larger websites:
- can take too long for many HTML requests
- may have detectors for crawlers
Twitter blocking: 1800 calls/minute, 15 min.
Reddit limits: 600 calls/minute is OK. (check Reddit Dev)
OK; 30/1 is the official
1 sec/call is bare level of politeness (small sites may not have infrastructure to check limits).
multiple calls to search API mean duplicated data (Tinder doesn't let you go back in time),
user view meant you can order things to avoid the duplicates.

tweepy pkg - very usable, although aimed @ application & developers - so take things
out of python classes.

* KAIST "LL" Twitter social graph @ the time

Reddit's usual API returns just most recent
But "cloned" takes timestamps (a backdoor!) Yes, use it!
Amit has exceptions mailed to him, (just using unix)
ex. We many servers.
Unexepected error codes can happen.

Keeping logs!

Amit ended up converting: C-Pickle> flatfile> SQL

Social-integrator

what can handle back-up data that can handle
(Just find what you are looking for).
Some uses: Beautiful Soup

URL-generation - finding the URLs? (URL target is sped to be considered private)

grep/awk, shell scripting
- Can post quickly & manually. Use API, not on the commandline.
- Less vulnerable to re-use.

also pandas, resp for random data.

just try to get everything to fit in memory.