd_min	Minimum count of words among independent claims at grant		Patents Only
pub_clm_ct	Number of independent claims at publication		PGPub Only
pub_dep_clm_ct	Number of dependent claims at publication		PGPub Only
pub_dep_wrd_avg**	Average count of words among dependent claims at publication		PGPub Only
pub_dep_wrd_ct**	Number of words among dependent claims at publication		PGPub Only
pub_dep_wrd_min**	Minimum count of words among dependent claims at publication		PGPub Only
pub_no	Pre-grant publication number		PGPub Only
pub_wrd_avg	Average count of words among independent claims at publication		PGPub Only
pub_wrd_ct	Number of words among independent claims at publication		PGPub Only
pub_wrd_min	Minimum count of words among independent claims at publication		PGPub Only

** While document-level dependent claim word count statistics are included in our datasets, these statistics
are not accurate measures of claim scope. The dependent claim word counts do not add the word counts of
referenced claims on which the given dependent claims are dependent. For more information, please see footnote
73.