pyresid

start

FullTextXML

BeautifulSoup

FullText

TextDict

spaCy

tokenize

tagger

parser

NER

Identification and Location

Protein Data Bank(PDB)API

PDB Entry List

Validate

PDB URI

RegEx

PDB MatchClass

Exact String

start index

end index

UniProt URI

Query

UniProt API

Query

PDB API

Download

mmCIF/PDB Structure

Validate + Associate

Residue MatchClass

Exact String

start index

end index

UniProt URI

Decompose Composite Matches

RegEx

Dictionary Lookup

Residue MatchClass

Exact String

start index

end index

Residues

RegEx

MatchClass

Exact String

start index

end index

SourceClass

Doc

PMC ID

FullText

TextDict

Doc

Section Names

Section Whitelist

FuzzyWuzzy

MatchClass

Exact String

start index

start Token

end index

end Token

Context Addition

Add Context

Residue MatchClass

Exact String

start index

end index

UniProt URI

Prefix

Postfix

Token

Sent

spaCy

SourceClass

PMC ID

Section Matches

Doc

FullText

Residue Matches

Output to Json

Data Product

Annotations

+PMID/PMCID
+Annotations
- prefix
- exact (residue)
- postfix
- section
- URI (protein)

Europe PubMed Central