CEVOpen

# Wikidata supporting open student research projects in plant chemistry

WikiDataCon 2021-10-31

**Presentation by**

CEVOpen Team

WIKI DATA CON

2021 - a sustainable future for Wikidata

# About CEVOpen

- Peter Murray-Rust – **C**ontentMine, TDM toolkit
- Gitanjali Yadav – **E**ssOilDB, a database
- Emanuel Faria– **V**erriclear, a company

Essential Oil Literature

WIKI DATA CON

2021 - a sustainable future for Wikidata

# Wikidata supporting open student research projects in plant chemistry

WikiDataCon 2021-10-31

Shweata N. Hegde

WIKI DATA CON

2021 - a sustainable future for Wikidata

# Presentation Outline

## First 5 Min. - Shweata

- Why are we mining plant chemistry?
- Overview of the workflow
- Introduce KARYA Interns

## 5 – 18 Min. – KARYA Interns

- Demo the tools
- Share experience and challenges

Pre-recorded

## 18 – 21 Min. – Sum up

WIKI DATA CON
2021 - a sustainable future for Wikidata

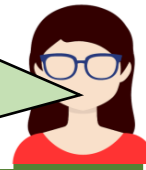
# About Me - Shweata N. Hegde (ಶ್ವೇತಾ ನ್. ಹೆಗಡೆ[KN])

- 3rd year  life science undergrad
- SciComm + Science Education + Open Science
- CEVOpen: Natural Language Processing project - `Docanalysis`

WIKI DATA CON
2021 - a sustainable future for Wikidata

# Why Plant Chemistry?

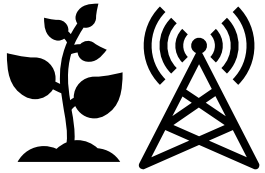
“Plants are better chemists than we are!“



Plants use chemicals to:

- communicate
- Defend and resist disease
- Attract pollinators
- …

And for literally everything else!

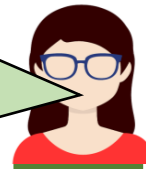


## TERPENES!

- >20000 compounds
- All have same starter molecule
- Based on what enzymes acts on the starting molecules the final product is different

# Why Plant Chemistry?

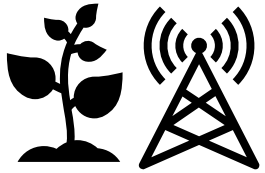
"Plants are better chemists than we are!"



Plants use chemicals to:

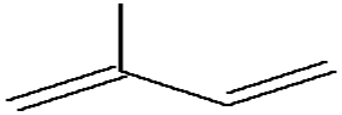
- communicate
- Defend and resist disease
- Attract pollinators
- …

And for literally everything else!



# TERPENES!

- >20000 compounds
- All have same starter molecule
- Based on what enzymes acts on the starting molecules the final product is different



isoprene

Enzymes

Aromatic!

…

What are the enzymes involved in the terpene synthesis?

What are its genes?

We are interested in TERPENES of crops

Deluge of information in scientific literature, thanks to recent advancement!

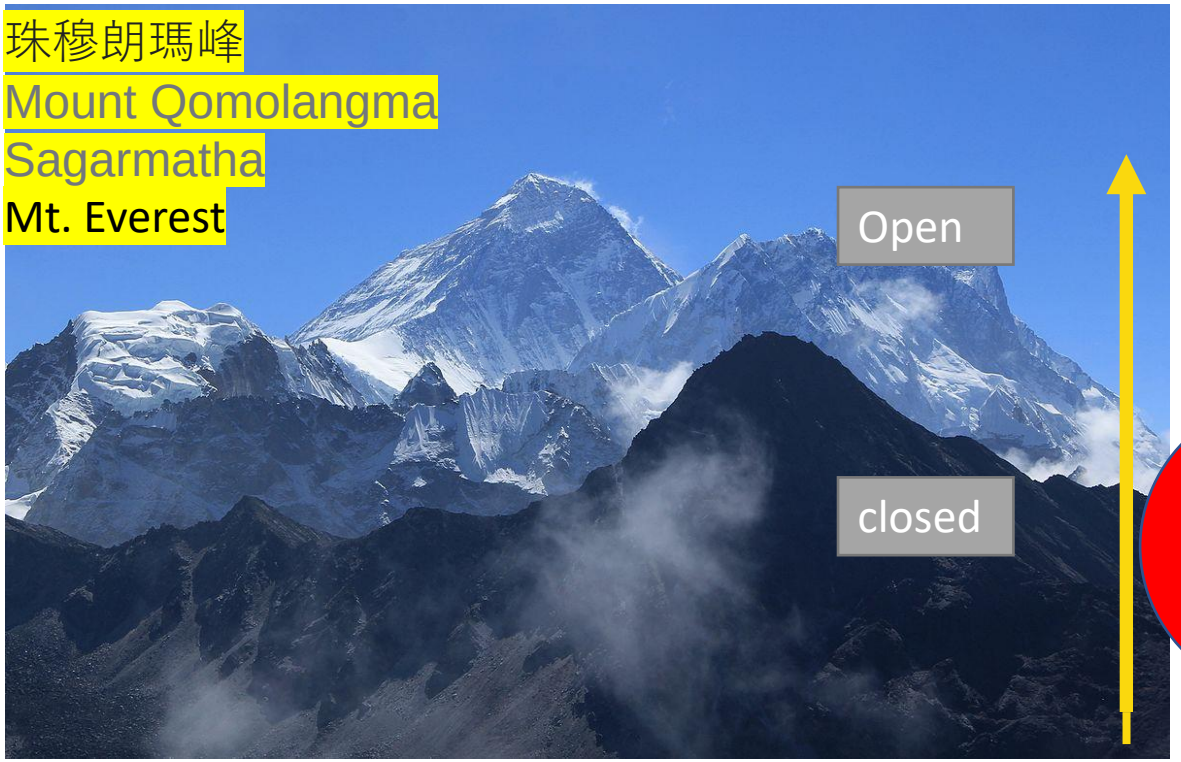
# But...

This Photo by Unknown Author is licensed under CC BY-SA

# SCIENTIFIC LITERATURE

How big is the scientific literature?

70 Million papers*?!



30% Open Access

Scientific literature

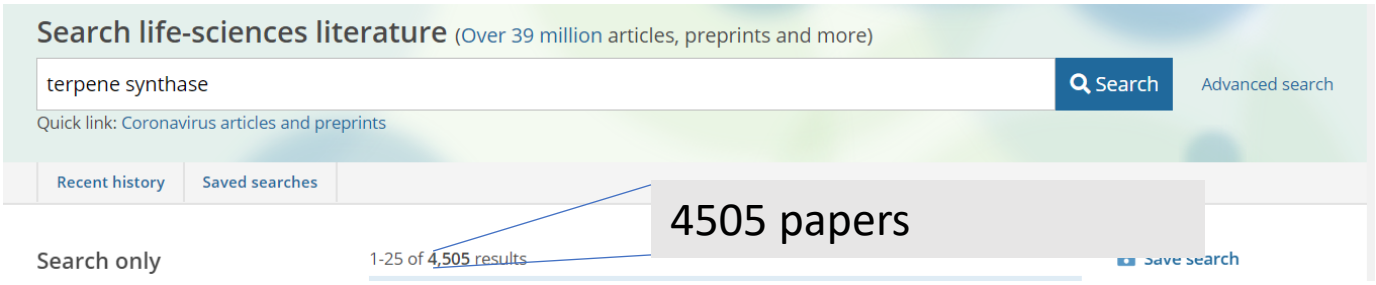
Can't read unless you pay!

One year's worth of publication in print

*https://blogs.lse.ac.uk/impactofsocialsciences/2020/01/17/not-yet-the-default-setting-in-2020-open-research-remains-a-work-in-progress/

P.S. this is just a back of the envelope calculation!

# SCOPING THE LITERATURE *MANUALLY*




4505 papers

I want to read the whole of scientific literature about terpene synthases…

1 Find and Download

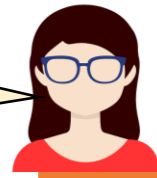
2 Read and annotate

3 Comprehend, interpret and infer

Many grad students spend 2 months time on their reviews!

# SCOPING THE LITERATURE *AUTOMATICALLY*

CEVOpen Tools can help!



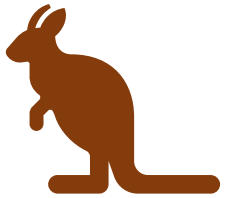
```
pygetpapers -q "terpene synthase" -n
INFO: Total number of hits for the query are 4505
```

1. Find and Download
2. Read and annotate
3. Comprehend, interpret and infer

pygetpapers

Europe PMC
Free information resource for biomedical and health researchers

WIKIDATA



Takes few minutes!

WIKIDATA

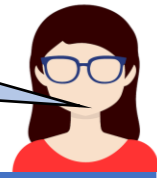
# Dictionaries are mining tools!

WIKI DATA CON

2021 - a sustainable future for Wikidata

# OpenVirus Dictionary

Multilingual!



WIKIDATA

Country

Diseases

Drugs

Funders

Viruses

Test & Trace

Zoonosis

Animal carriers of disease

Non-Pharmaceutical Interventions

Slide Credits: Ambreen Hamadani

**Credits:** MS PowerPoint 3D Models; Wikipedia (https://en.wikipedia.org)

# Plants Dictionary (CEVOpen)

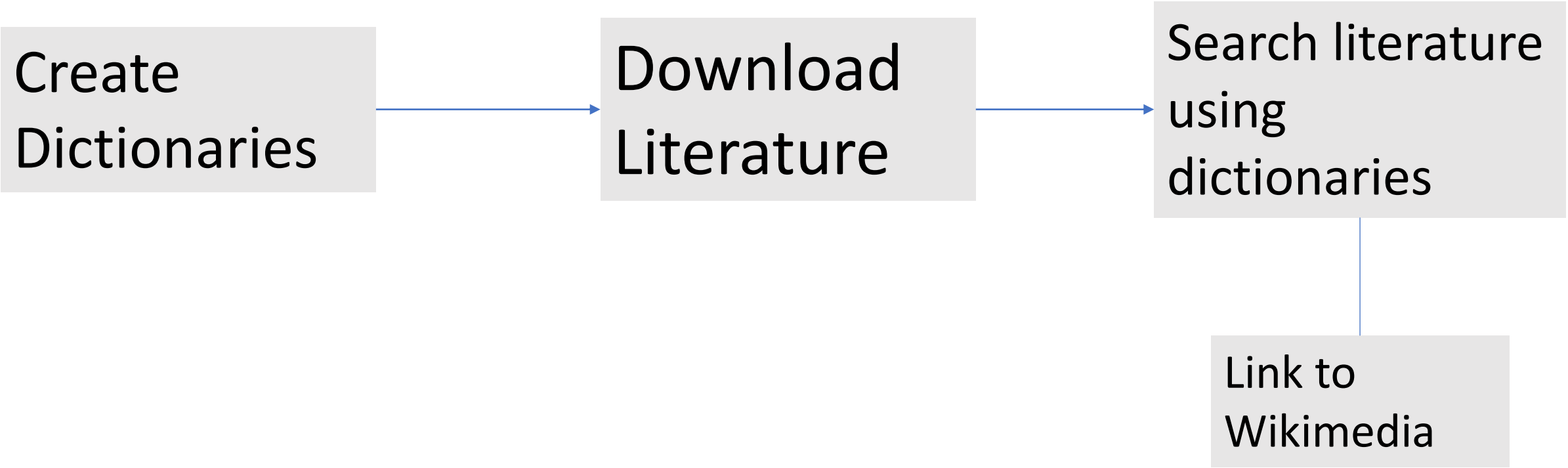
```
 <entry term="cone“
wikidataURL="http://www.wikidata.org/entity/Q22710"
wikidataID="Q22710"
wikipediaPage="https://en.wikipedia.org/wiki/Conifer_cone">
    <synonym>Cone, Conifer cones, Strobilus, pine cone,
pine-cone, pinecone, pine cones</synonym>

    <synonym xml:lang="zh">松球</synonym>
    <synonym xml:lang="de">Koniferenzapfen</synonym>
    <synonym xml:lang="hi">कोणधारी कोण</synonym>
    <synonym xml:lang="es">cono</synonym>
    <synonym xml:lang="fr">cône</synonym>
    <description xml:lang="fr">organe reproducteur des
conifères</description>
    <description xml:lang="es">estructura
vegetal</description>
  </entry>
```

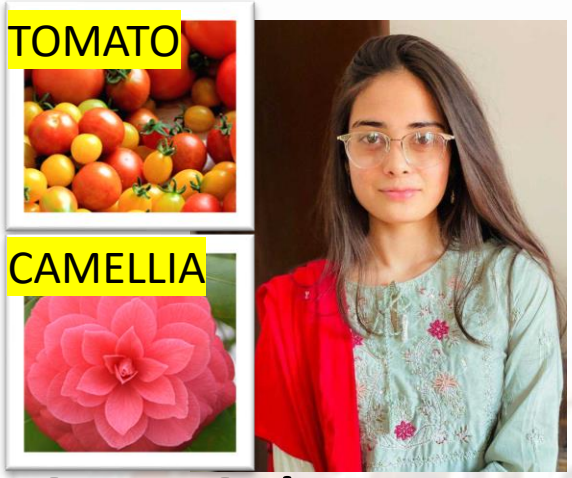
term

Multilingual entries

Credits: Vasant Kumar

# Workflow


Create Dictionaries
Download Literature
Search literature using dictionaries
Link to Wikimedia



**Prachi Singh**
Undergrad (Botany),
Janki Devi Bajaj Girls
Govt. College, Kota

**Anam Anjum**
Masters (Botany) Student,
Janki Devi Bajaj
Government Girls Science
College Kota

**Anjali Shekhawat**
Masters (Botany) Student,
Jai Narain Vyas University,
Jodhpur, Rajasthan, India.

**Kumari Rageshwari**
Masters (Biotech) Student,
Botany Department in
University of Rajasthan

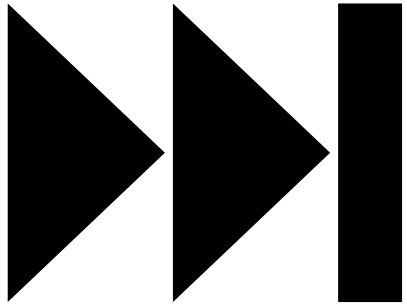
# Mining Plant Chemistry Literature

## KARYA (Knowledge Augmentation through Research in Young Aspirants) Interns

WIKI DATA CON
2021 - a sustainable
future for Wikidata

https://en.wikipedia.org/wiki/Terpene_synthase_N_terminal_domain , CC BY-SA

Next up:
Prachi Singh



WIKI DATA CON
2021 - a sustainable future for Wikidata

# Creating Dictionaries from Wikidata SPARQL

## Our Mining tool created from Wikidata!

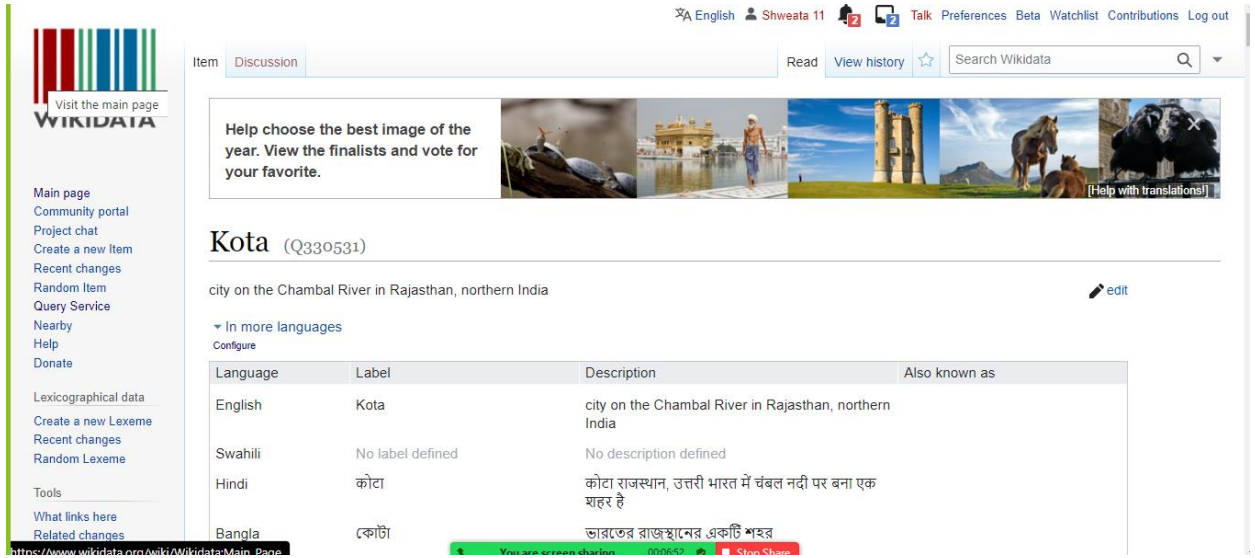
**Prachi Singh**
KARYA Intern

WIKI DATA CON
2021 - a sustainable future for Wikidata

# About Me - Prachi Singh (प्राची सिंह)

- From: Kota (Q330531), Rajasthan
- Bachelors (Botany) from Janki Devi Bajaj Girls Government College, Rajasthan, India.
- Joined CEVOpen in Sep. 2021

Working hard to learn – open source tools and open science



WIKI DATA CON

2021 - a sustainable future for Wikidata

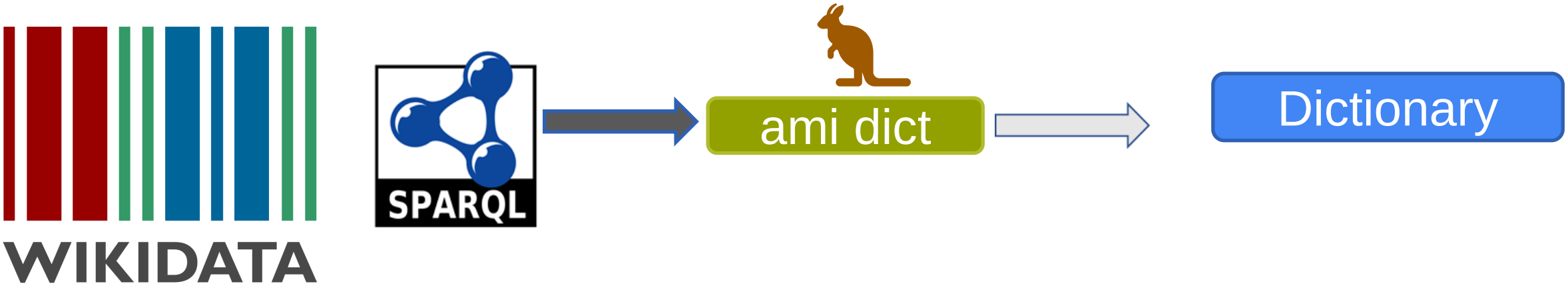
# Workflow

Create Dictionaries → Download Literature → Search literature using dictionaries

Link to Wikimedia

SPARQL method

# Dictionary creation using Wikidata




WIKI DATA CON
2021 - a sustainable future for Wikidata

# SPARQL Query

https://w.wiki/4HEb

```
#diterpenes

SELECT ?item ?itemLabel

WHERE

{

  ?item wdt:P31 wd:Q59447. # instance of diterpene

  SERVICE wikibase:label { bd:serviceParam wikibase:language "[AUTO_LANGUAGE],en". } # Helps get the label in your language, if not, then en language

}
```

WIKI DATA CON

2021 - a sustainable future for Wikidata

# SPARQL Demo

Do you need help creating a query? You can build queries without having to write SPARQL in the new Query Builder.

Wikidata Query Service | Examples | Query Builder | Help | More tools | English

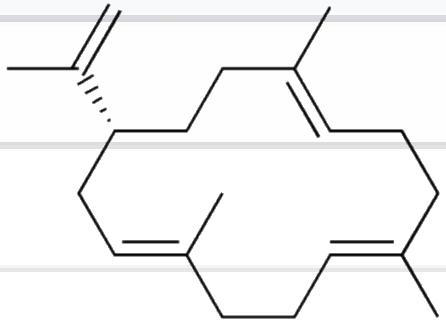
```
#diterpenes
SELECT ?item ?itemLabel
WHERE
{
  ?item wdt:P31 wd:Q59447. # diterpenes
  SERVICE wikibase:label { bd:serviceParam wikibase:language "[AUTO_LANGUAGE],en". } # Helps get the label in your language, if not, then en languag
}
```

WIKI DATA CON

2021 - a sustainable future for Wikidata

# SPARQL Results

| item | itemLabel |
| --- | --- |
| wd:Q5058639 | neocembrene |
| wd:Q74411014 | (-)-spiroviolene |
| wd:Q74411287 | pseudolaratriene |



# SPARQL Endpoint (XML)

Wikidata ID


```
- <results>
  - <result>
    - <binding name="diterpenes">
        <uri>http://www.wikidata.org/entity/Q5058639</uri>
      </binding>
    - <binding name="diterpenesLabel">
        <literal xml:lang="en">neocembrene</literal>
      </binding>
  </result>
  - <result>
```

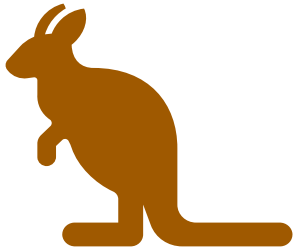

These go into the dictionary…

WIKI DATA CON

2021 - a sustainable future for Wikidata

# One last step:`ami-dict` to convert into dictionary

```
D:\diterpene_dictionary>amidict -vv
--dictionary diterpenes
--directory ami_terpene_dict
--input diterpene_sparql_out.xml
create --informat wikisparqlxml
--sparqlmap                                  # map SPARQL output to AMIDict
      wikidataURL=diterpenes,
      term=diterpenesLabel,
--transformName wikidataID=EXTRACT(wikidataURL,.*/(.*))
```



WIKI DATA CON

2021 - a sustainable future for Wikidata

# One last step: `ami-dict` to convert into dictionary

```
(c) Microsoft Corporation. All rights reserved.

D:\diterpene_dictionary>
```

# Final Dictionary!

Term

Wikidata ID

```
<entry wikidataID="Q5058639" wikidataURL="http://www.wikidata.org/entity/Q5058639"
term="neocembrene"/>
<entry wikidataID="Q74411014" wikidataURL="http://www.wikidata.org/entity/Q74411014"
term="(-)-spiroviolene”/>
<entry wikidataID="Q74411287" wikidataURL="http://www.wikidata.org/entity/Q74411287"
term="pseudolaratriene" />
```

Ready to go off and annotate the scientific literature!

WIKI DATA CON

2021 - a sustainable future for Wikidata

# My challenges

Java GitHub
EPMC Command line tools
Environment variables
Collaboration Wiki ...

Transition

## 6 weeks ago

Very little coding experience

## Now

Basic Python, Java, Git

WIKI DATA CON
2021 - a sustainable future for Wikidata

# Thanks for your attention!

## Useful Links

https://github.com/petermr/openVirus/wiki/openVirus:-SPARQL-queries-of-four-core-dictionaries

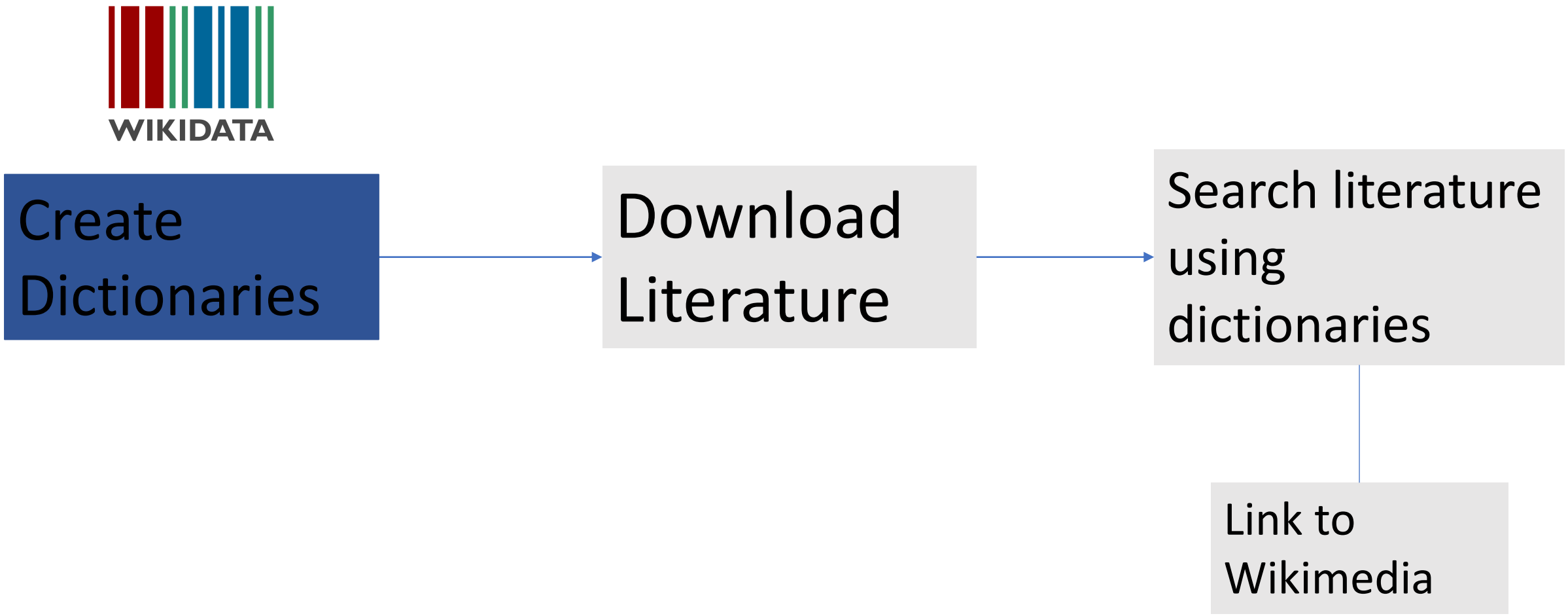
https://github.com/petermr/crops/wiki/Zea-mays-TPS-:-Prachi-Singh

https://www.wikidata.org/wiki/Wikidata:SPARQL%2520tutorial

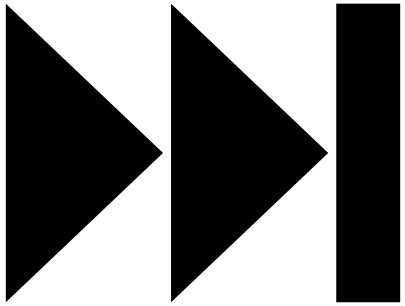
**Credits**
OpenVirus and CEVOpen Team

WIKI DATA CON
2021 - a sustainable future for Wikidata

# Workflow


WIKIDATA
Create Dictionaries
Download Literature
Search literature using dictionaries
Link to Wikimedia

Next up:
Anjali
Shekhawat



WIKI DATA CON
2021 - a sustainable
future for Wikidata

# Creating Dictionaries from scientific literature

## Enhanced with Wikidata

**Anjali Shekhawat**
KARYA Intern

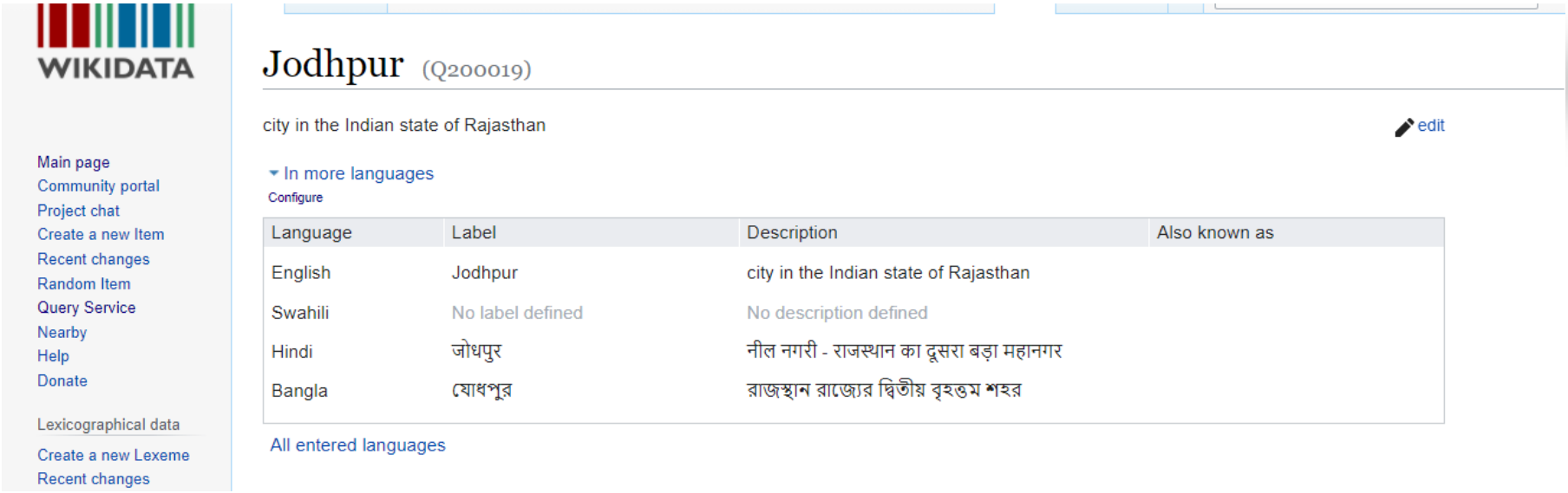
WIKI DATA CON
2021 - a sustainable future for Wikidata

# About Me – Anjali Shekhawat (अंजलि शेखावत)

- Masters (Botany) from , Jai Narain Vyas University of Jodhpur (Q200019), Rajasthan
- Joined CEVOpen in Sep. 2021

Learnt about:

- Wikidata and other open source tools at CEVOpen



WIKIDATA

Main page
Community portal
Project chat
Create a new Item
Recent changes
Random Item
Query Service
Nearby
Help
Donate

Lexicographical data
Create a new Lexeme
Recent changes

Jodhpur (Q200019)

city in the Indian state of Rajasthan

edit

In more languages
Configure

| Language | Label | Description | Also known as |
|---|---|---|---|
| English | Jodhpur | city in the Indian state of Rajasthan | |
| Swahili | No label defined | No description defined | |
| Hindi | जोधपुर | नील नगरी - राजस्थान का दूसरा बड़ा महानगर | |
| Bangla | যোধপুর | রাজস্থান রাজ্যের দ্বিতীয় বৃহত্তম শহর | |

All entered languages

WIKI DATA CON
2021 - a sustainable future for Wikidata

# Workflow

Create Dictionaries → Download Literature → Search literature using dictionaries

Terms linked to Wikidata

Link to Wikimedia

# What are Dictionaries – Annotating Tools

- Set of terms of our interest
- compounds, plants, country, disease, genes, and so on
- Used to annotate the scientific literature

... Previous studies on the floral volatile organic compounds of *M. officinalis* at different flowering stages revealed that terpenes were the main constituents, with the highest content of monoterpenes and sesquiterpenes such as camphene Q416775, *D-*limonene Q27888324 myrcene Q424577, and caryophyllene Q421614 (Wang et al., 2011). Terpenes are the largest class of floral components, of which the terpene synthase (TPS) genes are the key genes that determine the spatiotemporal release of volatile compounds (Gao et al., 2018). The involvement of TPS in the production of floral fragrance has been demonstrated in various plants (Muhlemann et al., 2014). A total of 40 putative TPS proteins in the *M. officinalis* genome were annotated according to the conserved domains (PF01397 and PF03936) using HMMER (Potter et al., 2018). We constructed a phylogenetic tree by aligning the TPS proteins Q6731965 Q141303 among *M. officinalis*, *M. biondii*, *L. chinense*, and *A. thaliana* to confirm the classification of TPS family proteins in *M. officinalis* (Figure 8B). According to the protein phylogenetic tree, the 40 TPS proteins were divided into five groups, which contain TPS-a, TPS-b, TPS-c, TPS-e, and TPS-f. …

Chemicals

Species

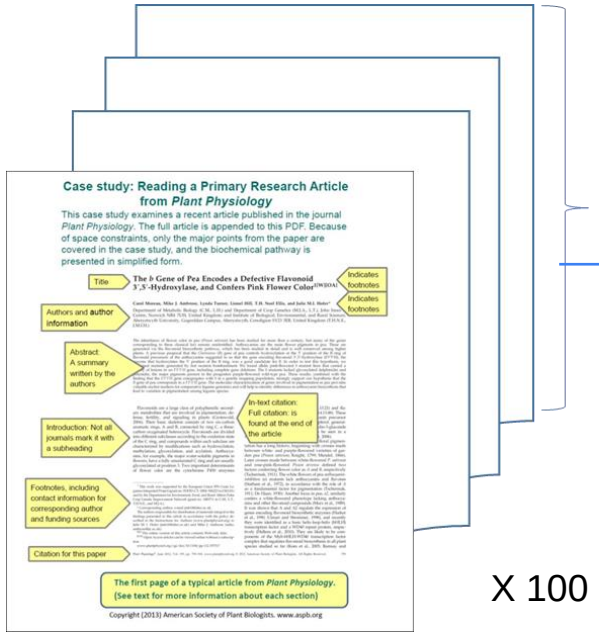
https://europepmc.org/article/MED/34505009

WIKI DATA CON

2021 - a sustainable future for Wikidata

# How do we create dictionaries?

## Mini-corpora




X 100

Europe PMC

Free information resource for biomedical and health researchers

pygetpapers

Terpene Synthase literature on Mentha (Mint)

Collected terpene synthase mentions

- diphosphate synthase
- valencene synthase
- terpene synthase 3
- β-ocimene synthase
- geranyl diphosphate
- monoterpene synthase

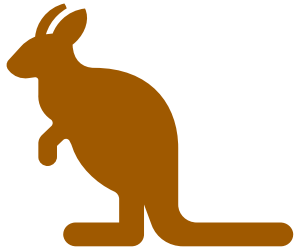
…

Text file

ami

WIKIDATA

Look-up Wikidata

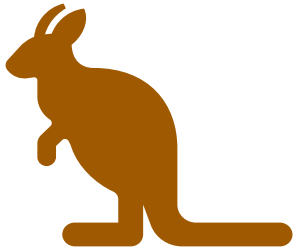


WIKI DATA CON

2021 - a sustainable future for Wikidata

# Last step:`ami-dict` to convert into dictionary

```
amidict -v
--dictionary terpene_synthase
--directory terpene_synthase_dir
--input terpene_synthase.txt
 create
--informat list
--outformats xml
```



WIKI DATA CON

2021 - a sustainable future for Wikidata

# Last step:`ami-dict` to convert into dictionary

ami-dict demo

# Final Dictionary!

term

Wikidata ID

```
…
<entry term="diphosphate synthase“ WikidataID = “Q18648706”/>
<entry term="valencene synthase WikidataID = “Q13405193”"/>
<entry term=“geranyl diphosphate” WikidataID = “Q23340835”"/>
…
```

Ready to go off and annotate the scientific literature!

WIKI DATA CON

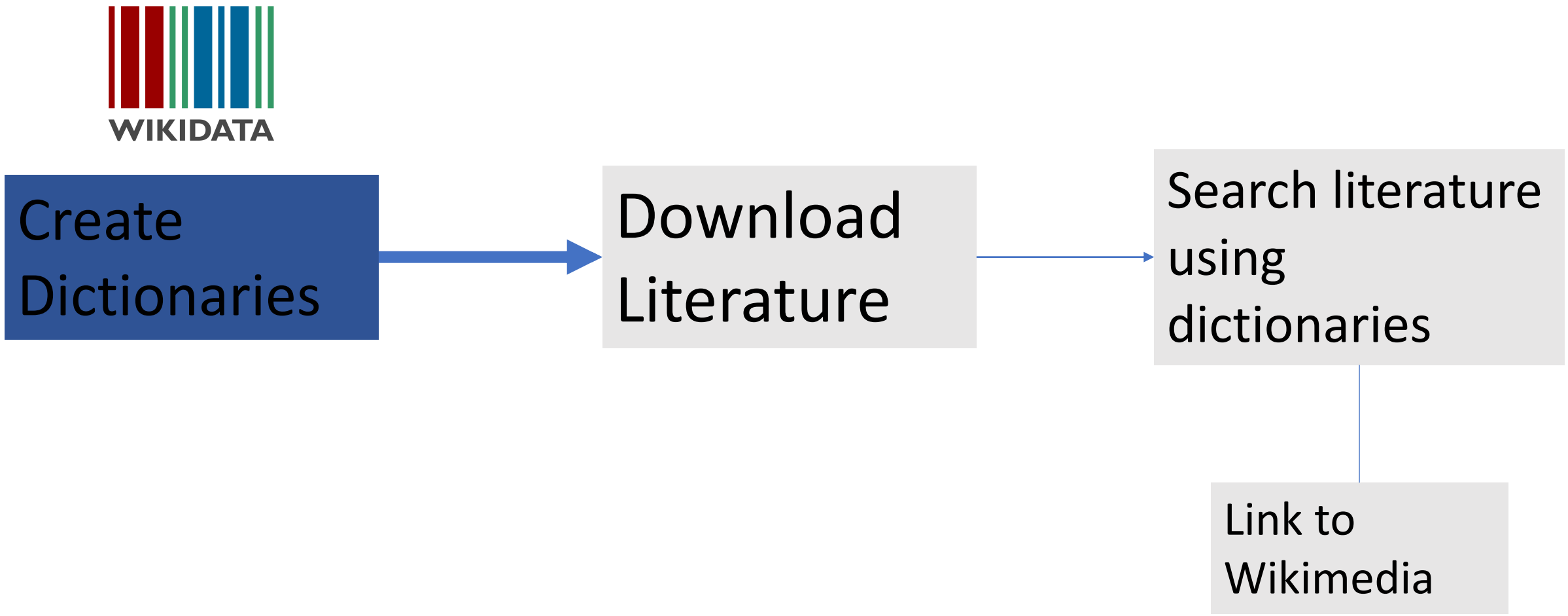
2021 - a sustainable future for Wikidata

# Learning Experience

- Teamwork
- Communally solving problems made learning to run tools fun!

WIKI DATA CON
2021 - a sustainable future for Wikidata

# Workflow


WIKIDATA
Create Dictionaries
Download Literature
Search literature using dictionaries
Link to Wikimedia

Next up:
Anam Anjum

WIKI DATA CON
2021 - a sustainable
future for Wikidata

# Creating Mini-Corpora

## How we find and automatically download papers


**Anam Anjum**

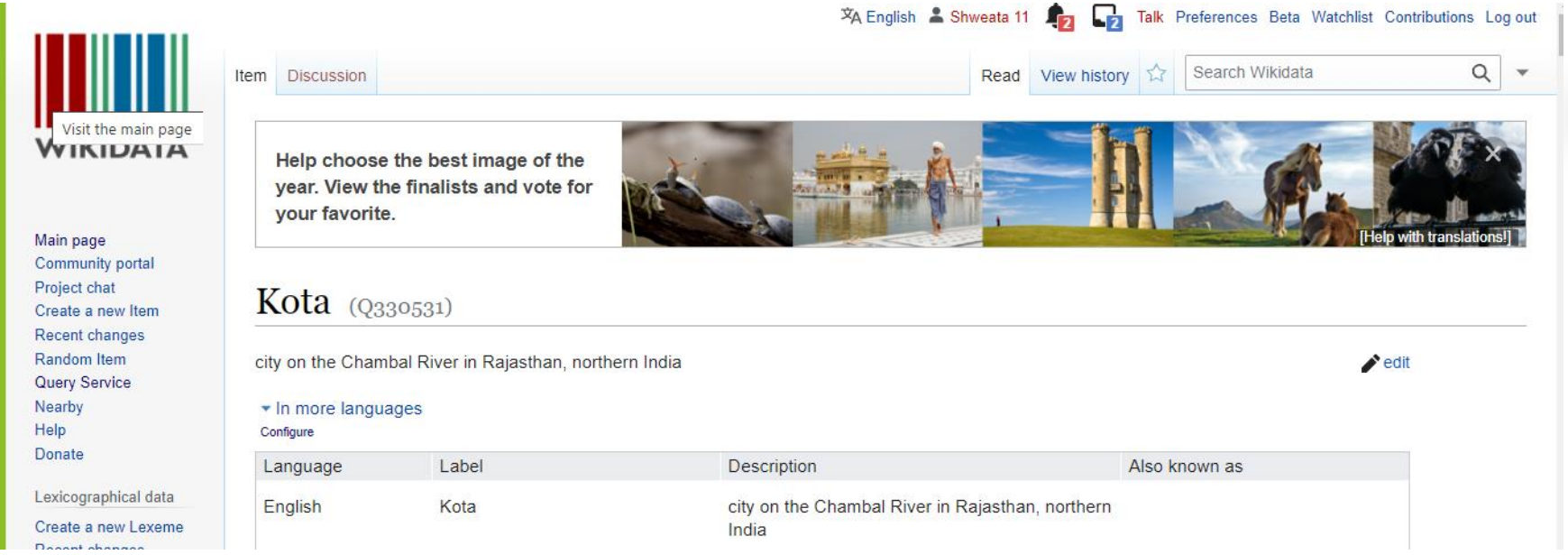
KARYA Intern

WIKI DATA CON

2021 - a sustainable future for Wikidata

# About Me – Anam Anjum - अनम अंजुम

- From: Kota (Q330531), Rajasthan
- Masters (Botany) from Janki Devi Bajaj Girls Government College, Rajasthan, India.
- Joined CEVOpen in Sep. 2021
- Work on creating dictionaries




WIKI DATA CON

2021 - a sustainable future for Wikidata

# Workflow

Create Dictionaries → Download Literature → Search literature using dictionaries

Link to Wikimedia

# 1. Mini-corpora

Europe PMC

Free information resource for biomedical and health researchers

Has millions of open access papers.

*

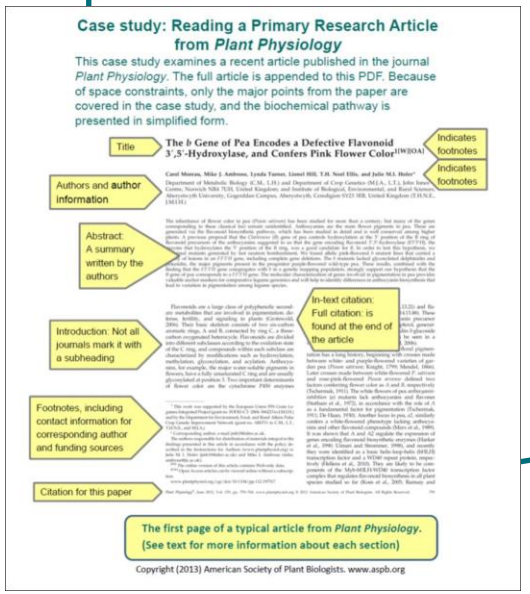
Open Repository

Preprint Servers

pygetpapers

Downloads hundreds of papers in minutes

Mini-corpora




X 1000

**On local machines.**

WIKI DATA CON

2021 - a sustainable future for Wikidata

Credits; Ayush Garg and Richard D Smith-Unna

http://www.seer.ufu.br/index.php/sociedadenatureza/issue/view/1922is licensed under CC BY

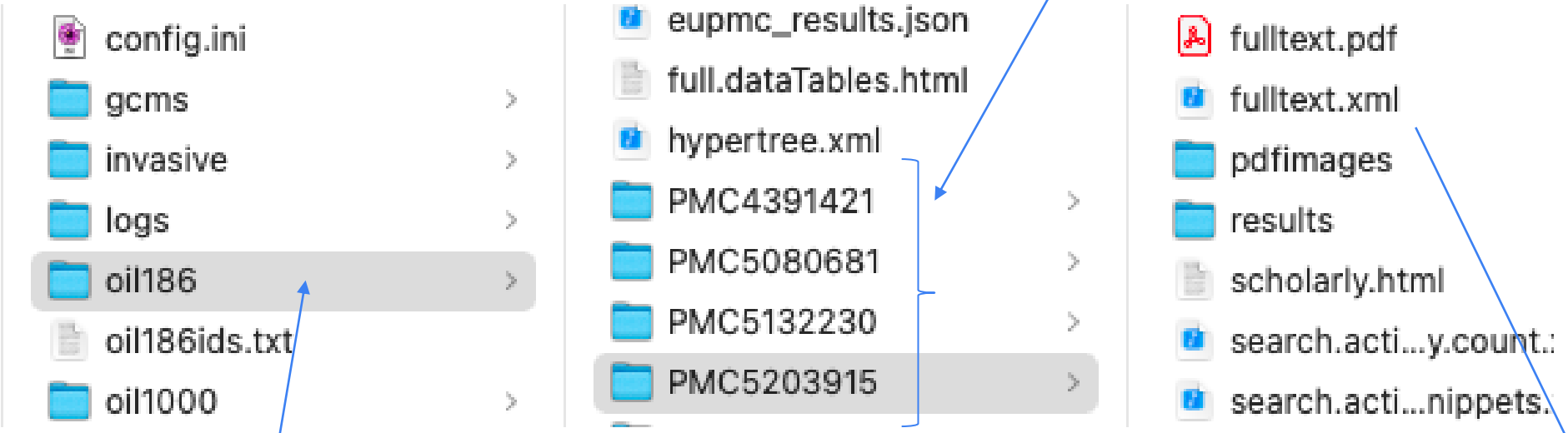
This Photo by Unknown Author is licensed under CC BY-NC

# pygetpapers query

```
pygetpapers -q "terpene synthase AND MENTHA(((SRC:MED OR SRC:PMC
OR SRC:AGR OR SRC:CBA) NOT (PUB_TYPE:"Review")))" -k 80 -x -o
TPS_mentha
```

WIKI DATA CON
2021 - a sustainable future for Wikidata

# Minicorpus




Folder containing papers

Mincorpus

Fulltext

# My progress

Plant science student

55 days

Literate about open tools

https://github.com/petermr/crops/wiki/TPS-Reserve-For-Camellia-and-Tomato-Crop

WIKI DATA CON

2021 - a sustainable future for Wikidata

# Experience

Creative team

Learn from mistakes

Help each other

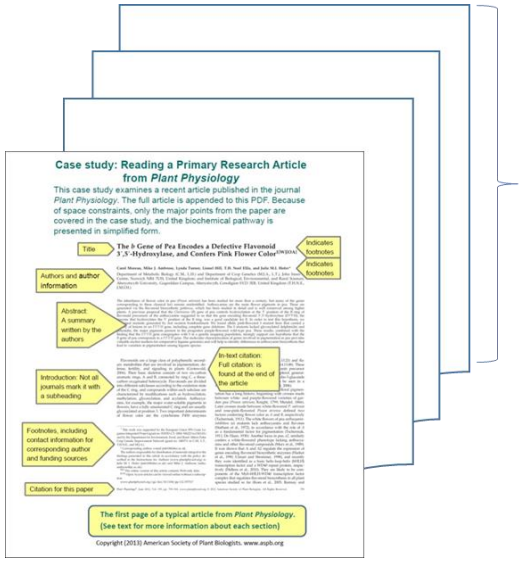
Team work

WIKI DATA CON

2021 - a sustainable future for Wikidata

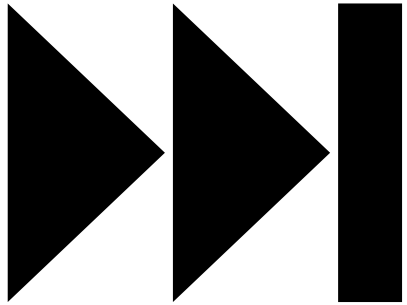
# Workflow

Create Dictionaries → Download Literature → Search literature using dictionaries




Link to Wikimedia

Next up: Kumari Rageshwari



WIKI DATA CON

2021 - a sustainable future for Wikidata

# ami-search

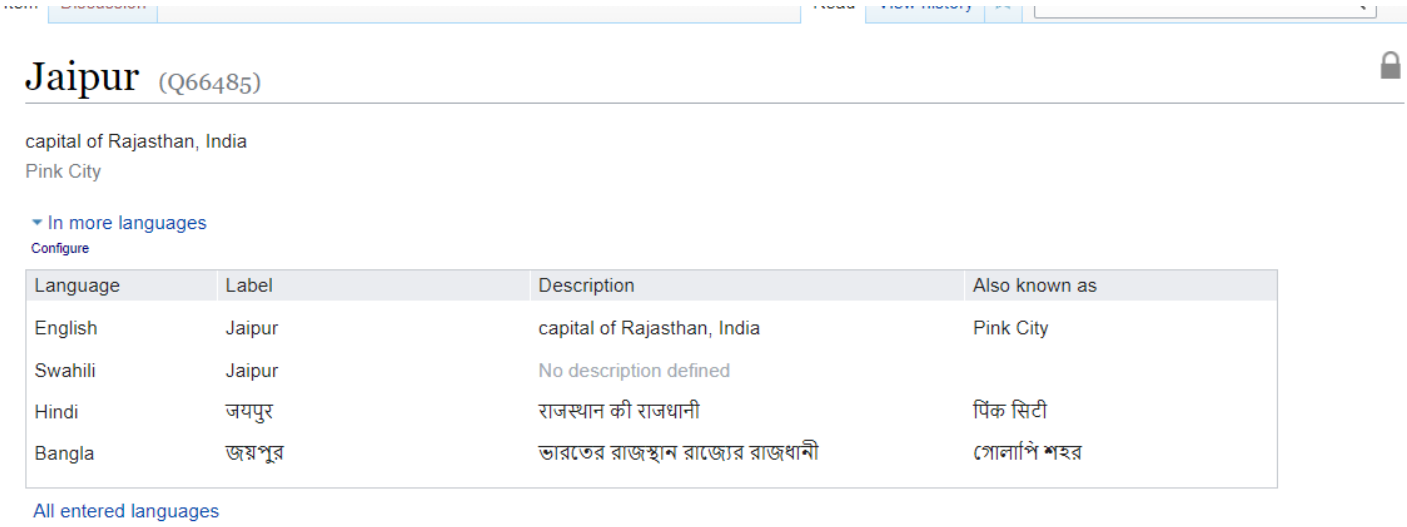
## How it all fits together

**Kumari Rageshwari**
KARYA Interns

WIKI DATA CON
2021 - a sustainable future for Wikidata

# About Me - Kumari Rageshwari (कुमारी रागेश्वरी)

- Masters (Biotechnology) from Botany Department, University of Rajasthan, Jaipur(Q66485)



Jaipur (Q66485)

capital of Rajasthan, India
Pink City

In more languages
Configure

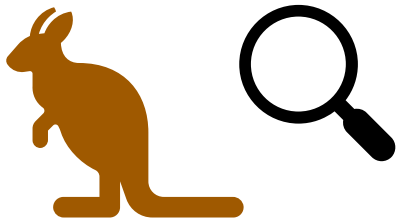
| Language | Label | Description | Also known as |
|---|---|---|---|
| English | Jaipur | capital of Rajasthan, India | Pink City |
| Swahili | Jaipur | No description defined | |
| Hindi | जयपुर | राजस्थान की राजधानी | पिंक सिटी |
| Bangla | জয়পুর | ভারতের রাজস্থান রাজ্যের রাজধানী | গোলাপি শহর |

All entered languages

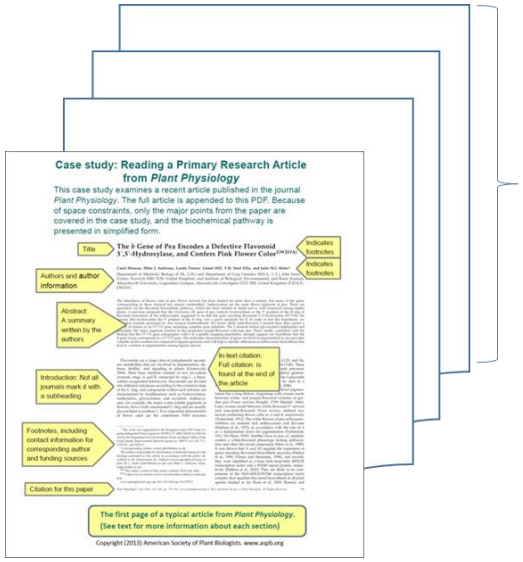
https://www.wikidata.org/wiki/Q66485

WIKI DATA CON
2021 - a sustainable future for Wikidata

# Workflow



Create Dictionaries → Download Literature → Search literature using dictionaries




Link to Wikimedia

# ami-search using eoTPS dictionary

minicorpus

```
ami -p "TPSvolatile"
search
  --dictionary eo_Gene.xml
```

dictionary

WIKI DATA CON
2021 - a sustainable
future for Wikidata

# ami-search using eoTPS dictionary

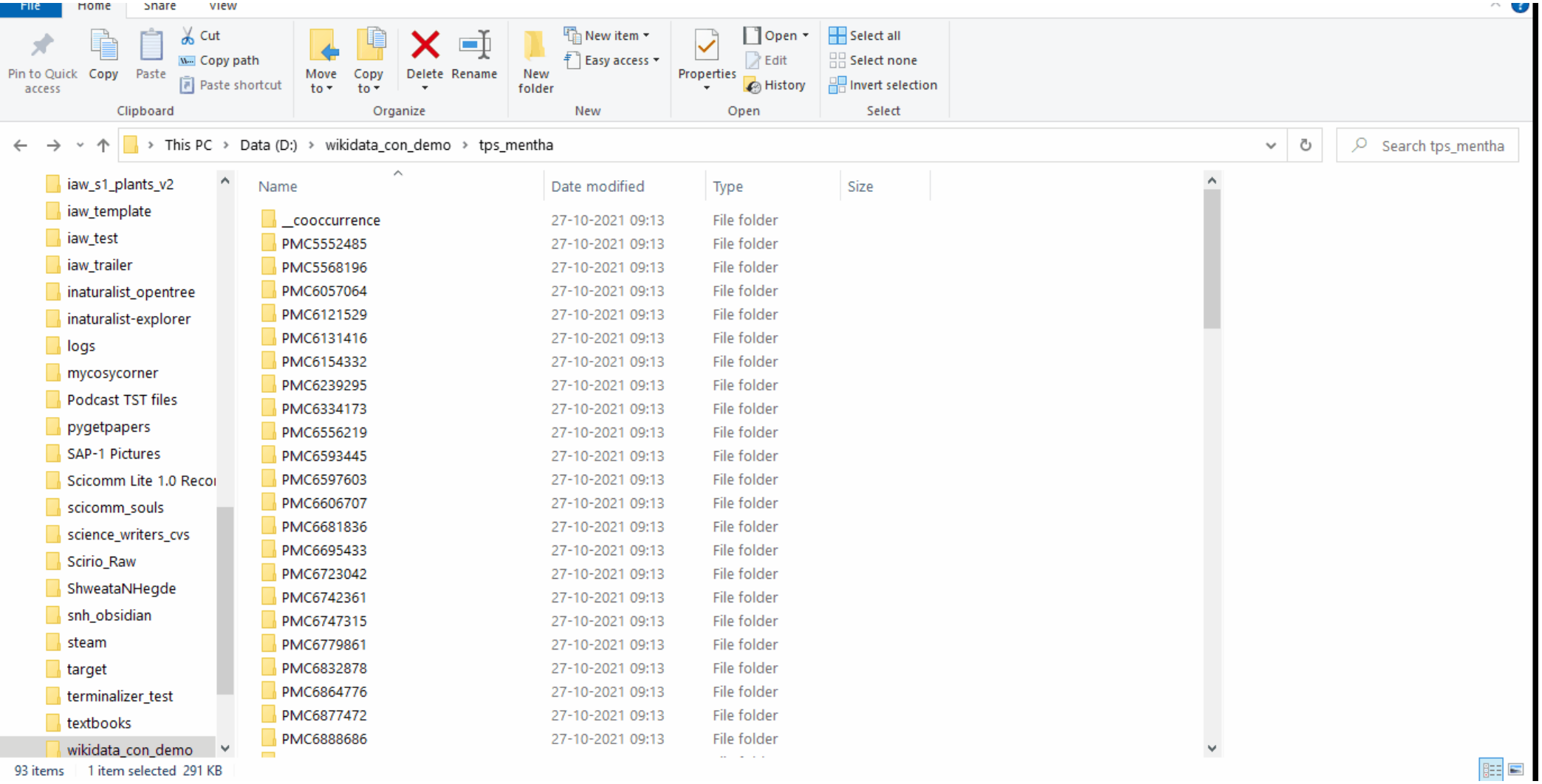
ami-search demo

```
Windows PowerShell
Copyright (C) Microsoft Corporation. All rights reserved.

Try the new cross-platform PowerShell https://aka.ms/pscore6
```

WIKI DATA CON
2021 - a sustainable
future for Wikidata

# Dashboard

## Semantic Papers




WIKI DATA CON

2021 - a sustainable future for Wikidata

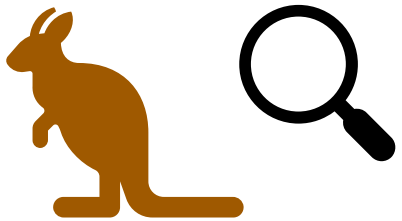
# Co-occurrence

mini_tpsenzyme
mini_eoplant

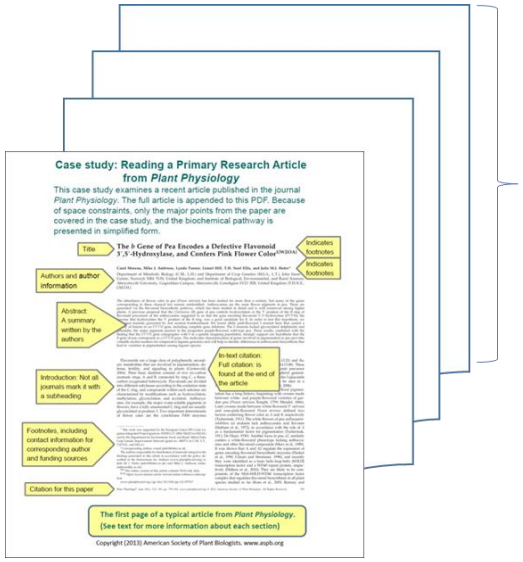
| | terpene synthas | gas | monoterpene syn | sesquiterpene s | limonene syntha | linalool syntha | geraniol syntha | myrcene synthas | diterpene synth | geranyllinalool | monoterpene syn | isoprene syntha | terpene synthas | ent-copalyl dip | geraniol syntha | terpinolene syn | isoprene syntha |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Arabidopsis tha | 17 | 11 | 6 | 6 | 3 | 3 | 2 | 4 | 3 | 1 | 1 | 0 | 0 | 0 | 1 | 1 | 0 |
| peppermint | 10 | 5 | 5 | 2 | 5 | 3 | 3 | 5 | 1 | 1 | 2 | 0 | 1 | 0 | 2 | 0 | 0 |
| Artemisia annua | 7 | 6 | 2 | 3 | 2 | 1 | 0 | 1 | 2 | 1 | 0 | 1 | 0 | 1 | 1 | 0 | 1 |
| spearmint | 6 | 4 | 4 | 2 | 6 | 2 | 3 | 3 | 2 | 0 | 1 | 1 | 0 | 1 | 1 | 0 | 1 |
| Cannabis sativa | 5 | 3 | 2 | 2 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 0 |
| C. sativa | 2 | 2 | 1 | 1 | 2 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| Arabidopsis tha | 7 | 4 | 2 | 2 | 3 | 1 | 1 | 1 | 2 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 0 |
| [illegible] | [illegible] | [illegible] | [illegible] | [illegible] | [illegible] | [illegible] | [illegible] | [illegible] | [illegible] | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 1 |
| [illegible] | [illegible] | [illegible] | [illegible] | [illegible] | [illegible] | [illegible] | [illegible] | [illegible] | [illegible] | 1 | 1 | 0 | 0 | 0 | 1 | 0 | 0 |
| Cannabis sativa | 2 | 1 | 1 | 1 | 2 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| Daucus carota | 2 | 2 | 2 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| Lavandula angus | 2 | 3 | 2 | 0 | 1 | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 |
| Vitis vinifera. | 2 | 2 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Trachyspermum a | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

WIKI DATA CON
2021 - a sustainable future for Wikidata

# Workflow



Create Dictionaries → Download Literature → Search literature using dictionaries




Link to Wikimedia

# How we work as a Team!

WIKI DATA CON

2021 - a sustainable future for Wikidata

# Documentation

We work in the OPEN!

– (03/09/2021)

Github Wiki page

## SCOPING OF LITERATURE AFTER DOWNLOADING FROM PYGETPAPERS

Instructions in detail onto this page - crop5 - petermr

### 1.Downloaded papers by running this command –

```
`
 -q "terpene synthase volatile Vitis vinifera AND (((SRC:MED OR SRC:PMC OR SRC:AGR OR SRC:CBA) NOT (PUB_TYPI
```

(-05/09/2021)

2.Collected volatiles of Terpene synthases, TPS etc. of Vitis vinifera from papers. (-06/07/21)

### 3.Downloaded supplementary papers by running this command (folder name is changed here, it's ViTPS instead of VitisTPS after facing issues due to similar named folder's presence)–

```
`-q "terpene synthase volatile Vitis vinifera AND (((SRC:MED OR SRC:PMC OR SRC:AGR OR SRC:CBA) NOT (PUB_TYI
```

WIKI DATA CON

2021 - a sustainable future for Wikidata

# Communal Problem Solving

petermr / crops Public

Watch 2 | Star 1 | Fork 2

Code | Issues 2 | Pull requests | Actions | Projects | Wiki | Security | Insights

**Label issues and pull requests for new contributors**

Now, GitHub will help potential first-time contributors discover issues labeled with good first issue

Dismiss

Filters | is:issue is:open | Labels 9 | Milestones 0 | New issue

2 Open | 1 Closed | Author | Label | Projects | Milestones | Assignee | Sort

- Py4ami installation Error — 5
  #4 opened 10 days ago by Anam04anjum
- Testing dictionary — 1
  #3 opened 19 days ago by prachi0508

WIKI DATA CON

2021 - a sustainable future for Wikidata

# Thanks for your attention!

Every progress reported on wiki – https://github.com/petermr/crops/wiki/Progress-outline-of-Vitis-vinifera-:-Kumari-Rageshwari

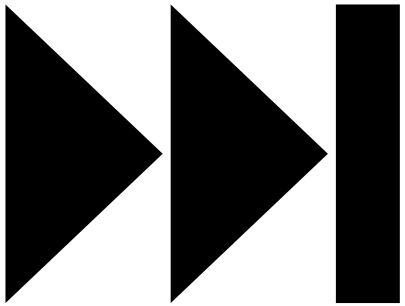
Full data table and link to wikipedia - https://github.com/petermr/crops/blob/main/Vitis%20vinifera/full.dataTables.html

Ami installation procedure documented – https://github.com/petermr/openVirus/wiki/Documentation-of-getpapers-&-ami

**Credits**
OpenVirus and CEVOpen

WIKI DATA CON
2021 - a sustainable future for Wikidata

# Back to Shweata



WIKI DATA CON

2021 - a sustainable future for Wikidata

# Open Notebook Philosophy

- ✓ Meet Biweekly (virtually):

  **Abuzz with Ideas**

- ✓ Update our progress

  **Show it to the World**

Solve problems communally

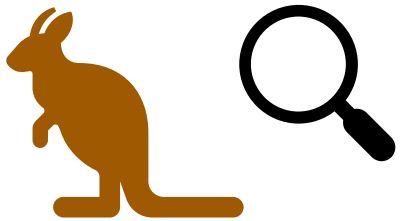
| ISSUES | Ambreen | Vaishali | Priya | Rajan | Vanisha | S | | |
|---|---|---|---|---|---|---|---|---|
| corpus (between 200-950) | Ready | Ready | Ready | Ready | Ready | Ready | Ready | Ready |
| dictionary | Ready | Ready | Ready | Ready | Ready | Ready | Ready | In progress |
| annotated (Pos/Neg) viral epidemics | 8.90% | 40/10 | 230/20 | 111/116 | In progress | 36/14 | 121/29 | 5/35 |
| specialist subsets (e.g. methods) e.g. sections | Ready | Ready | In progress | Ready | In progress | Ready | Ready | Haven't Started |
| notebooks | Ready | In progress | In progress | In progress | In progress | In progress | In progress | Haven't Started |
| machine learning / NLP | In progress | Haven't Started | Haven't Started | In progress | Haven't Started | Haven't Started | In progress | In progress |
| display | In progress | In progress | Haven't Started | Haven't Started | Haven't Started | Haven't Started | In progress | In progress |
| language variants | Ready | Haven't Started | In progress | In progress | Haven't Started | In progress | Haven't Started | Haven't Started |
| ami search | Ready | Ready | In progress | Ready | In progress | In progress | Ready | Haven't Started |

Credits: MS PowerPoint Icons

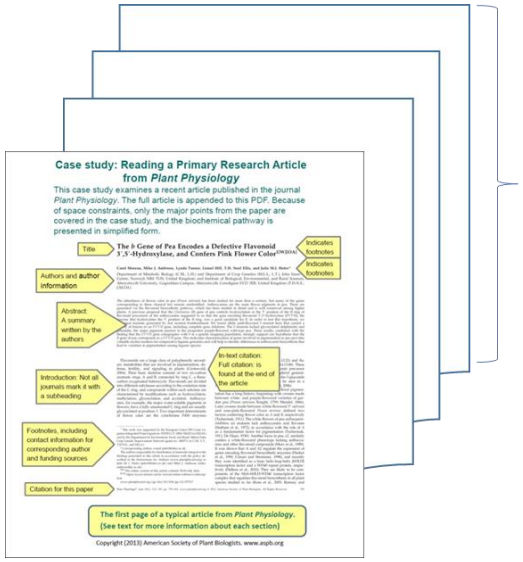
Slide courtesy: Ambreen

*Some of us use only mobile devices*

# Workflow



Create Dictionaries → Download Literature → Search literature using dictionaries

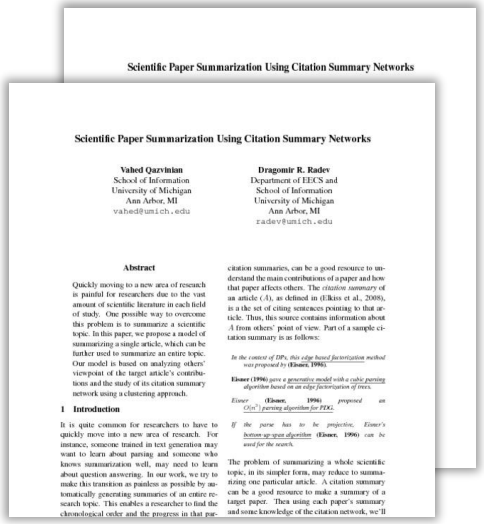



Link to Wikimedia

# Docanalysis: Feedback Loop

Dictionary

**Natural Language Processing and Machine Learning**




10000 papers

AUTOMATIC
1 paper/s

. L-phenylalanine is converted to 2-phenylacetaldehyde by phenyl acetaldehyde synthase (PAAS). Subsequently, 2-phenylacetaldehyde is reduced to 2-phenylethanol by phenyl acetaldehyde reductase (PAR). Next, 2-phenylethanol is acetylated to 2-phenylethyl acetate by acetyl-coenzyme a: geraniol/citronellol acetyl transferase (AAT)...

WIKIDATA

WIKI DATA CON
2021 - a sustainable future for Wikidata

Ayush Garg, Shweata Hegde, Daniel Mietchen

https://github.com/petermr/docanalysis

https://europepmc.org/article/MED/34630447#free-full-text

# Our Team

**Dr Peter Murray-Rust**
*Reader Emeritus,*
*University of Cambridge*

**Dr Gitanjali Yadav**
*NIPGR (India), University of Cambridge (UK)*

**Lakshmi Devi Priya**
*BTech Student, Govt College of Technology*

**Kareena Singh**
*MSc Scholar,*
*Fergusson College*

**Dr. Ambreen Hamadani**
*PhD Scholar, Veterinarian*
*SKUAST-K*

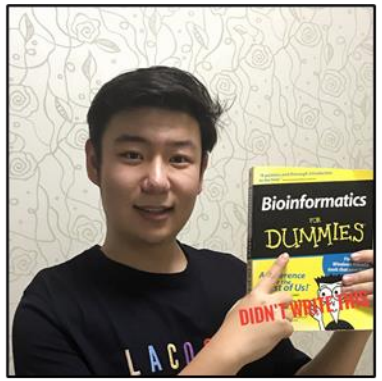
**Sana Saifi**
*B.Sc. Student*
*University of Delhi*

**Pruthivrajan Karunakaran**
*BTech Student*
*Dr. MGR Edu. & Res. Institute*

**Vaishali Arora**
*MSc Scholar,*
*University of Delhi*

CC BY

# Our Team




**Charles Zeyang Li**
*BA Hons, University of Cambridge*

**Vanisha Arora**
*MSc Scholar, Osmania University*

**Dheeraj Dhingani**
*BSc Student RR College Alwar*

**Pooja Pareek**
*MSc Scholar, Maharaja Ganga Singh University*

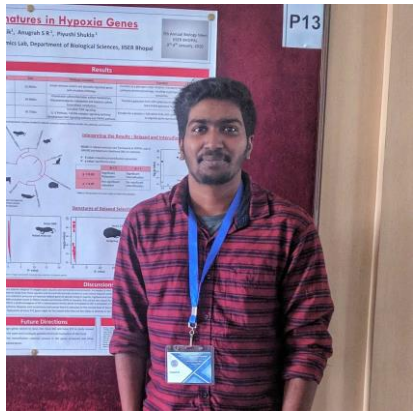
**Urja Biswas**
*MSc Scholar, University of Rajasthan*

**Om Prakash Mehra**
*MSc Scholar, MSJ college Bharatpur, Rajasthan*

**Jitu Ram Bhargav**
*MSc Scholar, University of Rajasthan*

**Simranleen Singh**
*MSc (Chemistry), Fergusson College*

CC BY




**Anugrah S. R.**
*BS-MS(Biological Sciences) IISER Bhopal*

**Shweata N. Hegde**
*BSc Student*
*Regional Inst. Of Edu. Mysore*

**Ayush Garg**
*GIIS Singapore*

**Kanishka Parashar**
MSc Biotech
Jamia Millia Islamia,Delhi

**Vasant Kumar**
M.Sc. Biotech,
Himachal Pradesh University, Summerhill

**Talha Hasan**
MSc Toxicology
Jamia Hamdard University

**Radhu Kantilal Ladani**
Msc Bioinformatics
Rajiv Gandhi Institute, Pune.

**Prachi Singh**

Undergrad (Botany), JDB Girls Govt. College, Kota

**Anam Anjum**

Masters (Botany) Student, JDB Government Girls Science College Kota

**Anjali Shekhawat**

Masters (Botany) Student, Jai Narain Vyas University, Jodhpur, Rajasthan, India.

**Kumari Rageshwari**

Masters (Biotech) Student, Botany Department in University of Rajasthan

Anubhab Chakraborty

Undergrad at IISER Bhopal

Ankit Anil Lade

Undergrad at IISER Bhopal

**Utkarsha N. Pande**

Undergrad at IISER Bhopal

# Thanks for your attention!

Thanks to NIPGR and DST Rajasthan
Mentors, openVirus, CEVOpen Team, Global Volunteers and Wikimedia Community!

Credits
Image credits: Microsoft Powerpoint, Wikimedia Commons

WIKI DATA CON
2021 - a sustainable future for Wikidata