Europe PMC

> 39 million articles

docanalysis demo

Automatically download 1000s of papers, extract scientific data

Ayush Garg,

High School,

Singapore

Shweata N. Hegde,

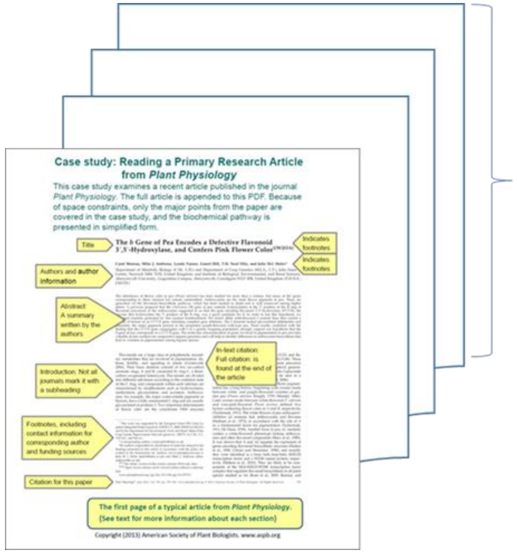
3rd year undergrad (plant science & education)

India

Early Career Researchers

MADICES

Query “XANES AND EXAFS AND XRD”



Case study: Reading a Primary Research Article from *Plant Physiology*

This case study examines a recent article published in the journal *Plant Physiology*. The full article is appended to this PDF. Because of space constraints, only the major points from the paper are covered in the case study, and the biochemical pathway is presented in simplified form.

Title

The *b* Gene of Pea Encodes a Defective Flavonoid 3′,5′-Hydroxylase, and Confers Pink Flower Color

Indicates footnotes

Indicates footnotes

Authors and author information

Abstract A summary written by the authors

In-text citation Full citation: is found at the end of the article

Introduction: Not all journals mark it with a subheading

Footnotes, including contact information for corresponding author and funding sources

Citation for this paper

The first page of a typical article from *Plant Physiology*. (See text for more information about each section)

Copyright (2013) American Society of Plant Biologists. www.aspb.org

> 1, 000 papers

`docanalysis`

Dictionary of X-ray Spectroscopy

Personalized search system

Structured data

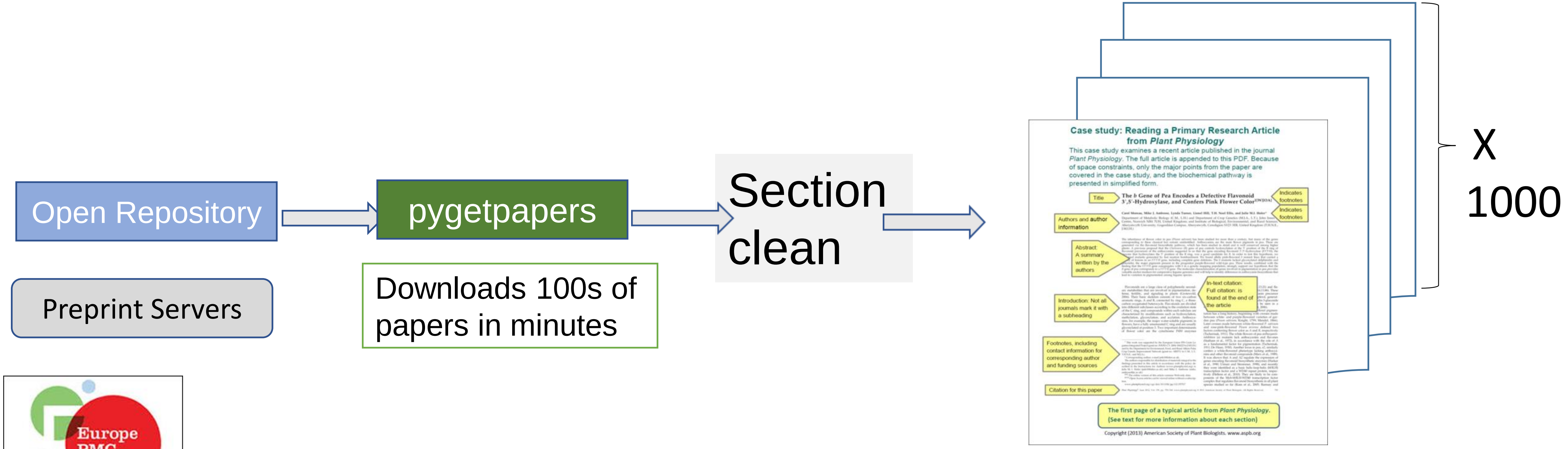
docanalysis – a command line tool

Automatically download 1000s of papers, extract scientific data

```
pip install docanalysis
```

Step 1: Download papers

```
docanalysis --run_pygetpapers -q “"XANES AND EXAFS AND XRD" " -k 10 --project_name xanes_madices
```



Open Repository
pygetpapers
Section clean
Preprint Servers
X
1000
Title
Authors and author information
Abstract: A summary written by the authors
Introduction: Not all journals mark it with a subheading
Footnotes, including contact information for corresponding author and funding sources
Citation for this paper
Indicates footnotes
Indicates footnotes
In-text citation: Full citation: is found at the end of the article

Downloads 100s of papers in minutes

Case study: Reading a Primary Research Article from *Plant Physiology*

This case study examines a recent article published in the journal *Plant Physiology*. The full article is appended to this PDF. Because of space constraints, only the major points from the paper are covered in the case study, and the biochemical pathway is presented in simplified form.

The *b* Gene of Pea Encodes a Defective Flavonoid 3′,5′-Hydroxylase, and Confers Pink Flower Color[[W][OA]]

The first page of a typical article from *Plant Physiology*. (See text for more information about each section)

Copyright (2013) American Society of Plant Biologists. www.aspb.org

On local machines.

Europe PMC

Free information resource for biomedical and health researchers

Has millions of open access papers.

Step 2: Section papers

```
docanalysis --project_name xanes_madices --run_sectioning
```

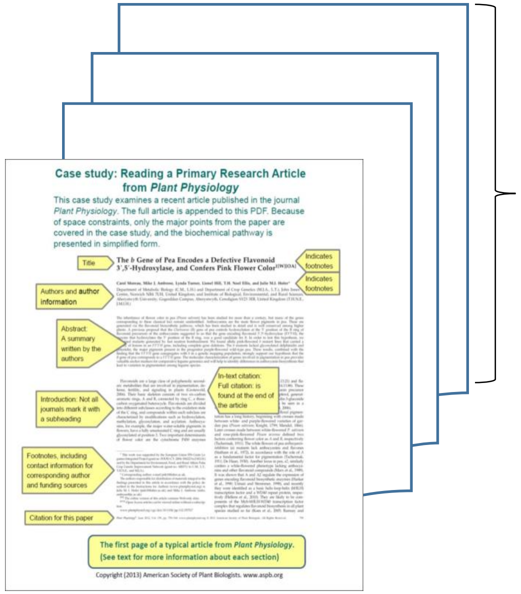
- Article Sections

- <**JATS**> : Journal Article Tag Suite

Slide credits: Ambreen Hamadani

Step 3: Extract Entities

```
docanalysis --project_name xanes_madices --run_sectioning --output entities_20220219.csv
```



Case study: Reading a Primary Research Article from *Plant Physiology*

This case study examines a recent article published in the journal *Plant Physiology*. The full article is appended to this PDF. Because of space constraints, only the major points from the paper are covered in the case study, and the biochemical pathway is presented in simplified form.

Title

The *b* Gene of Pea Encodes a Defective Flavonoid 3',5'-Hydroxylase, and Confers Pink Flower Color[1][W][OA]

Indicates footnotes

Indicates footnotes

Authors and author information

Abstract: A summary written by the authors

In-text citation: Full citation: is found at the end of the article

Introduction: Not all journals mark it with a subheading

Footnotes, including contact information for corresponding author and funding sources

Citation for this paper

The first page of a typical article from *Plant Physiology*. (See text for more information about each section)

Copyright (2013) American Society of Plant Biologists. www.aspb.org

SciSpacy/other NLP tools

AUTOMATIC
1 paper/s

The experimental procedures followed the rules

Background: The purpose of this study is to analyze the surface morphology and elemental composition of zirconia implants before and after photofunctionalization.

100s of Entities

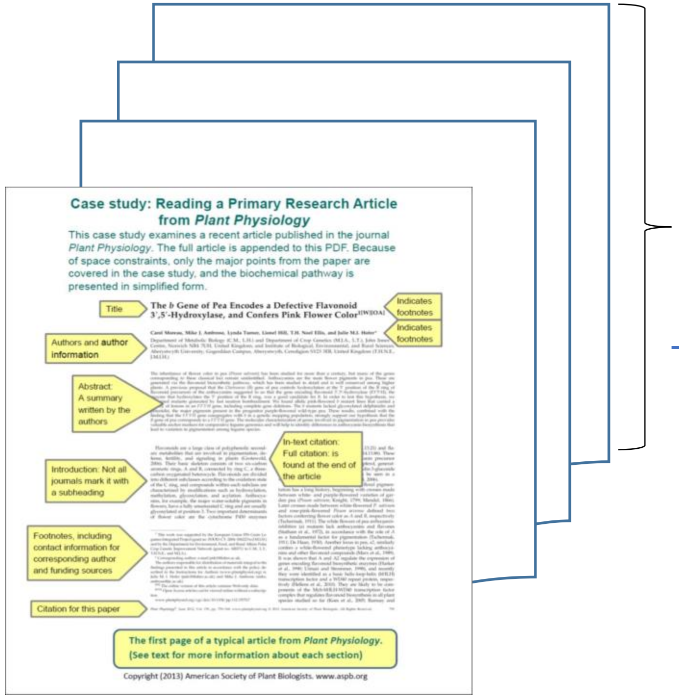
10 papers on XANES
AND EXFAS AND XRD

Ayush Garg, Shweata Hegde, Daniel Mietchen

https://github.com/petermr/docanalysis

Step 4: Make dictionary

```
docanalysis--project_name xanes_madices --make_ami_dict entities_20220209
```



Case study: Reading a Primary Research Article from *Plant Physiology*

This case study examines a recent article published in the journal *Plant Physiology*. The full article is appended to this PDF. Because of space constraints, only the major points from the paper are covered in the case study, and the biochemical pathway is presented in simplified form.

Title
Indicates footnotes
Indicates footnotes
Authors and author information
Abstract: A summary written by the authors
In-text citation: Full citation: is found at the end of the article
Introduction: Not all journals mark it with a subheading
Footnotes, including contact information for corresponding author and funding sources
Citation for this paper

The *b* Gene of Pea Encodes a Defective Flavonoid 3′,5′-Hydroxylase, and Confers Pink Flower Color[1][W][OA]

The first page of a typical article from *Plant Physiology*. (See text for more information about each section)

Copyright (2013) American Society of Plant Biologists. www.aspb.org

10 papers on XANES AND EXAFS AND XRD

```
'surface', 30,
 'redox', 25,
'NIB', 18,
'sXAS', 16,
'mRIXS', 15
'Fe', 13,
'TM', 13,
'DNN', 12,
'MSE', 11,
'capacity', 11,
'surfaces', 10,
'Mn', 9,
'spectra', 7,
'Ti', 7,
'N', 7,
'Al', 7,
'XANES spectra', 6…
```

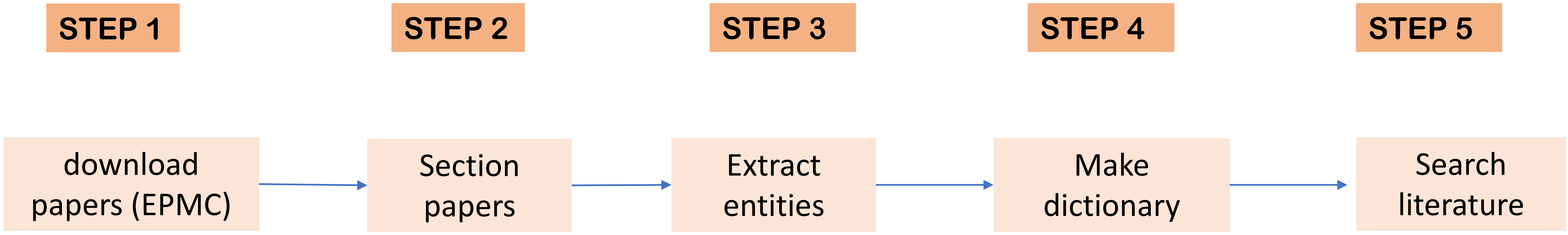
Snippet of the dictionary

Annotate the literature

docanalysis – Command line Tool

Automatic!

< 2 secs/paper



STEP 1

download papers (EPMC)

STEP 2

Section papers

STEP 3

Extract entities

STEP 4

Make dictionary

STEP 5

Search literature

`--run_pygetpapers`

`--run_sectioning`

`--output`

`--make_ami_dict`

Ayush Garg

Peter Murray-Rust

https://allenai.github.io/scispacy/

```
pip install docanalysis
```

```
docanalysis --run_pygetpapers -q "XANES AND EXAFS AND
XRD" -k 10 --project_name xanes_exafs_xrd --
run_sectioning --output entities_202202019.csv --
make_ami_dict entities_20220209
```

Special thanks to Ayush Garg, Anubhab Chakraborty and Peter Murray-Rust