

# Repo cleanup notes

The BertRegressor model is now called "general\_regressor"

Doesn't seem that transition away from word splitters has occurred yet?

Missing allennlp blocks?

## March 27th

Currently, I don't have a way to install futil from the package Brian gave me. I think he mentioned that it's missing entry points which maybe means that it lacks the `setup.py` file.

# Switching over to DataLoader for the Homogeneous batcher

There are three things the homogeneous batch loader was doing:

- All data has to be added to a batch, the last batch cannot be dropped
- Batch size needs to be configurable
- Each batch only contains one dataset type
- The key in the instances that determines the dataset type is configurable

## I think a new sampler is required

I wrote a new batch sampler that generates homogenous batches using the dataset type parameter specified in the instances.

This can be used with the off the shelf data loader and does not require creating a new data loader type.

# What's the deal with the new cache directory setup?

It looks like `_setup_cache_files` is deprecated in `allennlp.training`. Not sure what mechanism handles cache stuff now.

According to the prerelease notes:

“Dataset caching is now handled entirely with a parameter passed to the dataset reader, not with command-line arguments. If you used the caching arguments to `allennlp train`, instead just add a `"cache_directory"` key to your dataset reader parameters.

---

Revision #7

Created Tue, Mar 17, 2020 5:16 PM by [kenneth](#)

Updated Wed, Apr 8, 2020 2:34 PM by [kenneth](#)