

# Sophie Chen

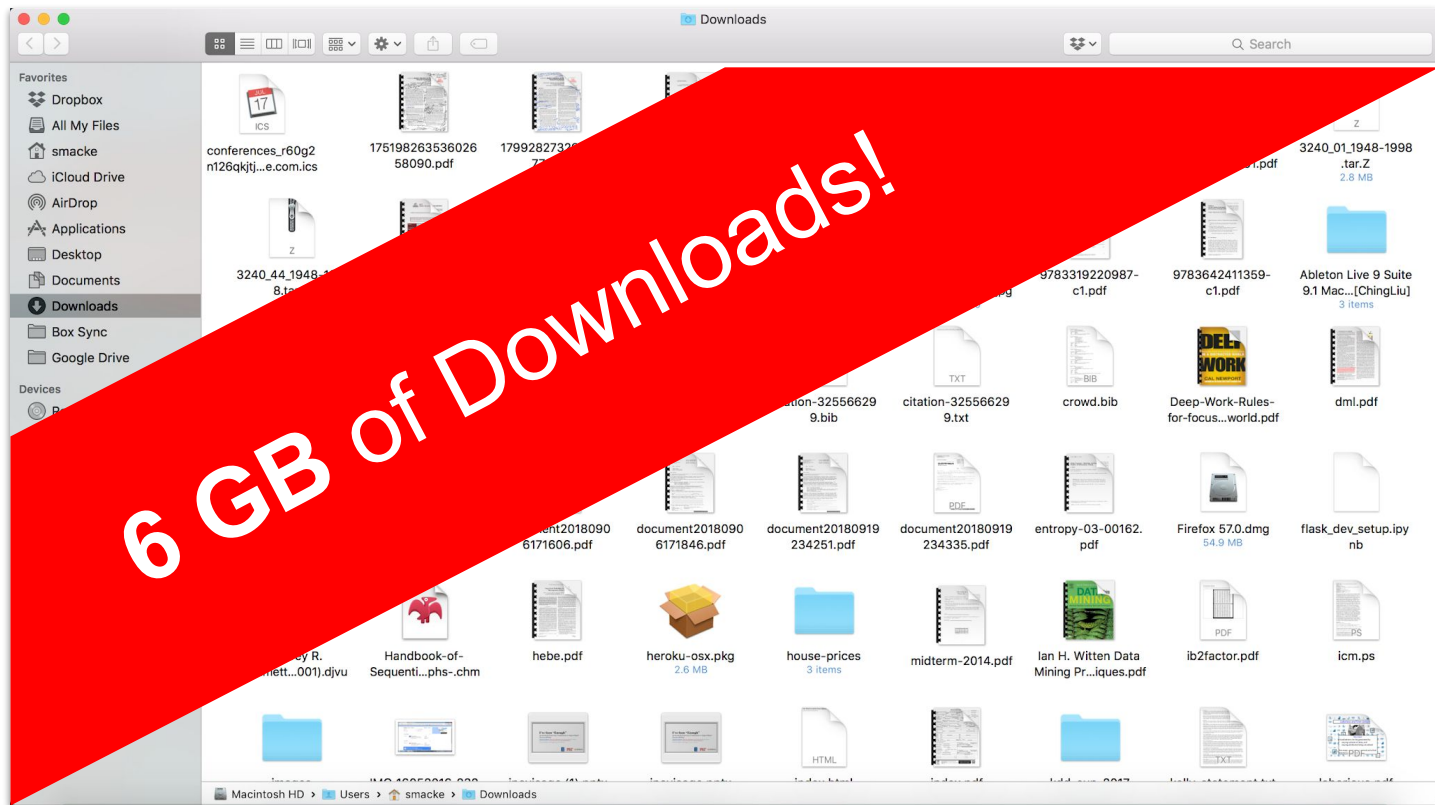


## *AI Folder Genie*

**Download Files where they Belong!**

# *Disorganized* File Structure!

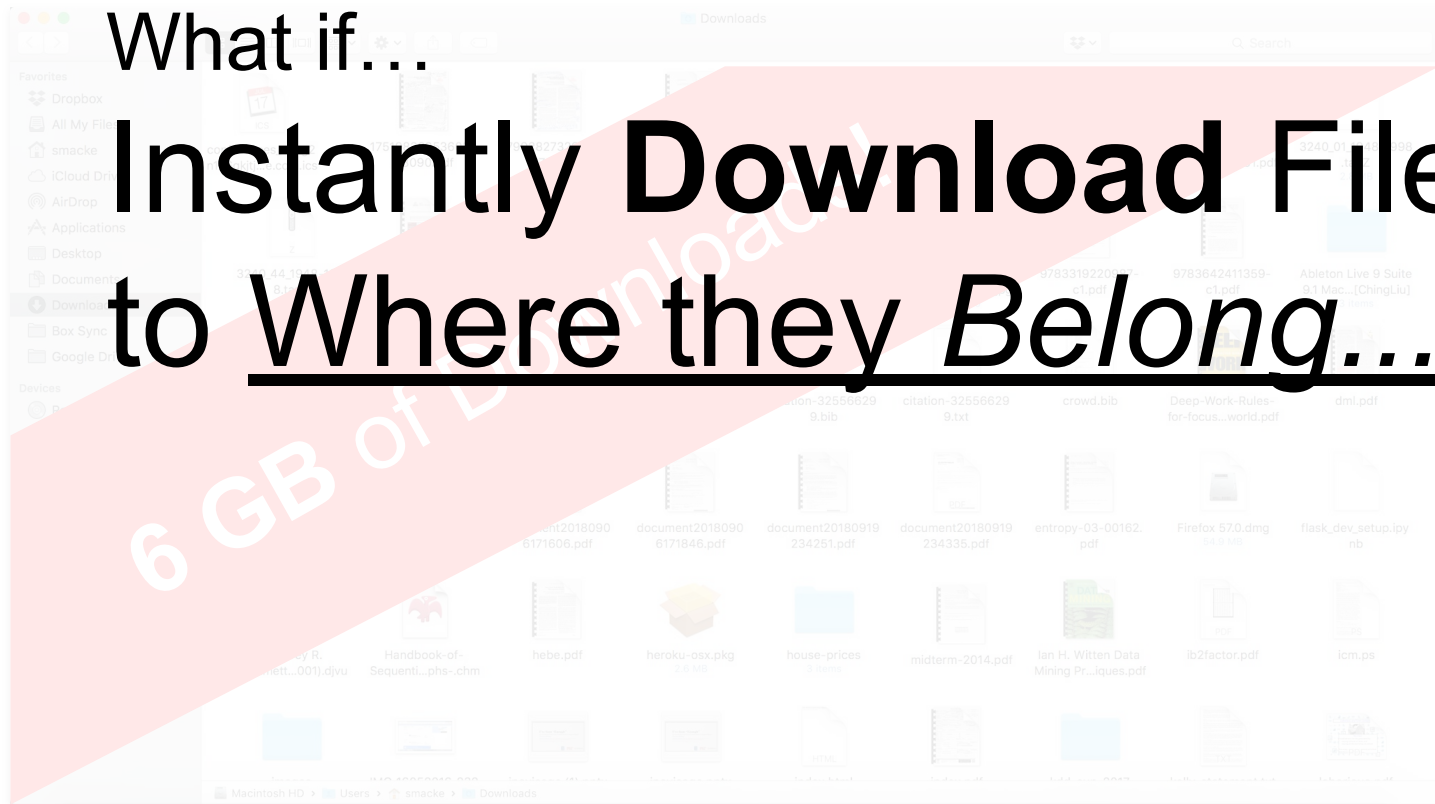
## Too many files in Downloads Folder



*Disorganized* File Structure!  
Too many files in Downloads Folder

What if...

**Instantly Download Files  
to Where they *Belong*...**





*AI Folder Genie*



available in the  
chrome web store

Predict the ***Best*** Folder/Sub-Folder  
for *YOUR* Download,  
for *YOUR* File System!

# Text Classification

A circular graphic with a dark blue background. Inside the circle is a white laptop with a blue screen displaying the text "16 K files". To the left of the laptop is a circular inset showing a portrait of a smiling woman with glasses and brown hair.

16 K files

- **File** names (**features**)
- **Folder** names (**labels**)

Processed **file names** with  Natural Language Analyses with NLTK and 

'CP 2006 Theoretical potential energy surfaces for excited mercury trimers.pdf'

'cp', '2006', 'theoret', 'potenti', 'energi', 'surfac', 'for', 'excit', 'mercuri', 'trimer', 'pdf'

# Extract the Features



**CS**  
ipynb  
output



**Pictures**  
img  
jpg



**Videos**  
mp4  
v1

# Hierarchical Labels

## Files in:

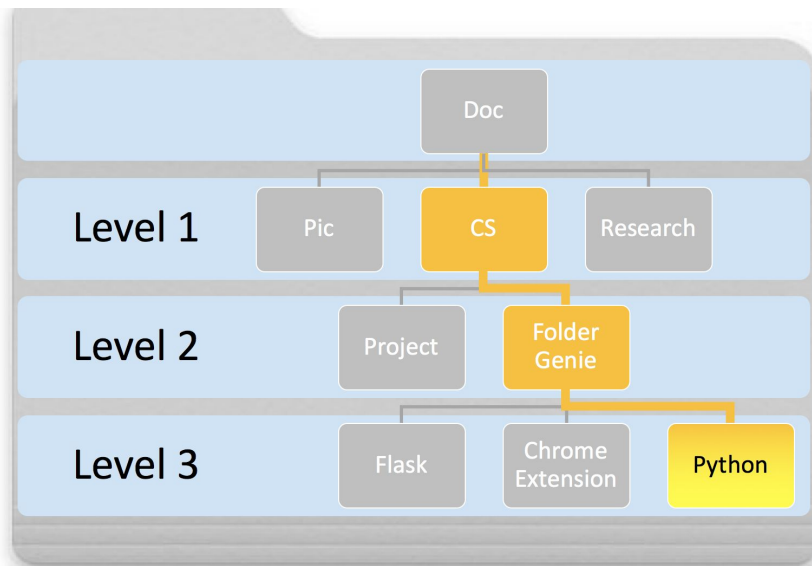
Doc/CS/FolderGenie/Python

## Hierarchical labels:

Level 1: CS

Level 2: CS/FolderGenie

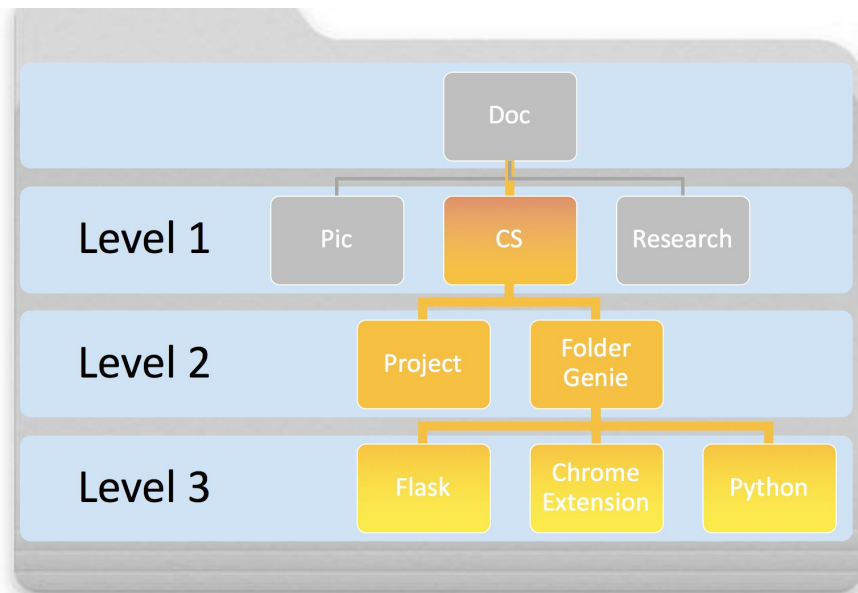
Level 3: CS/FolderGenie/Python





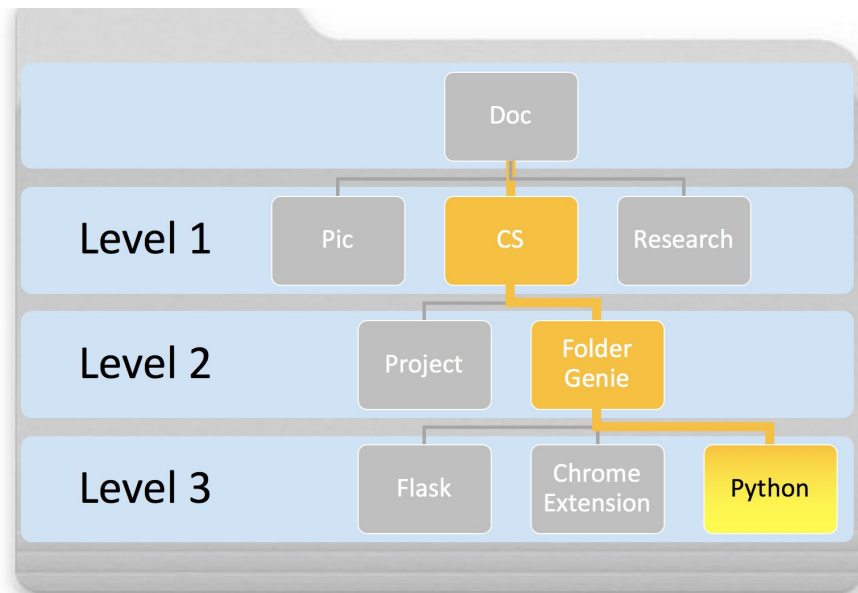
# Hierarchical Classification

**Softmax classifier** trained at each folder

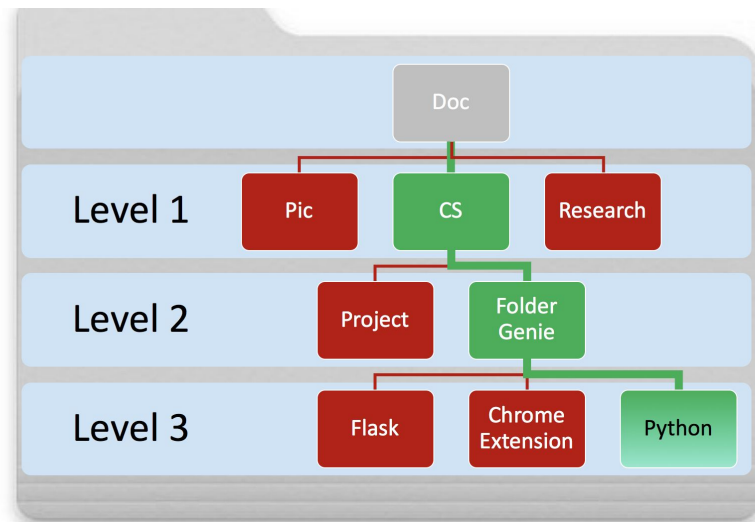
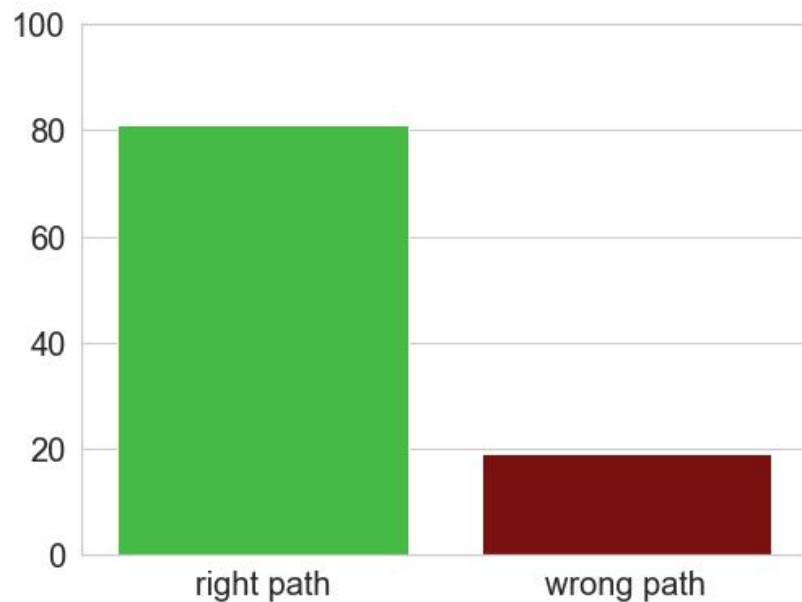


# Hierarchical Classification

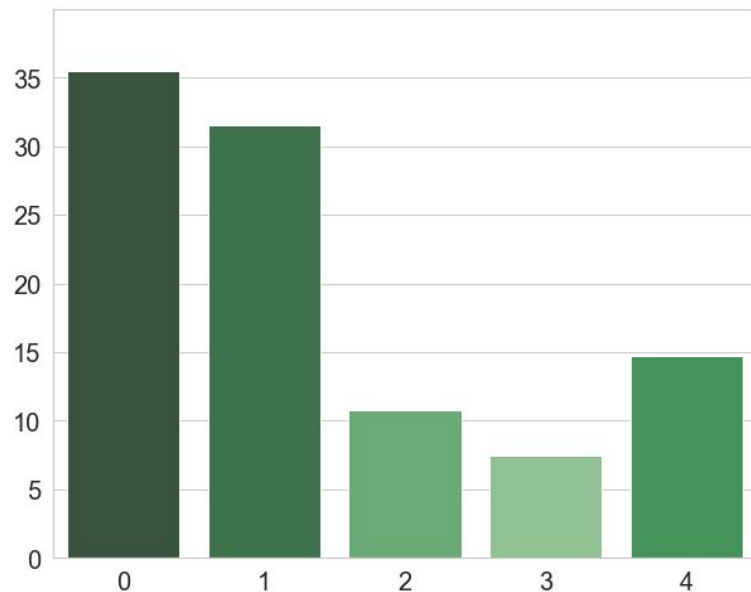
Predictions only advance deeper when confident



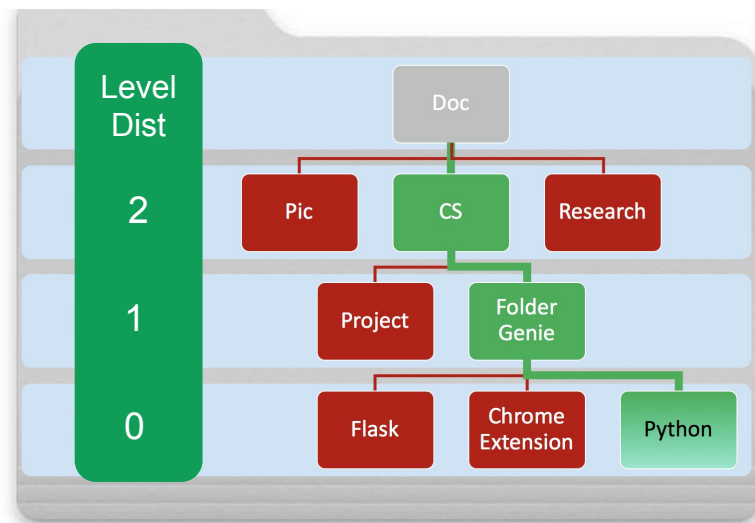
81% go to *Right Path!*



69% go to within 1 level distance!



Level distance = actual - predicted



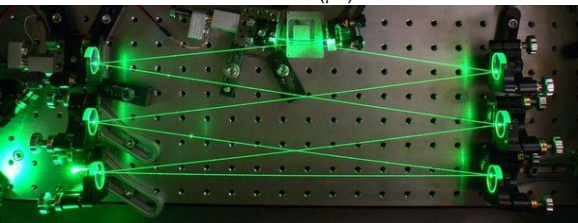
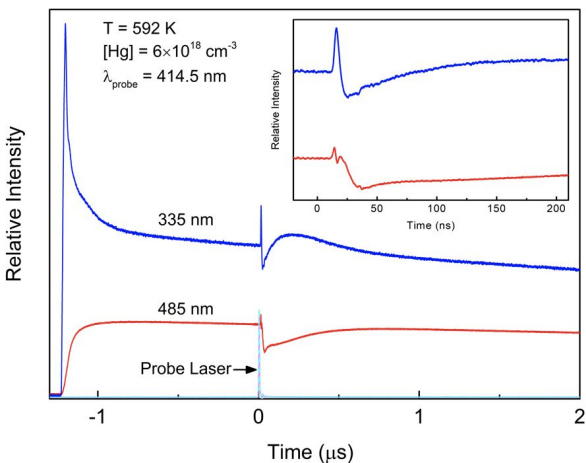
local file folder levels: 9 in total  
first 5 considered for predictions

# Sophie Chen



PhD in Electrical & Computer  
Engineering, Spectral Analysis

Github: SophieGarden  
Linkedin: Sophie-Chen-Data



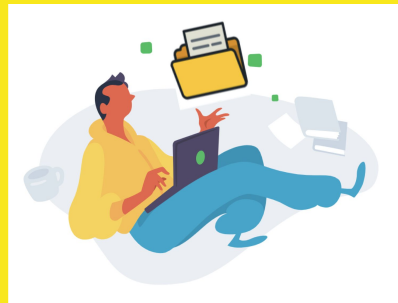
# Thank you for attending my talk!

## Welcome to try

### *AI Folder Genie*



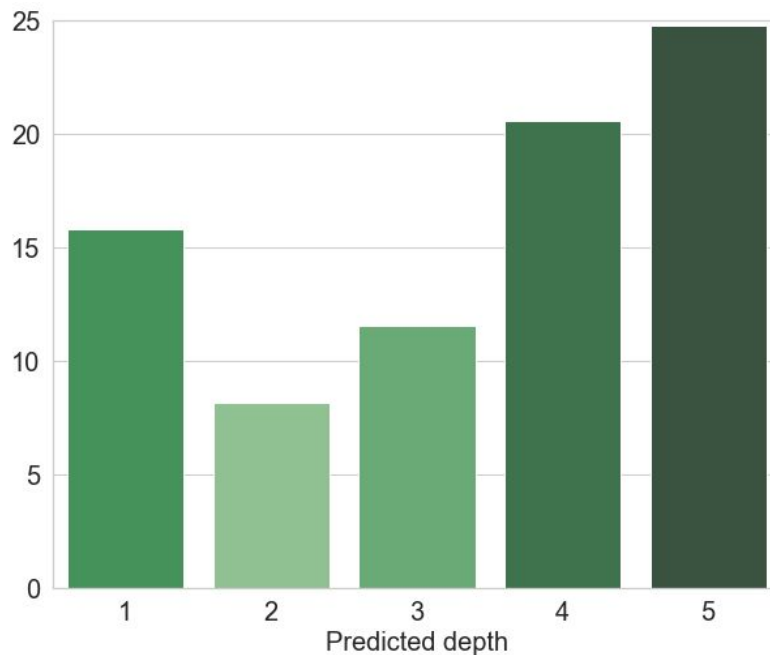
available in the  
chrome web store



# Future Plans

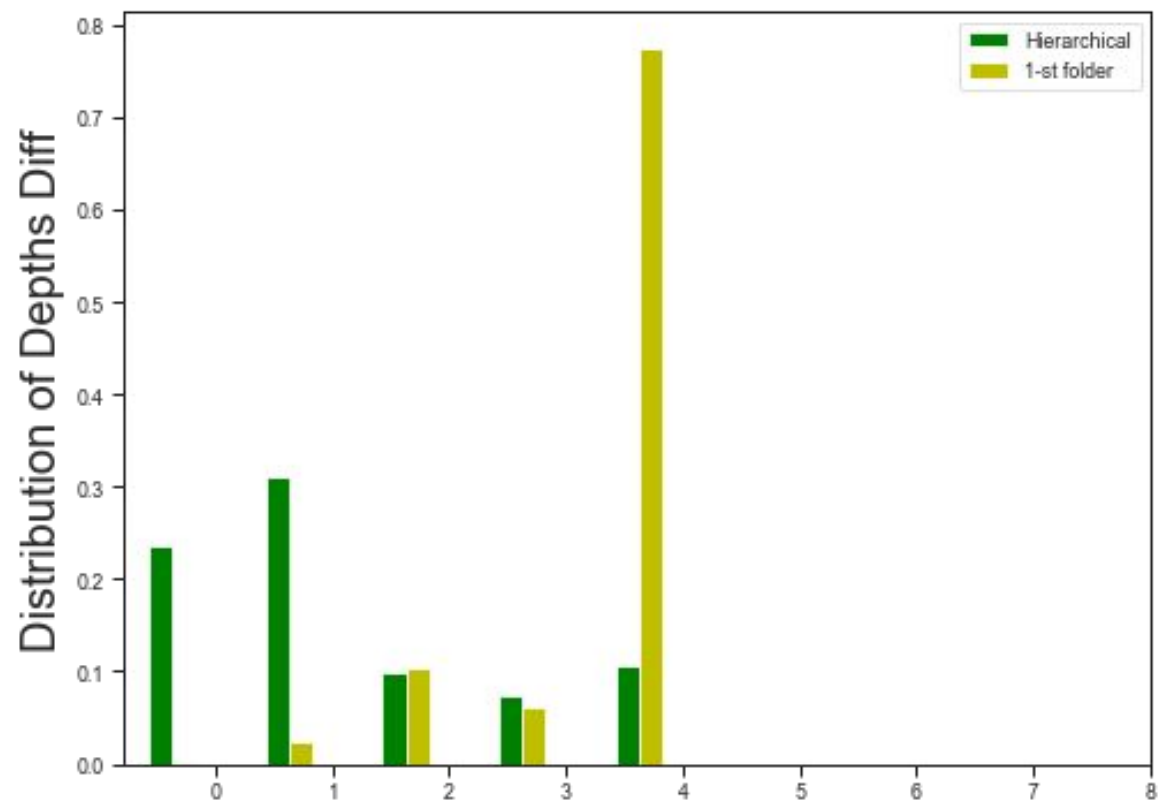
- Chrome Extension + Desktop App: sync between all devices, Dropbox...
- Output a list of choices
- Features: metadata, NLP folder names, included folder names as part of features

# Predicted Depths



local file folder levels: 9 in total;  
first 5 considered for predictions





'CP 2006 Theoretical potential energy surfaces for excited mercury trimers.pdf'

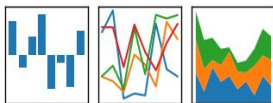
'cp', '2006', 'theoret', 'potenti', 'energi', 'surfac', 'for', 'excit', 'mercuri', 'trimer', 'pdf'



Natural Language  
Analyses with NLTK

pandas

$$y_{it} = \beta' x_{it} + \mu_i + \epsilon_{it}$$



**Original File Name**

**Remove - \_ .**

**Tokenize**

**Stem**

**Remove Stop Words**

**tf-idf**

# Feature Engineering

['CP-2006-Theoretical-potential-energy-surfaces for excited mercury trimers.pdf']

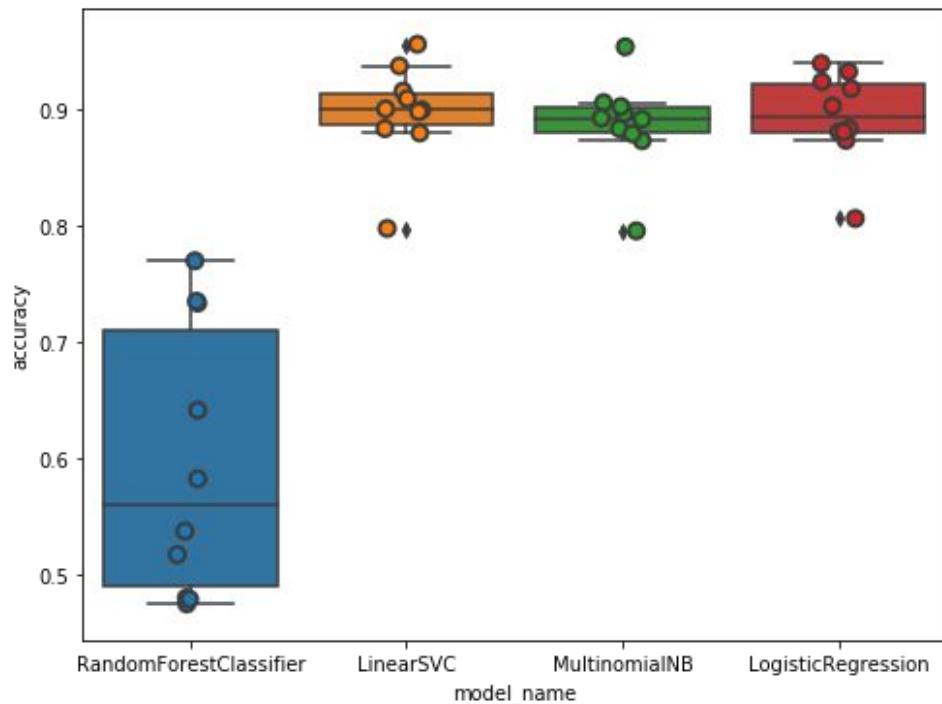
['cp-2006-theoretical-potential-energy-surfaces for excited mercury trimers.pdf']

[['cp', '2006', 'theoretical', 'potential', 'energy', 'surfaces', 'for', 'excited', 'mercury', 'trimers', 'pdf']]

[['cp', '2006', 'theoret', 'potenti', 'energi', 'surfac', 'for', 'excit', 'mercuri', 'trimer', 'pdf']]

['cp 2006 theoret potenti energi surfac for excit mercuri trimer pdf']

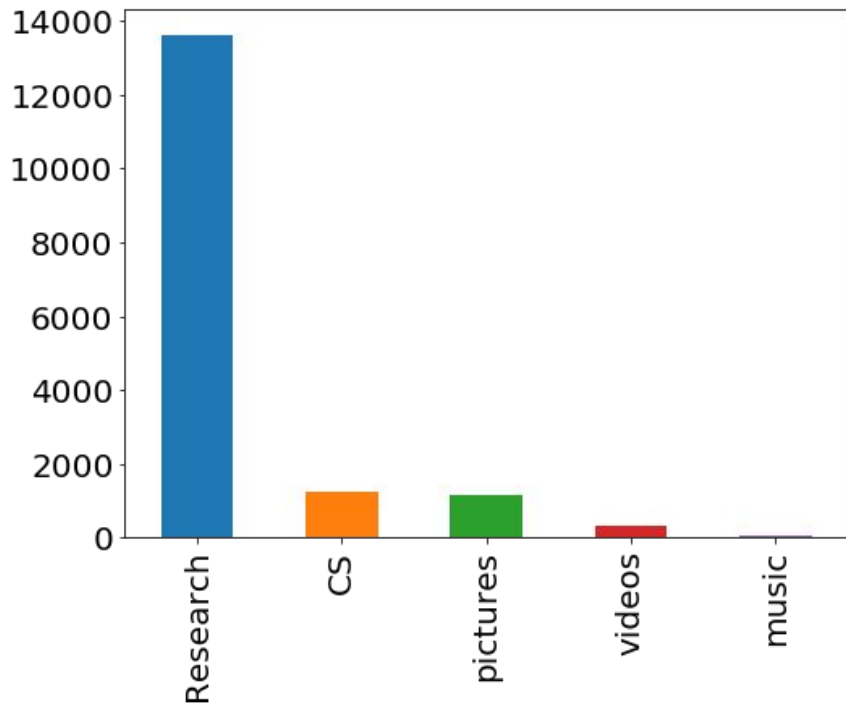
# Multi-labeled Logistic Regression Text Classification



Logistic regression:  
Fast, stable, **with predict\_prob**

**1st-depth AUC >0.9**

# Text Classification: Softmax Regression



## Text Classification:

- Logistic Regression
- 1st-Folder AUC 90%

## Challenges:

- Varied depth of folders
- 2st-Folder AUC 65%
- Files live in interior nodes

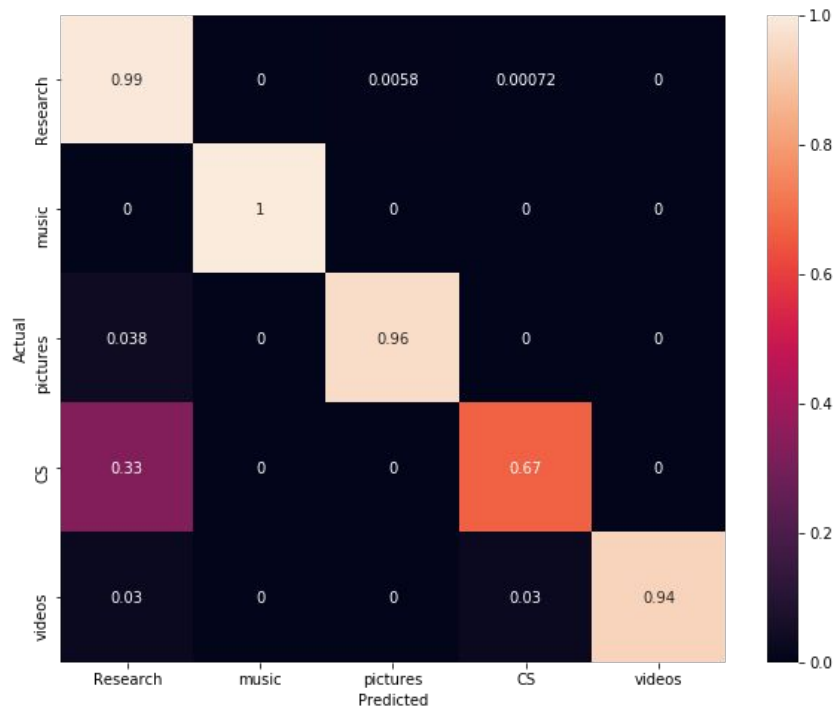
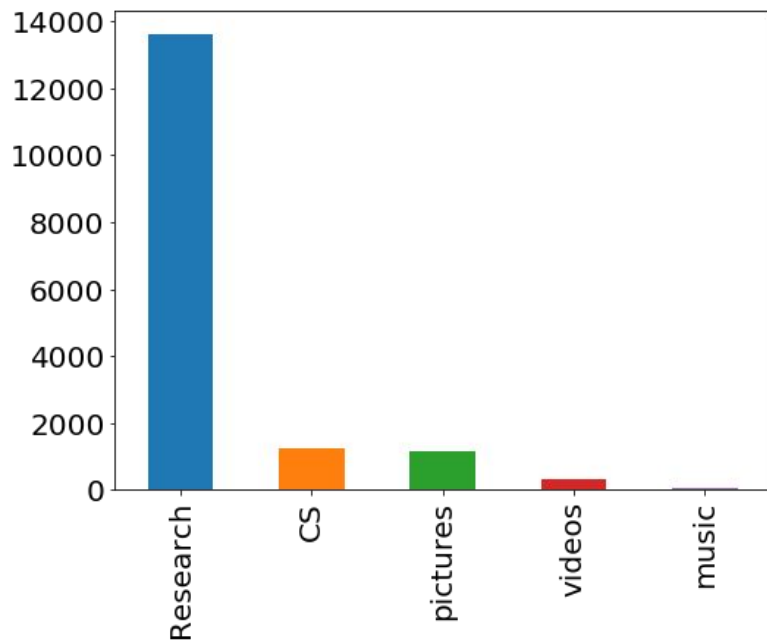
## Solutions:

- Hierarchical Classification

# Text Classification: Softmax Regression

- 16400+ files in total

- 1st split accuracy > 90%



# How does it work?

Data

Feature Engineering

Label Engineering

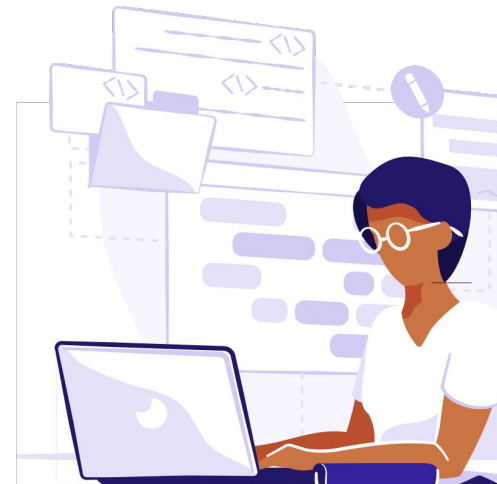
Classifier Picking

Hierarchical Classification (wrote my own algorithm)

Move files (symbolic links)


# User Cases

- Email Attachments
- Social Media Photos
- Cloud Shared Files
- Github Codes
- Digital Music
- ...





# AI Folder Genie Chrome Extension



AI Folder Genie for Downloading Files 1.0.0



AI Folder Genie: Magically download files to the right folder!

ID: liaikibdkljfddhlnfbmpacfmmapomdoc

Inspect views [background page \(Inactive\)](#)

Details

Remove

## Set Paths

### Root Directory:










e.g. /Users/Documents/


### Local Downloads Directory:

e.g. /Users/Downloads/

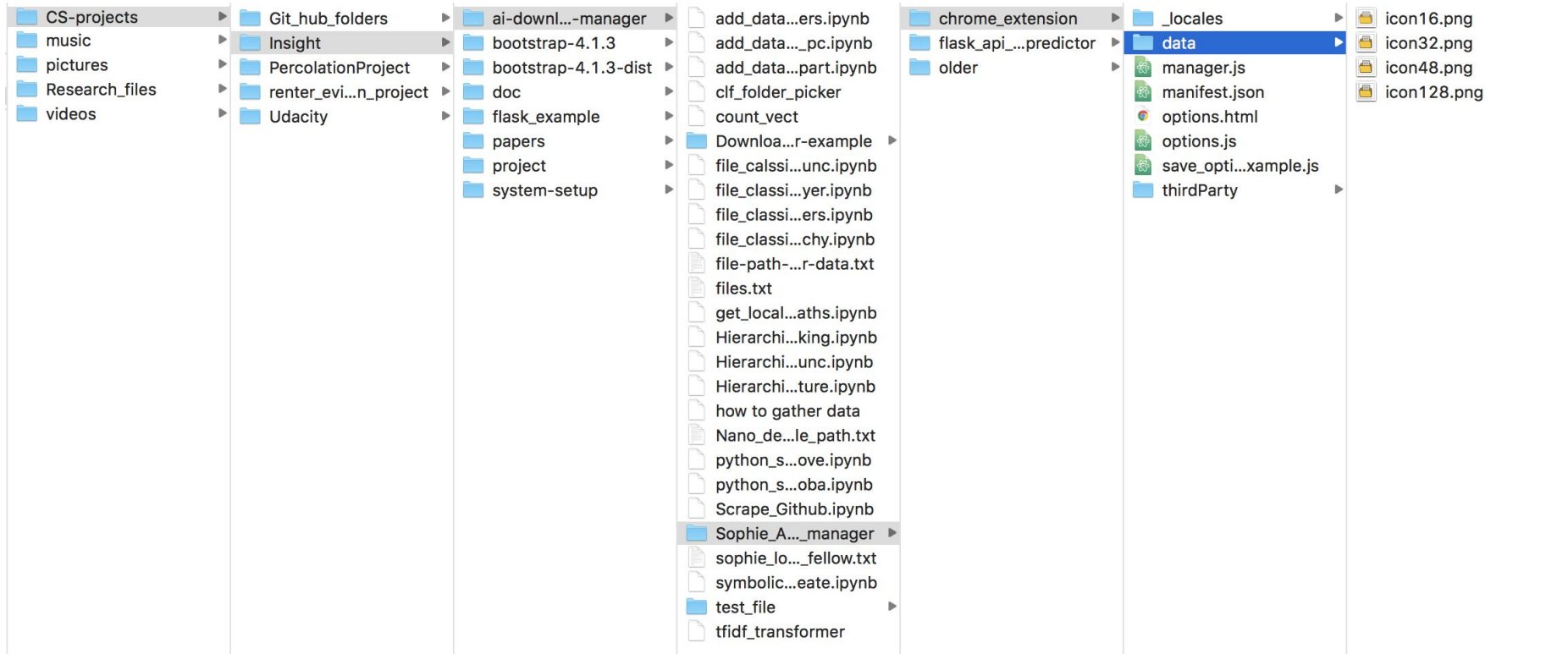
Save

Downloads

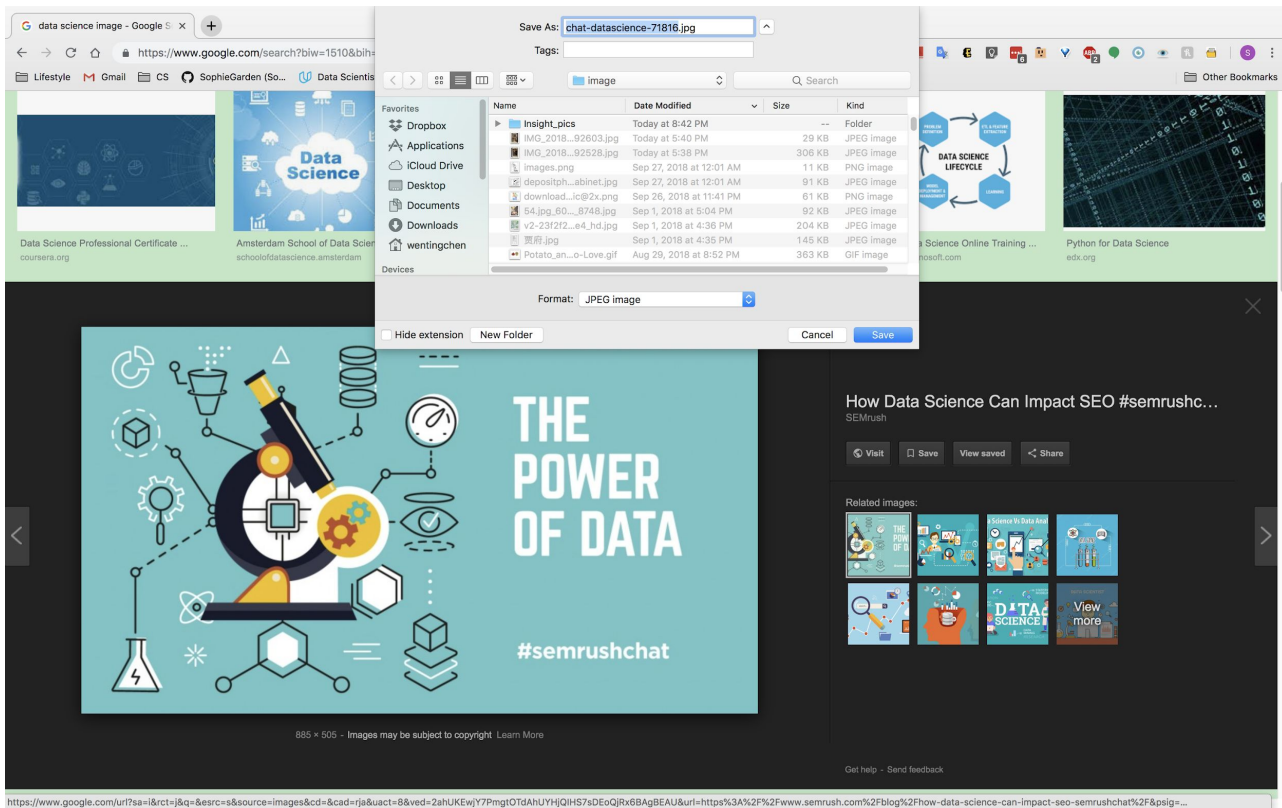


Today	Size	Kind
 symlink_folder	43 bytes	Alias

# Sophie's Folder Structure



# 2 levels: /pictures/image/



# 1 level: /Research\_files/

