

Table S1. Biomedical information extraction-based features extracted by scispaCy.

Features	Models
Entrez Gene and Protein Ontology	craft
Sequence Ontology	craft
NCBI Taxonomy	craft
Chemical Entities of Biological Interest	craft
Gene Ontology	craft
Cell Ontology	craft
DNA	jnlpba
Cell Type	jnlpba
Cell Line	jnlpba
RNA	jnlpba
Protein	jnlpba
Disease Names	bc5cdr
Chemicals Names	bc5cdr
Amino Acids	bionlp13cg
Anatomical Systems	bionlp13cg
Cancer	bionlp13cg
Cell	bionlp13cg
Cellular Components	bionlp13cg
Developing Anatomical Structure	bionlp13cg
Gene or Gene Product	bionlp13cg
Immaterial Anatomical Entity	bionlp13cg
Multi-tissue Structure	bionlp13cg
Organ	bionlp13cg
Organism	bionlp13cg
Organism Subdivision	bionlp13cg
Organism Substance	bionlp13cg
Pathological Formation	bionlp13cg
Simple Chemical	bionlp13cg

Tissue	bionlp13cg
--------	------------

Table S2: Biomedical information extraction-based features extracted by MetaMap.

Features	Full Names
aapp	Amino Acid, Peptide, or Protein
acab	Acquired Abnormality
acty	Activity
aggp	Age Group
amas	Amino Acid Sequence
amph	Amphibian
anab	Anatomical Abnormality
anim	Animal
anst	Anatomical Structure
antb	Antibiotic
arch	Archaeon
bacs	Biologically Active Substance
bact	Bacterium
bdsu	Body Substance
bdsy	Body System
bhvr	Behavior
biof	Biologic Function
bird	Bird
blor	Body Location or Region
bmod	Biomedical Occupation or Discipline
bodm	Biomedical or Dental Material
bpoc	Body Part, Organ, or Organ Component
bsoj	Body Space or Junction
celc	Cell Component
celf	Cell Function
cell	Cell

cgab	Congenital Abnormality
chem	Chemical
chvf	Chemical Viewed Functionally
chvs	Chemical Viewed Structurally
clas	Classification
clna	Clinical Attribute
clnd	Clinical Drug
cnce	Conceptual Entity
comd	Cell or Molecular Dysfunction
crbs	Carbohydrate Sequence
diap	Diagnostic Procedure
dora	Daily or Recreational Activity
drdd	Drug Delivery Device
dsyn	Disease or Syndrome
edac	Educational Activity
eehu	Environmental Effect of Humans
elii	Element, Ion, or Isotope
emod	Experimental Model of Disease
emst	Embryonic Structure
enty	Entity
enzy	Enzyme
euka	Eukaryote
evnt	Event
famg	Family Group
ffas	Fully Formed Anatomical Structure
fish	Fish
fndg	Finding
fngs	Fungus
food	Food

ftcn	Functional Concept
genf	Genetic Function
geoa	Geographic Area
gngm	Gene or Genome
gora	Governmental or Regulatory Activity
grpa	Group Attribute
grup	Group
hcpp	Human-caused Phenomenon or Process
hcro	Health Care Related Organization
hlcs	Health Care Activity
hops	Hazardous or Poisonous Substance
horm	Hormone
humn	Human
idcn	Idea or Concept
imft	Immunologic Factor
inbe	Individual Behavior
inch	Inorganic Chemical
inpo	Injury or Poisoning
inpr	Intellectual Product
irda	Indicator, Reagent, or Diagnostic Aid
lang	Language
lbpr	Laboratory Procedure
lbtr	Laboratory or Test Result
mamm	Mammal
mbrt	Molecular Biology Research Technique

mcha	Machine Activity
medd	Medical Device
menp	Mental Process
mnob	Manufactured Object
mobd	Mental or Behavioral Dysfunction
moft	Molecular Function
mosq	Molecular Sequence
neop	Neoplastic Process
nnon	Nucleic Acid, Nucleoside, or Nucleotide
npop	Natural Phenomenon or Process
nusq	Nucleotide Sequence
ocac	Occupational Activity
ocdi	Occupation or Discipline
orch	Organic Chemical
orga	Organism Attribute
orgf	Organism Function
orgm	Organism
orgt	Organization
ortf	Organ or Tissue Function
patf	Pathologic Function
phob	Physical Object
phpr	Phenomenon or Process
phsf	Physiologic Function
phsu	Pharmacologic Substance
plnt	Plant
podg	Patient or Disabled Group
popg	Population Group
prog	Professional or Occupational Group

pros	Professional Society
qlco	Qualitative Concept
qnco	Quantitative Concept
rcpt	Receptor
rept	Reptile
resa	Research Activity
resd	Research Device
rnlw	Regulation or Law
sbst	Substance
shro	Self-help or Relief Organization
socb	Social Behavior
sosy	Sign or Symptom
spco	Spatial Concept
tisu	Tissue
tmco	Temporal Concept
topp	Therapeutic or Preventive Procedure
virs	Virus
vita	Vitamin
vtbt	Vertebrate

Table S3: Biomedical information extraction-based features extracted by spaCy customized model.

Features
Pathogen
Medical Condition names
Medicine names

Table S4: Named-entity recognition features.

Features
Person
FAC
GPE
Product
Work of art
Language
Time
Money
Cardinal
NORP
Organization
Location
Event
Law
Date
Percent
Quantity
Ordinal

Table S5: Linguistic features.

Features
Num of stopwords
Num of @
Num of numeric values
Num of lowercase characters
Text language
Character count
Average word length
Num of ?
Num of /
Num of #
Num of uppercase characters
Num of all uppercase characters
Word count
Sentence count
Average sentence length

Table S6: Sentiment Features.

Features
Positive sentiment score
Negative sentiment score
Neutral sentiment score
Compound sentiment score
Text polarity