GSOC TIMELINE

Author: Recep
Date: 17.06.2019
Mentors: Pierre Ruyssen,
Marcin Michalski,
Etienne Pot

27 May - 17 June

Note: 2-16 June Final exam week

- Researching
  - Jupyter Notebooks Extensions
  - Gsutil
- Issue
- Docs
  - KaggleBuilder
- Pull Requests
  - Add kaggle downloader extension to jupyter notebook #447
  - Add gsutil support to downloader. #627
- Reviews
  -

17- 23 June

- Researching
  - TFRecords,
  - Data validation

- Issue
  - Better Sequence display in documentation #689
  - Bug in document_datasets.py.

- Docs
  - KaggleBuilder
- Pull Requests
  - Add the missing links in index. #698
  - Add a comment warning at the top of the generated .md files #697

24-30 June

- Researching
  - FACETS,
  - Data Visualization
- Issue
  - Allow TFDS datasets defined outside TFDS #704 (assigned)
  - Dataset Versioning #721
- Docs
  - Data Visualization
  - Easier Dataset Implementation (Created at 7 June)
- Pull Requests
  - Expand Sequence list #692
  - FACETS Visualization #714
  - Delete unfound reference `from __future__ import google_type_annotations' #718
- Reviews
  - coco2017 and coco2017_panoptic #716
  - Read directly from archives on datasets with many records. #701
  - mnist dataset falls back to its original website #713

1-7 July

- Researching
  - Auto-generation of fake data
- Issue
  - Dataset Versioning #721
- Docs
  - GSOC TIMELINE
- Pull Requests
  - Broken link on add_dataset.md #730

8-14 July

- Researching
    - Auto-generation of fake data
- Issue
    - Dataset Versioning #721
    - day2night #752
    - Direct onedrive url (for Resisc45) #631
- Docs
    -
- Pull Requests
    -
- Reviews
    - Implemented Mozilla Common Voice Dataset (Multilingual #161
    - Added Stanford Online Products Dataset #238
    - FACETS Visualization #714
    - Expand Sequence list #692
    - Improving documentation in text_feature.py #750
    - Add PatchCAMELYON dataset #258
    - mnist dataset falls back to its original website #713
    - coco2017 and coco2017_panoptic #716

15-21 July

- Researching

    - TFRecords
- Issue
    - download mnist using python 3.7.4 #769
    - Allow TFDS datasets defined outside TFDS #704
    - Loading "super_glue/copa" fails #797
- Docs
    - Automate the Fake Data Generator and Testing
- Pull Requests
    - Auto fake data generator #785
    - Fix checksum input #786
    - Add dataset-name to download_dir. #795
    - Optimizing Imports with Pycharm #796
- Reviews
    - Small fix for assert statement in Flores BuilderConfig #763
    - Adding examples to DatasetBuilder.as_dataset() #784
    - Use TFDV to compute statistics in TFDS #782
    - mnist dataset falls back to its original website #713
    - Added a new Dataset in translate directory #793
    - Add Quickdraw Sketch RNN Dataset #361
    - Added Dep of Tensorflow_io.lmdb to LSUN #468
    - Add `show_examples(ds)` option for image datasets. #670

22-28 June

- Researching
    -
- Issue
    - Cannot download all datasets except mnist using command tfds.load(). #807
    - Can't download IMDB dataset from Microsoft Windows #800
    - [data request] binary-mnist #758
    -

- Docs
    - Offline Mode of TFDS
- Pull Requests
    - Change `tf.io.gfile` instead of os functions #804
- Reviews
    - Include __version__ in __all__ #798
    - Add `show_examples(ds)` option for image datasets. #670
    - add mini_imagenet dataset #263
    - Add binarized_mnist to TFDS. #809
    - add dataset: yelp_polarity_review #582
    - Would like to contribute a new Dataset to translate directory #814

29 July - 4 Aug

- Researching
    -
- Issue
    - Unit testing does not work on Windows #817
    - Allow TFDS datasets defined outside TFDS #704
- Docs
    -
- Pull Requests
    - Add tfds.add_checksums_dir #834
- Reviews

    - Add AFLW2000-3D Dataset #359
    - Added Mura Dataset #397
    - Added dataset for Cars196 --Issue 202 #294
    - Fixes #207: Added Support for Visual Dialog Dataset. #307
    - Add Quickdraw Sketch RNN Dataset #361
    - Added Deepweeds dataset #370
    - Add initial unit test and test data #391
    - Add lfw #379
    - MIT Scene Parse 150 dataset #398
    - Add `show_examples(ds)` option for image datasets. #670
    - Added support for Cartoon Set #436

5 Aug - 11 Aug

- Researching
    - Threading
    - Requests: HTTP for Humans™¶ (python lib)
- Doc
    - Better Download Experience
- Issue
    - Kaggle silently compresses some archives #844
    -
- Pull Requests
    - Launch S3 on all text datasets. #864
    - Update translate dataset to s3 #870

    - Add usage of checksum_dir and EXAMPLE_DIR doc. #873
    - Fix DatasetBuilderTestCase fake example dir path #879
    - Fix broken link on README.md. #888
- Reviews
    - Add Fruits360 #835
    - add cifar-10.1 to tfds #839
    - Add english-tamil parallel corpus to translate #818
    - Fixes #317: Added Support for Stanford Question Answering Dataset - 2.0 #323

12 Aug - 18 Aug

- Researching
    - Urllib3 (python lib)
    -
- Issue
    -
- Pull Requests
    - Add versions to list builders. #897
    - Add translate/wmt S3 Version #923
    - Better Download Experience #921
- Reviews
    - Added cityscapes dataset #225

- add the300w_lp dataset #890
- Fix the link to the list of available datasets in the documentation #918
- Launch S3 to structured datasets #874
- Add usage of checksum_dir and EXAMPLE_DIR doc. #873