

Fixing spelling mistakes in Debian packages with 578 github pull requests

Edward Betts <edward@4angle.com>

codespell

- tool for finding spelling mistakes in source code
- dictionary of 42,145 spelling errors

annoints → anoints
banannas → bananas
daed → dead
engagment → engagement
extered → exerted
interrput → interrupt
misdemenors → misdemeanors
ocassions → occasions
perheaps → perhaps
referece → reference

asthetically → aesthetically
carreer → career
demorcracy → democracy
enviornment → environment
frequentily → frequently
manifestations → manifestations
neolithic → neolithic
oppositition → opposition
pricipal → principal
relaly → really

attributred → attributed
compensantion → compensation
eles → else
excecuting → executing
holf → hold
miscelaneous → miscellaneous
obsessed → obsessed
payed → paid
publisehr → publisher
unsurprized → unsurprised

UDD: Ultimate Debian Database

- use UDD PostgreSQL database to find GitHub upstream
- multiple places to check for GitHub URLs in UDD
- homepage field, watch file, upstream metadata

Found some errors

- 6,801 GitHub URLs found, of those 6,426 were usable
- submitted 14 Debian bugs for moved repos or incorrect URLs
- 3,441 have spelling mistakes

```
./changelog:108: controlling  => controlling
./changelog:461: criterias  => criteria
./changelog:462: criterias  => criteria
./changelog:464: criterias  => criteria
./changelog:473: developement => development
./changelog:585: misspelled  => misspelled
./backupchecker/configurations.py:49: availabe  => available
./backupchecker/generatelist/generatelistforzip.py:220: dependant  => dependent
./backupchecker/checkbackups/checkzip.py:88: dependant  => dependent
```

Submit pull requests

- skip .po files, and vendor directories
- only submit pull requests for packages with a single mistake
- didn't bother with a role account, used my regular github account
- python modules: PyGithub and GitPython
- result was 578 GitHub pull requests

Upstream responses

- most upstream very happy
- some upstream complain that contributor rules not followed