Sophie Chen

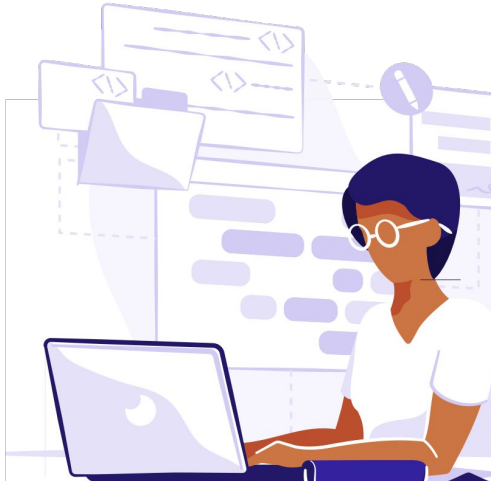
AI Folder Genie

Download Files where they Belong!

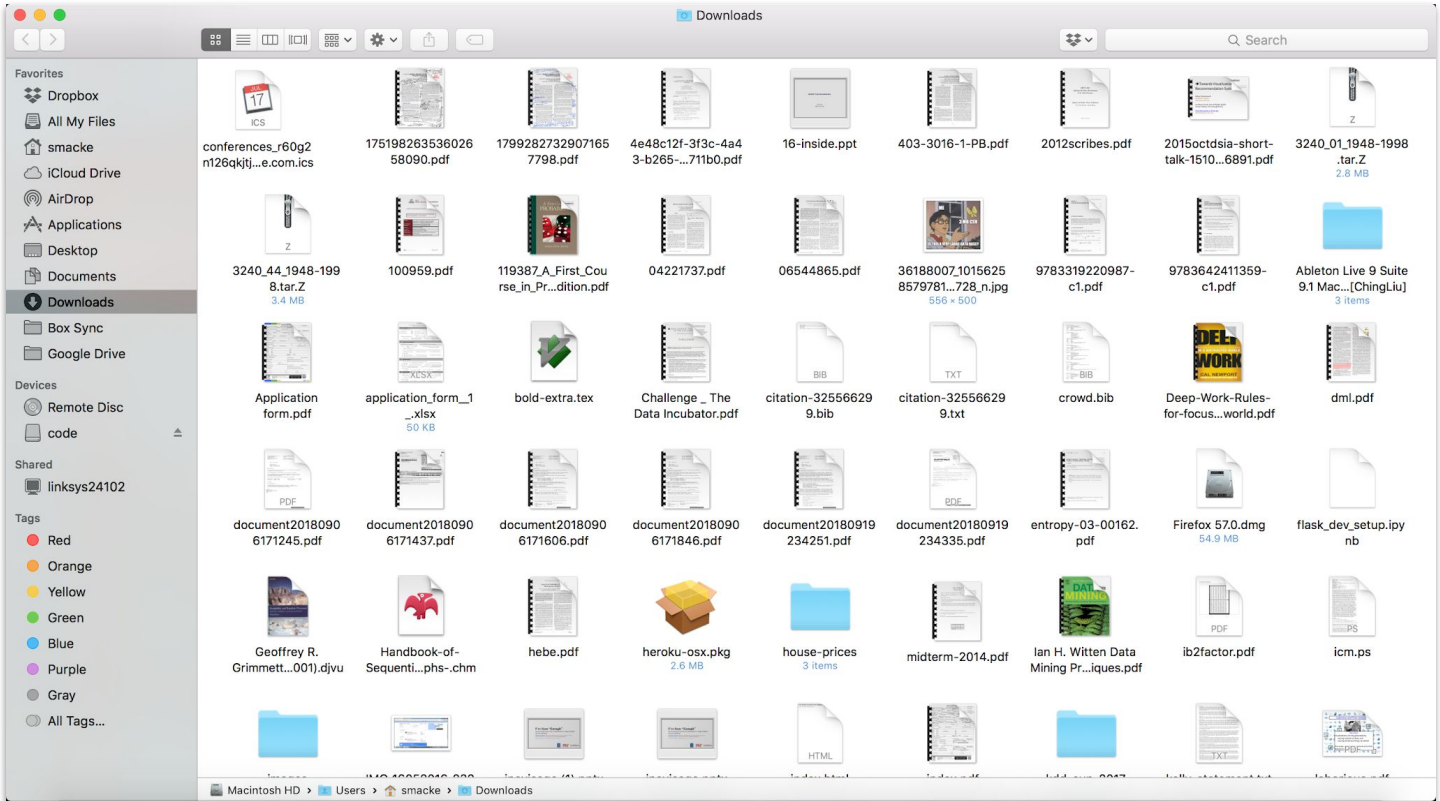
Download Every Day

Downloads :

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

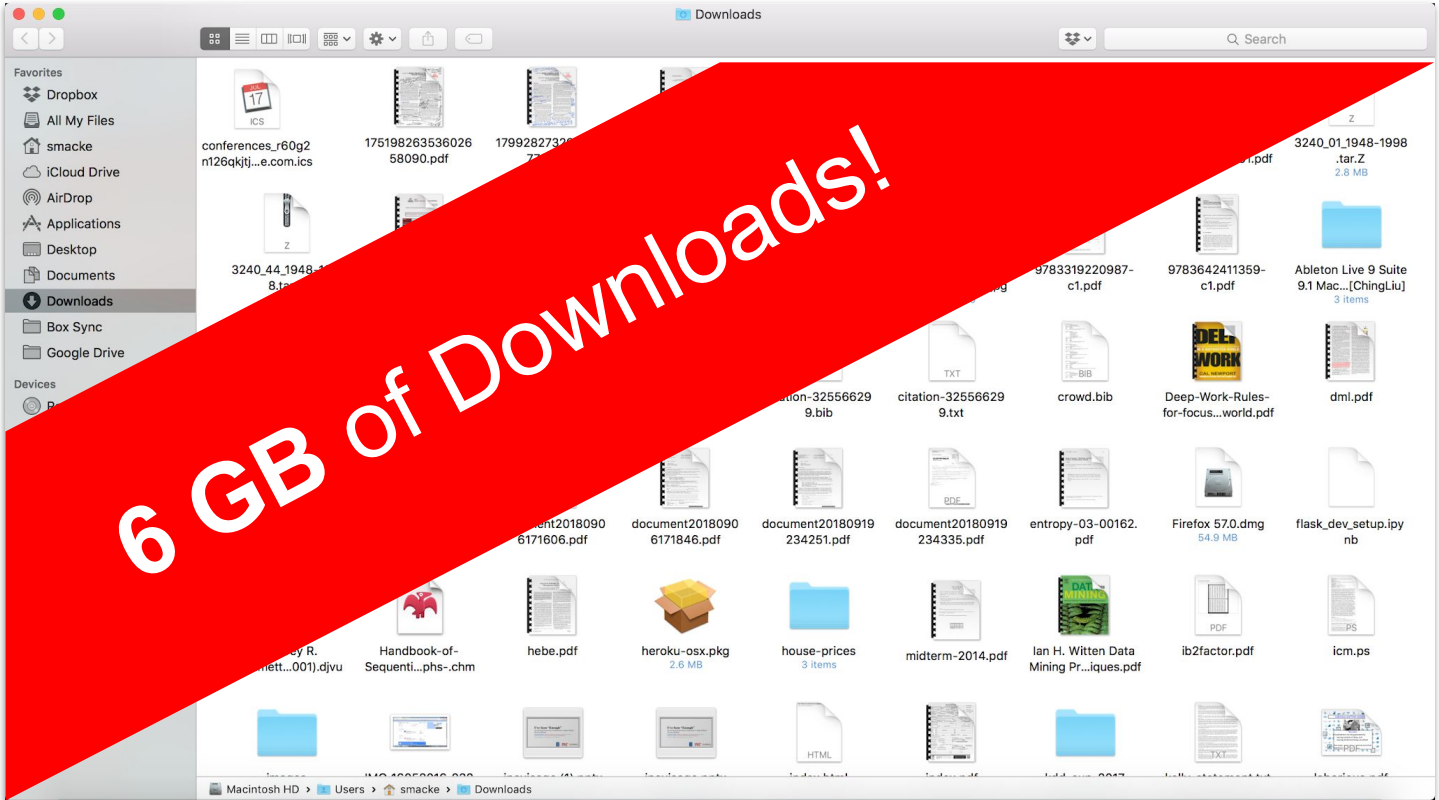


All go to *The Downloads Folder*



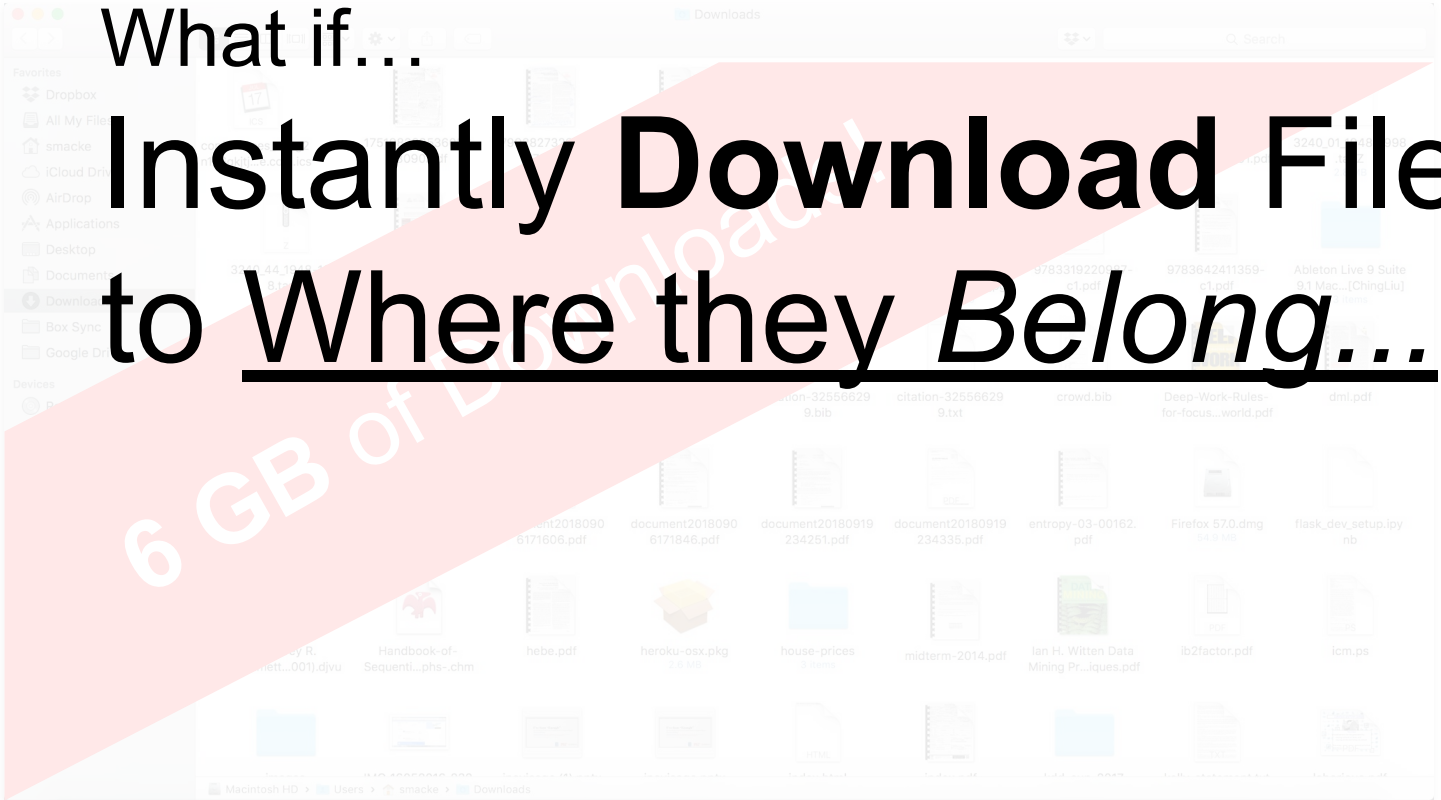
Disorganized File Structure!

Too many files in Downloads Folder



Disorganized File Structure!

Too many files in Downloads Folder



What if…

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

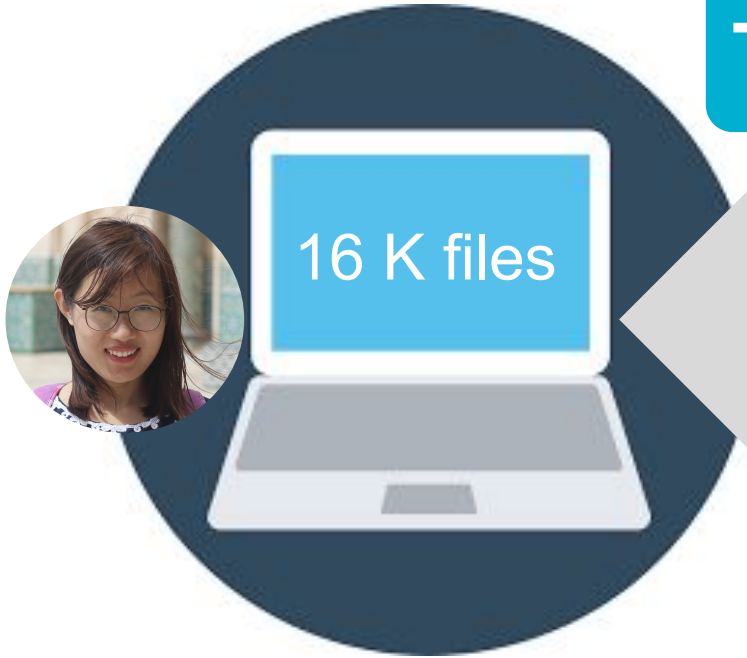
AI Folder Genie

available in the chrome web store*

Predict the ***Best*** Folder/Sub-Folder

for *YOUR* **File Download**,

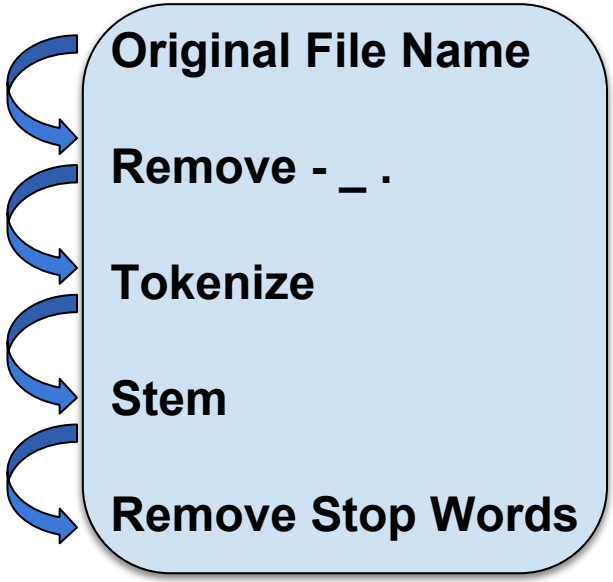
for *YOUR* **File System**!



Text Classification

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

Processed file names with Natural Language Analyses with NLTK and scikit learn



tf-idf

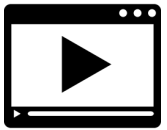
'CP 2006 Theoretical potential energy surfaces for excited mercury trimers.pdf'

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

Extract the Features

CS/Insight
hierarch
bootstrap

Pictures/Image
jpg
larg



videos/wild_china
wild
china

Naive Level-Flattened Labels

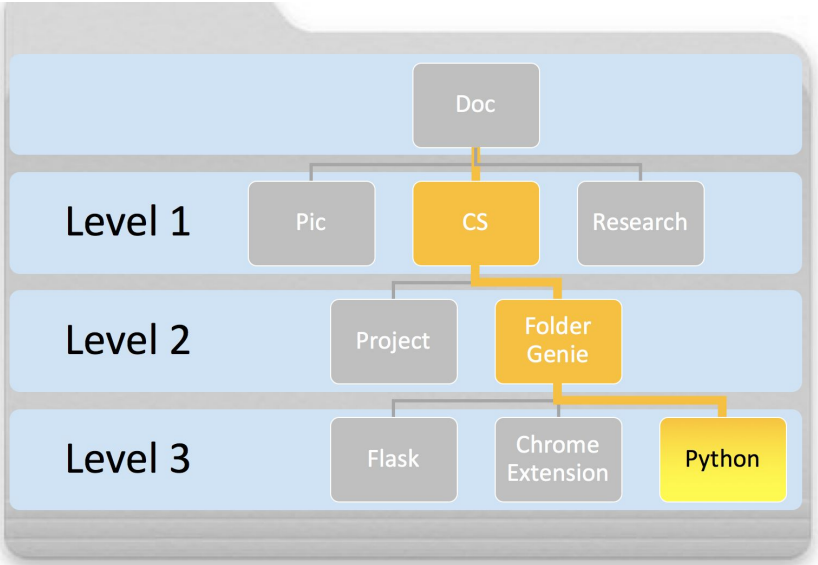
Files in:

`Doc/CS/FolderGenie/Python`

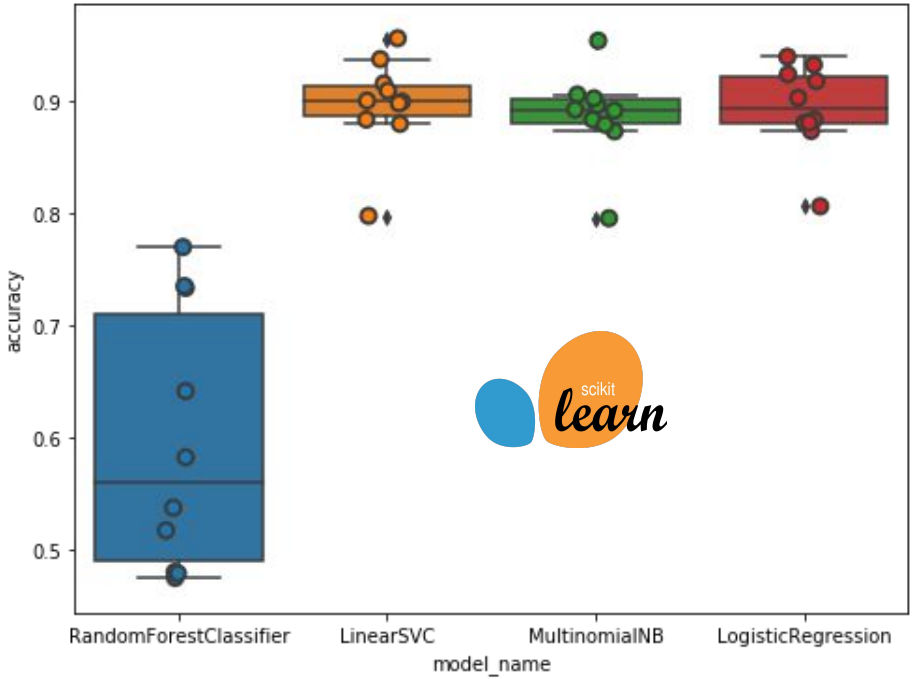
Level-flattened labels:

Level 3: `CS/FolderGenie/Python`

Flatten to desired level 3



Naive Level-Flattened Classification



Multilabel classifier

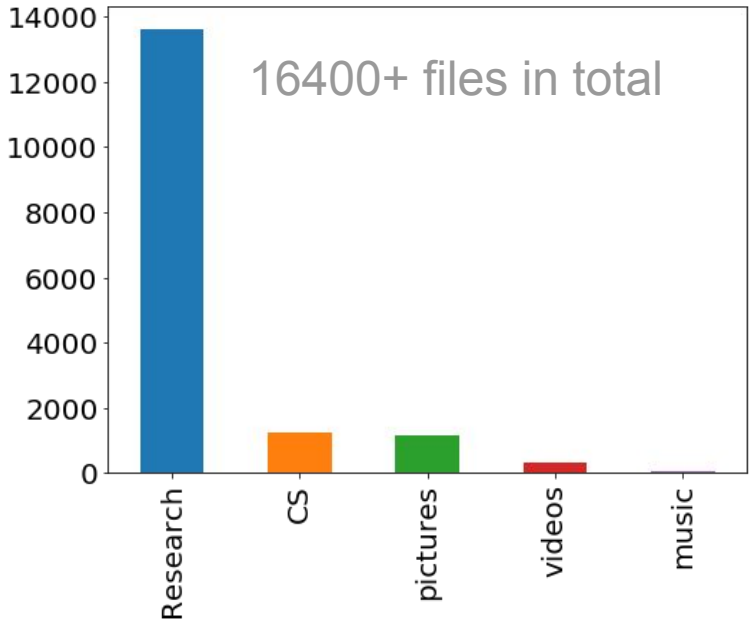
Logistic regression:

Fast, stable, with predict_prob

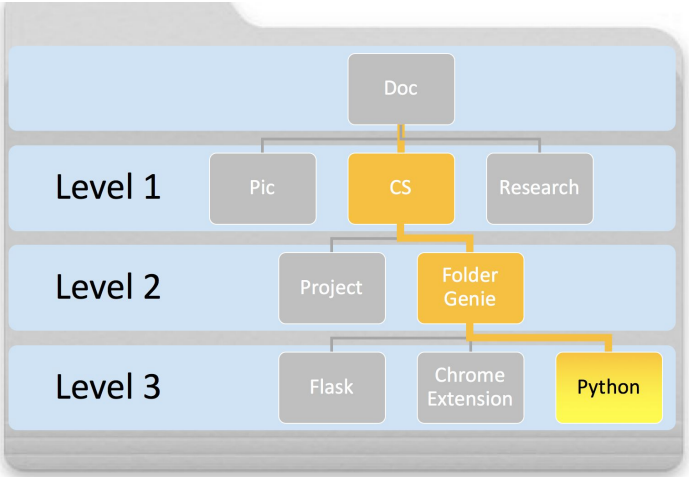
1st-depth accuracy > 90%

Shortcomings

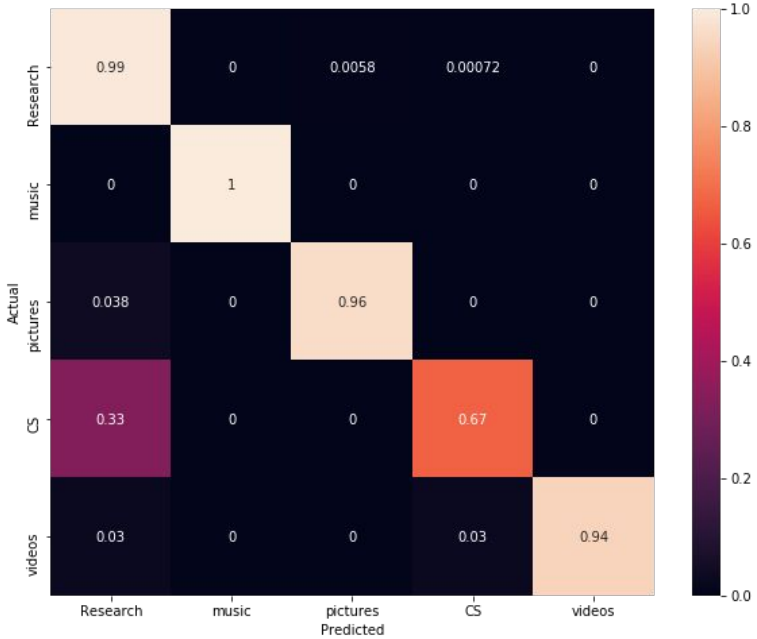
Unbalanced data



Varied prediction difficulty



Deeper Level Predictions have Low Accuracy



Problem:

- 1st level accuracy 90%
- *2nd level* accuracy 65%

Solution:

- Hierarchical Classification

Hierarchical Labels

Files in:

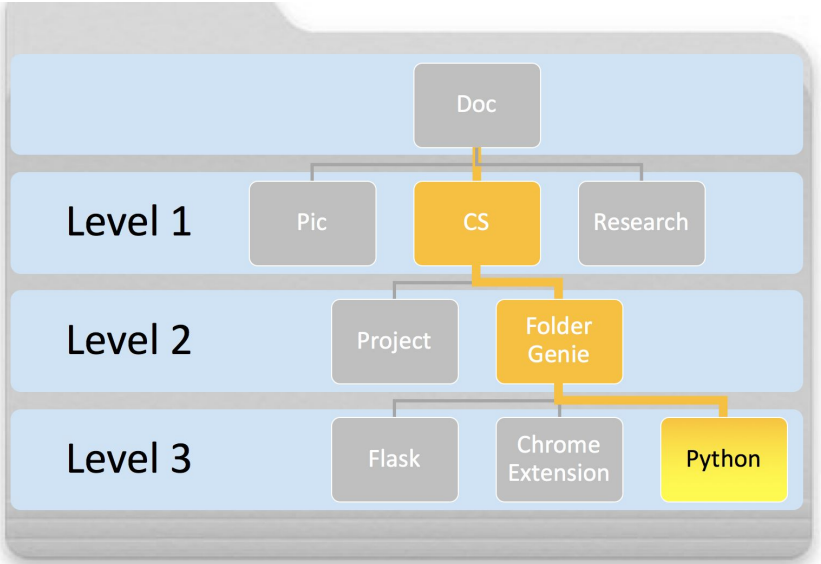
`Doc/CS/FolderGenie/Python`

Hierarchical labels:

Level 1: `CS`

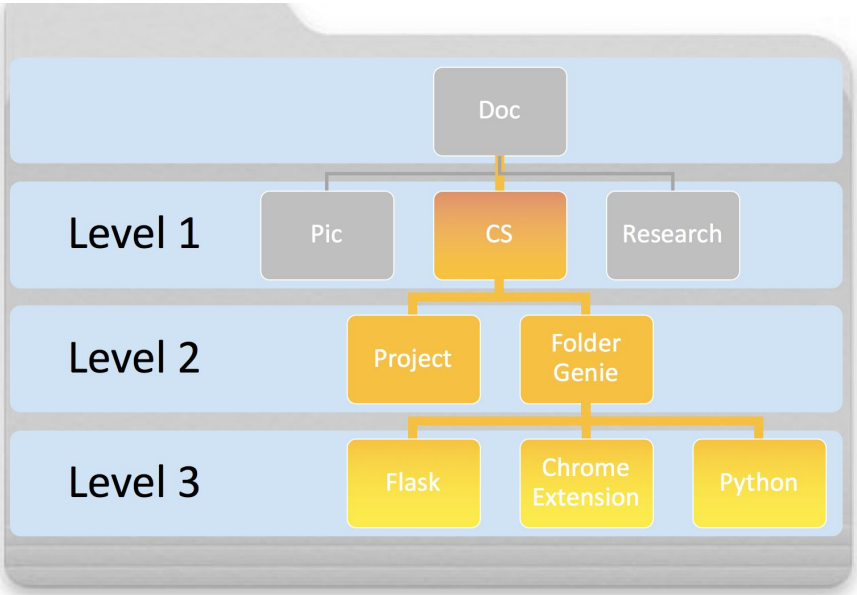
Level 2: `CS/FolderGenie`

Level 3: `CS/FolderGenie/Python`



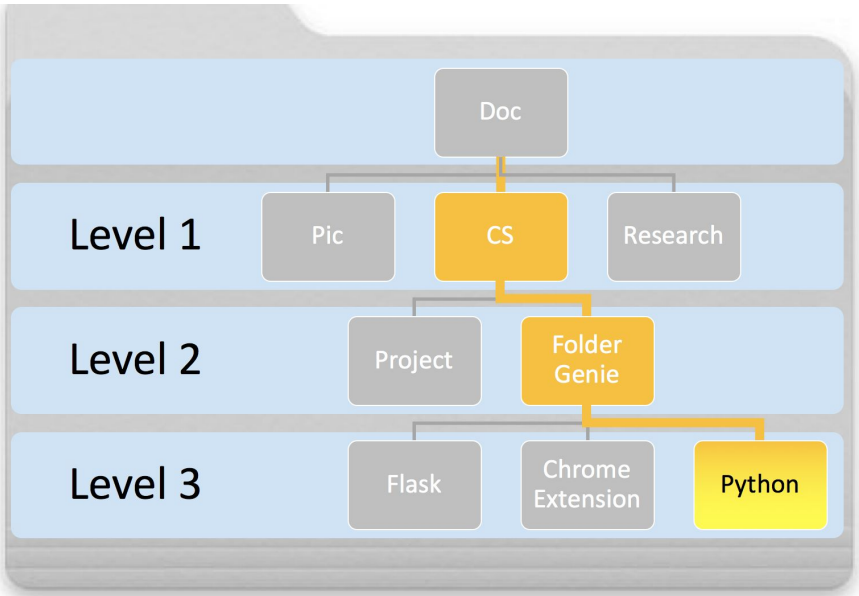
Hierarchical Classification

Multilabel classifier trained at each folder

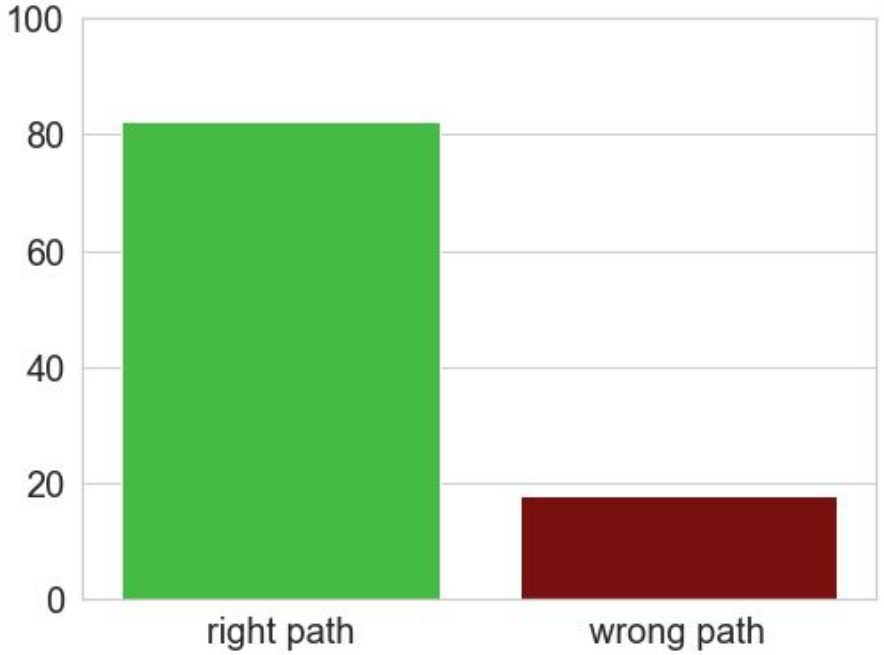


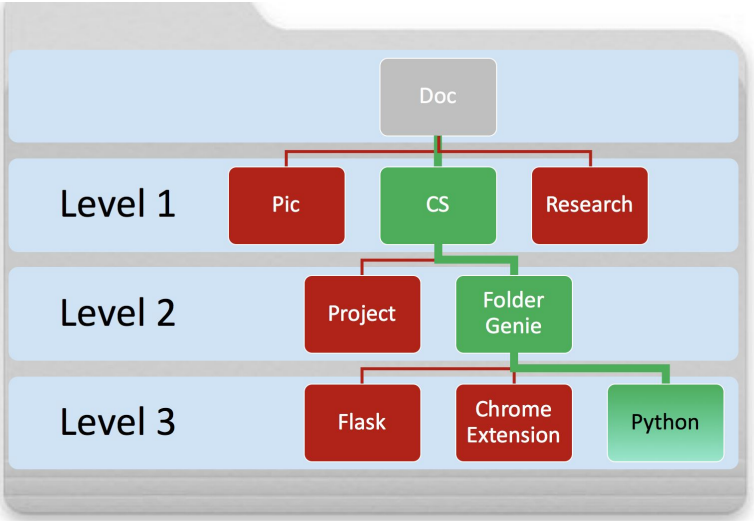
Hierarchical Classification

Predictions only advance deeper when confident

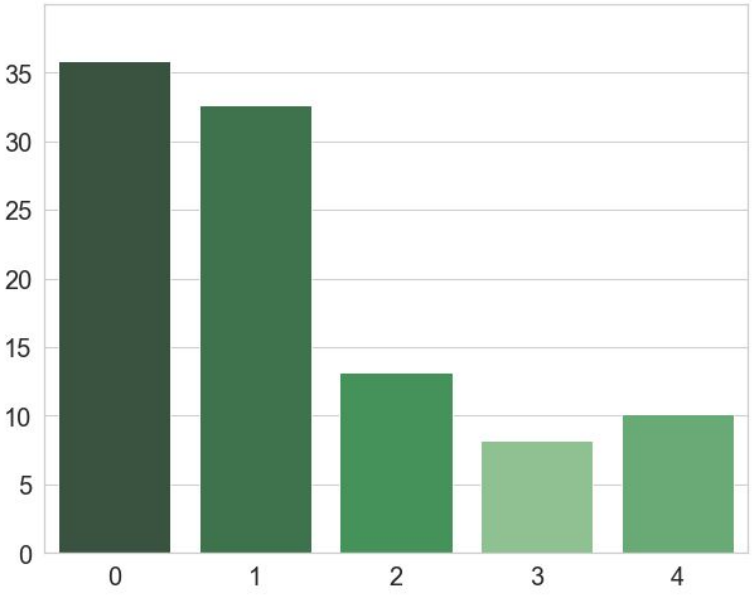


82% go to ***Right Path!***

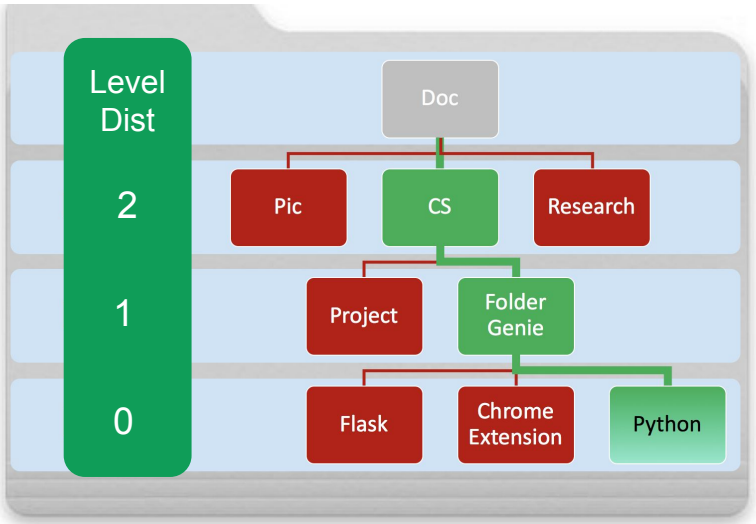




69% go to ***within 1 level distance!***



Level Distance = Actual - Predicted



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

Thank you for attending my talk!

Welcome to try

Available soon

AI Folder Genie

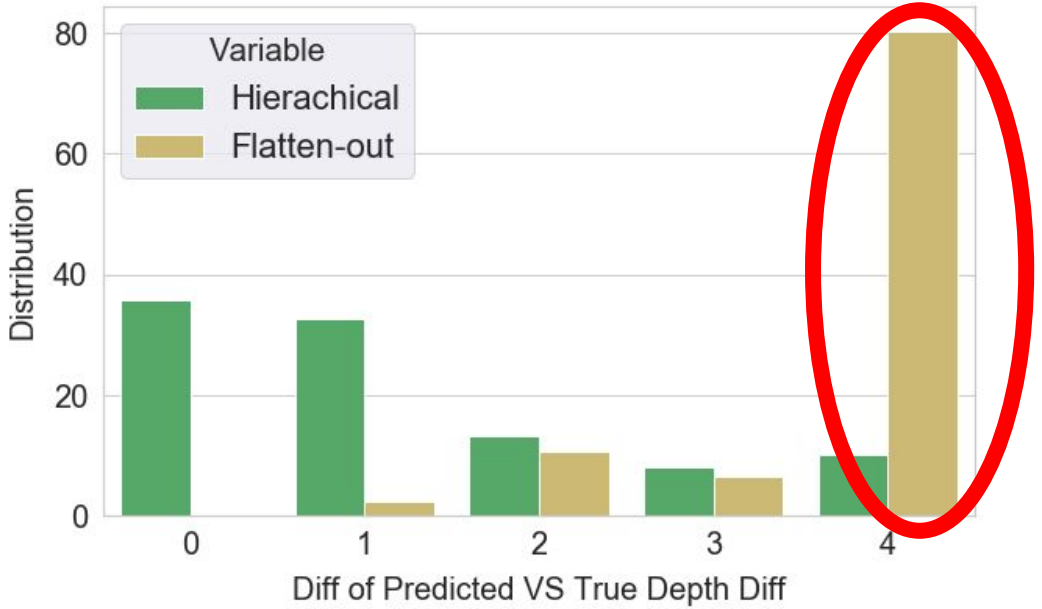
available in the chrome web store

Q&A

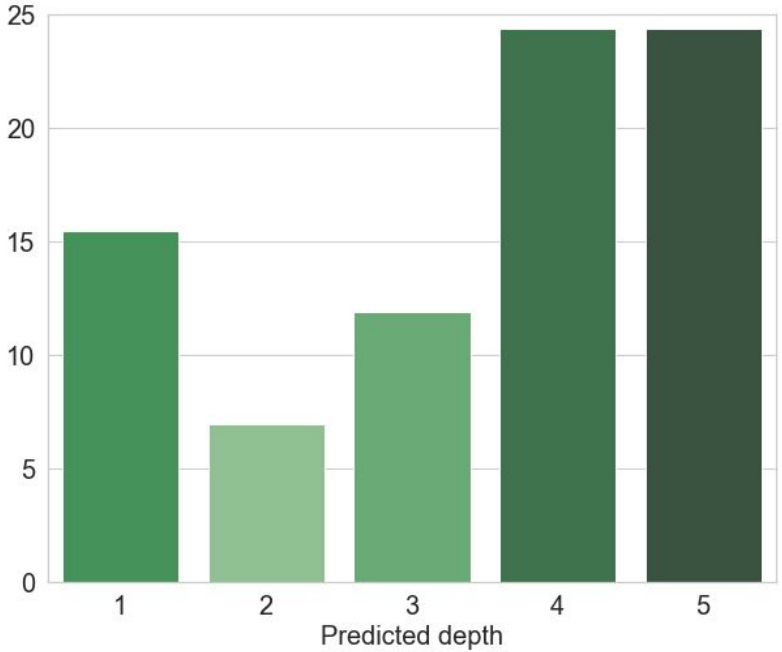
Future Plans

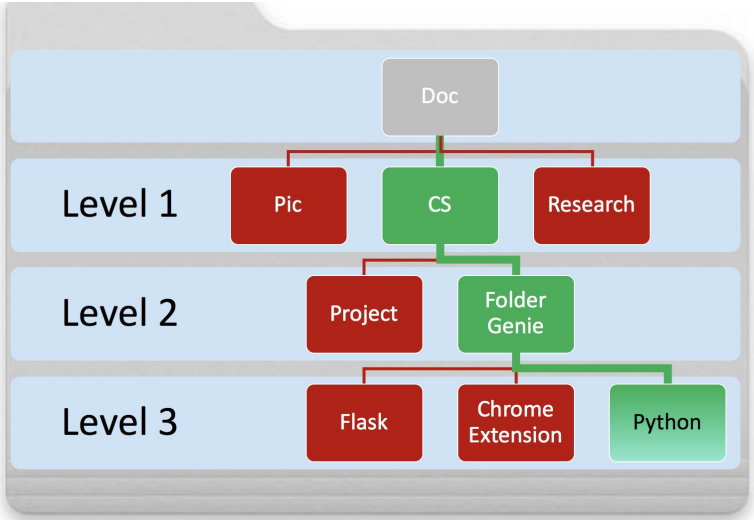
- Features: metadata, NLP folder names, included folder names as part of features
- Chrome Extension + Desktop App: sync between all devices, Dropbox…
- Output a list of choices

Comparison of 2 Alg: Depths Diff



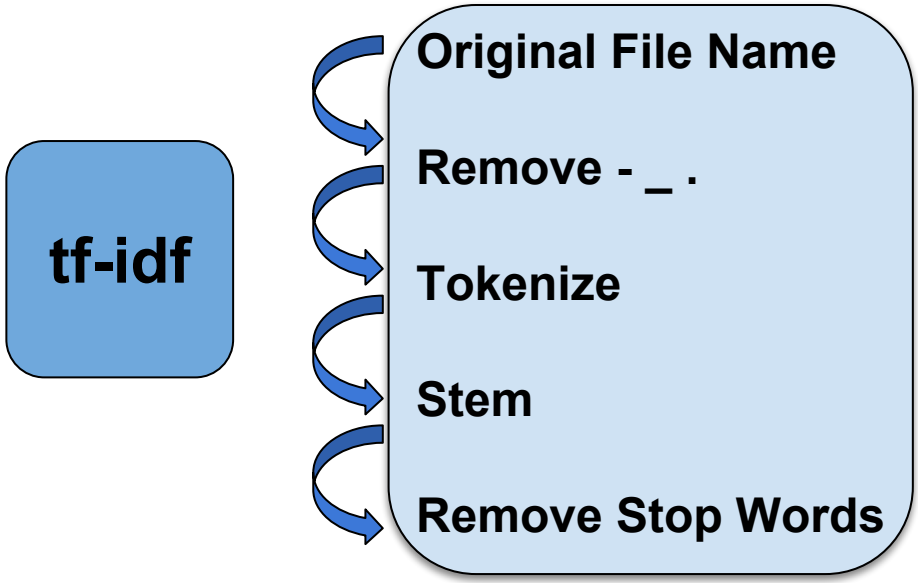
Predicted Depths





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

Features Engineering Steps



- ['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