

Get started

Open in app



Follow

573K Followers



This is your **last** free member-only story this month. [Sign up for Medium and get an extra one](#)

How to Organize your Data Science Articles with Github

Overwhelmed by New Information? Now you can Keep Track and Create Customized Notes for your Articles with Ease



Khuyen Tran · Apr 11, 2020 · 6 min read ★



Retrieved from [The Next Web](#)



like Medium have an efficient recommendation system that does a good job of recommending the articles on which we cannot resist clicking. One article after another appears to be useful. So we keep reading, and either closing the tab or saving the article for future reference. Since we come across many articles every day, we must be informed by new information, right? Not quite so.

Quality is more important than quantity. One home run is much better than two doubles. — Steve Jobs

The best way for us to **absorb new knowledge is to use it**. No matter how many technical articles we read, if we don't apply the new knowledge, we will retain almost nothing. We might retain something now but will forget it a week later. But think about it. In a fast pace and overload society, what is our common strategy for reading an article?



My articles get piled up in my Instapaper folder



phones to read articles, or we want to try out the codes and tools when we need it. When we look back at the piles of articles we have saved, we are overwhelmed and find it time-consuming to go through those articles to find one useful article for our current project. So what is a solution?

The Problems with the Existing Tools

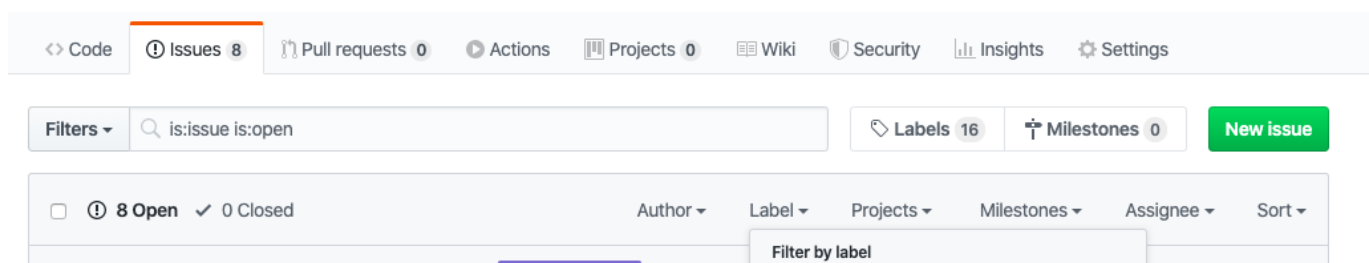
There are many tools out there that allow us to save your articles and organize them in categories. But there are a couple of drawbacks when using these tools to save the articles:

- We simply save the articles because they are interesting but not quite sure how and when to apply them
- We need to reread the articles in one particular category to find the one with the information that we are looking for, while that information can be something small such as snippets of codes
- If we have some questions or opinions on the articles, we have to either go back to the article's website or post the questions online, which is time-consuming, and so on

The good news is that it does not take much for us to organize your articles! With Github and a little twist in the use of the platform, we can organize the articles efficiently.

Welcome to Github and its Issues tab

Github is a light-weight platform that brings together developers to discover, share codes and build better software. Github has an Issues tab that allows users to request an issue on the existing code. We will utilize this tab for another purpose: Organize and summarize our articles. Something like this:





<input type="checkbox"/> Version Control for Data Science — Tracking Machine Learning mo Data Science Tools Version Control #6 opened 16 hours ago by khuyentran1401	<input type="checkbox"/> Build an app to generate photorealistic faces using TensorFlow and Machine Learning Deployment Streamlit TensorFlow #5 opened 16 hours ago by khuyentran1401	<input type="checkbox"/> Scraping Flight Data Using Python Web Scraping #4 opened 16 hours ago by khuyentran1401	<input type="checkbox"/> A Visual and Interactive Guide to the Basics of Neural Networks Data Neural Network #3 opened 17 hours ago by khuyentran1401	<input type="checkbox"/> Jupyter Superpower — Interactive Visualization Combo with Pytho Jupyter Notebook #2 opened 19 hours ago by khuyentran1401	<input type="checkbox"/> Tokenization and Text Data Preparation with TensorFlow & Keras NLP TensorFlow #1 opened 19 hours ago by khuyentran1401
---	---	---	--	--	---

☐ Data Mining

☐ Data Preprocessing

☐ Data Science Tips

☐ Data Science Tools

☐ Data Visualization

☐ Feature Engineering

☐ Jupyter Notebook

☐ Machine Learning Deployment

☐ Neural Network

☐ NLP

Use + click/return to exclude labels.

Explore this repository [here](#). These are why you should start using and contributing to this repository:

- Organize your articles so that it is **easy to find your article for future reference**
- **Quickly review the key points** of the articles without reading the whole articles
- The art of summarizing will help you to **retain and comprehend the information better**
- **Save the time** of going through a low-quality article by looking at the articles others enjoy
- **Ask questions or express your opinions** on the articles in the comment section of Issues

I am a Newbie to Github. Can I still Use it?

Yes of course! This repository is built so that it is easy for anyone to start with. Whenever you find an interesting article that you want to save, click on the Issues button on the top left corner next to Code, then click new Issue on the top right corner.

Get started

Open in app



Author ▾

Label ▾

Projects ▾

Milestones ▾

Assignee ▾

Sort ▾

A template will be shown. All you need to do is to fill in the template. Imagine this like your notebook, you want to quickly save the essential information of one article so it is easy for you to quickly retain the information in the future or for others to look at and know what that article about.



|Title

Related Issues **Beta** [Try it.](#)

Write

Preview

AA B i “ <> 🔗 ☰ ☷ ✓ @ 📌 ↶

TL;DR

Article Link

Author

Key Takeaways

*
*

Useful Code Snippets

```python

#Add your codes here

import numpy as np

Attach files by dragging & dropping, selecting or pasting them.



Styling with Markdown is supported

Submit new issue

Remember, contributions to this repository should follow our [GitHub Community Guidelines](#).

Template


These are quick descriptions of the sections in this template:



• [Link to the article.](#)

- **Author**
- **Key Takeaways** — parts of the articles that you find useful.
- **Useful Code Snippets** — useful codes that you want to save for later
- **Useful Tools** — useful tools that you want to try out or integrate into your workflow
- **Comments/Questions** — Your thoughts or questions on the articles.
- Image (optional) — give you the visual description of what the article about

This is how the template should look like after hitting “Submit new issue”



khuyentran1401 commented 2 minutes ago • edited ▾

Owner 🗨️ ⋮

### TL;DR

This article looks at how to tokenize and prepare text data with TensorFlow and Keras preprocessing tools

### Article Link

<https://www.kdnuggets.com/2020/03/tensorflow-keras-tokenization-text-data-prep.html>

### Author

Matthew Mayo

### Key Takeaways

### Useful Code Snippets

```
Tokenize our training data
tokenizer = Tokenizer(num_words=num_words, oov_token=oov_token)
tokenizer.fit_on_texts(train_data)

Get our training data word index
word_index = tokenizer.word_index

Encode training data sentences into sequences
train_sequences = tokenizer.texts_to_sequences(train_data)

Get max training sequence length
maxlen = max([len(x) for x in train_sequences])

Pad the training sequences
train_padded = pad_sequences(train_sequences, padding=pad_type, truncating=trunc_type,
```

### Useful Tools

Assignees

khuyentran1401

Labels

NLP

TensorFlow

Projects

None yet

Milestone

No milestone

Linked pull requests

Successfully merging a pull request may close this issue.

None yet

Notifications

Customize

Unsubscribe

You're receiving notifications because you're watching this repository.

1 participant

Lock conversation

Pin issue ⓘ

Transfer issue



for yourself to look back! If an article has many pieces of information, some of which you already know and too lazy to write down in the template. Just write down whatever you feel like writing. It is better to write something than give it up all the way.

Last but not least, use the label to tag the category of the article accordingly.



You can also use the comments section as a place to discuss, comment, ask questions or give feedback on the article.





## Conclusion

For every minute spent organizing, an hour is earned.  
— Benjamin Franklin

In the society of overloaded information, we don't need to be one of those who get overwhelmed by new information and give it up all the way. We just need to twist our approach so that we can be more efficient at what we often do. Get started with [this repository](#) to organize and retain more information. I hope with this platform, we can together build a community where everybody is eager to learn and share useful knowledge.

I like to write about basic data science concepts and play with different algorithms and data science tools. You could connect with me on [LinkedIn](#) and [Twitter](#).

Star [this repo](#) if you want to check out the codes for all of the articles I have written. Follow me on Medium to stay informed with my latest data science articles like these:

### How to Share your Jupyter Notebook in 3 Lines of Code with Ngrok

Imagine having your Friends Working with your Local Jupyter Notebook in a Remote Machine

[towardsdatascience.com](https://towardsdatascience.com)

### How to Monitor And Log your Machine Learning Experiment Remotely with HyperDash

The Training Takes a Long Time to Finish, but You Need a Bathroom Break...

[towardsdatascience.com](https://towardsdatascience.com)



Get started

Open in app



I Struggled to Dedicate Time for Data Science. But Finding new Strategies Enables me to Boost my Learning Rate and...

towardsdatascience.com

### Maximize your Productivity with Python

You create a to-do list to be productive but end up wasting your time on non-important tasks. What if you could create...

towardsdatascience.com

---

## Sign up for The Variable

By Towards Data Science

Every Thursday, the Variable delivers the very best of Towards Data Science: from hands-on tutorials and cutting-edge research to original features you don't want to miss. [Take a look.](#)

Get this newsletter

Data Science

Organize

Programming

Github

Reading

About Write Help Legal

Get the Medium app

Download on the

GET IT ON

---

Get started

Open in app

