## **User input**

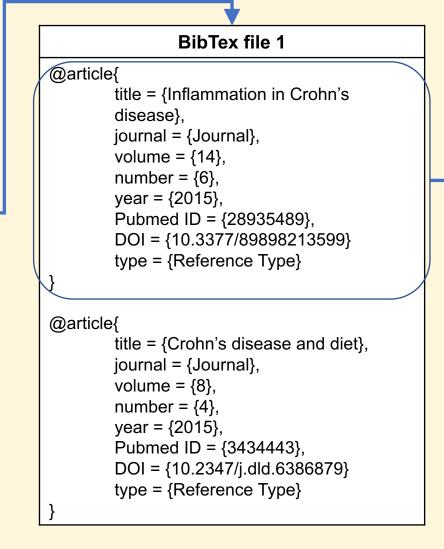
The whole process is repeated for every file of the folder

## BibTex Files/

- BibTex file 1.bib
- BibTex file 2.bib
- •
- BibTex file N.bib

# Open the BibTex file

The process is repeated for every article of the file



## **Extract the Pubmed ID**

**Pubmed ID** 

28935489

Download the associated abstract from PubMed Database and annotate the text using EXTRACT tagger

#### **Abstract**

Crohn's disease is a type of inflammatory bowel disease that can affect any part of the digestive system, but most commonly occurs in the last section of the small intestine or the large intestine. Escherichia coli has been increasingly implicated in the pathogenesis of the Crohn's disease. Sulfasalazine, an anti-inflammatory medicine, is a type of treatment option for Crohn's disease patients.

#### **Extract the identified terms**

Result for 28935489							
BibTex file 1	28935489	pathogenesis	Biological process	GO:0009405			
BibTex file 1	28935489	Sulfasalazine	Chemical compound	CIDs05359476			
BibTex file 1	28935489	Crohn's disease	Disease	DOID:8778			
BibTex file 1	28935489	inflammatory bowel disease	Disease	DOID:0050589			
BibTex file 1	28935489	Escherichia coli	Organism	562			
BibTex file 1	28935489	large intestine	Tissue	BTO:0000706			
BibTex file 1	28935489	small intestine	Tissue	BTO:0000651			

### **Description of pytag internal pipeline structure**

Result					
BibTex file 1	28935489	pathogenesis	Biological process	GO:0009405	
BibTex file 1	28935489	Sulfasalazine	Chemical compound	CIDs05359476	
BibTex file 1	28935489	Crohn's disease	Disease	DOID:8778	
BibTex file 1	28935489	inflammatory bowel disease	Disease	DOID:0050589	
BibTex file 1	28935489	Escherichia coli	Organism	562	
BibTex file 1	28935489	large intestine	Tissue	BTO:0000706	
BibTex file 1	28935489	small intestine	Tissue	BTO:0000651	
BibTex file 1	3434443	proteobacteria	Organism	1224	
BibTex file 1	3434443	peroxisome	Cellular Component	go:0005777	
BibTex file 2	3434443	l4a	Genes/Proteins	ybr031w	
BibTex file 2	789564653	i -12	Molecular Function	go:0005143	
BibTex file 2	789564653	tap water	Environment	envo:00003096	

•

BibTex file N	67457564	skin	Tissue	bto:0001253
BibTex file N	67457564	diarrhoea	Disease	doid:13250
BibTex file N	67457564	vitamin d	Chemical Compound	cids05280795

All numbers are fake.
The document just
describes a plan for
processing data.