

# Language\_detection\_using\_Statistical\_Language\_Models

July 8, 2020

*Student name* - Name : **Bhavya-deepthi.kanaparthi**

!chmod +x /content/SRILM\_installation.bash !./SRILM\_installation.bash

[7]:

```
!chmod +x /content/fastText_installation.bash
!./fastText_installation.bash
```

mkdir: created directory '//root/Tools/fastText/'

//root/Tools/fastText /content

--2020-07-08 10:36:24--

https://github.com/cservan/website/raw/master/tools/fastText.tar.gz Resolving

github.com (github.com)... 140.82.112.4

Connecting to github.com (github.com)|140.82.112.4|:443...connected. HTTP request sent, awaiting response... 302 Found

Location: https://raw.githubusercontent.com/cservan/website/master/tools/fastText.tar.gz [following]

--2020-07-08 10:36:24--

https://raw.githubusercontent.com/cservan/website/master/tools/fastText.tar.gz Resolving

raw.githubusercontent.com (raw.githubusercontent.com)... 151.101.0.133, 151.101.64.133, 151.101.128.133, ...

Connecting to raw.githubusercontent.com

(raw.githubusercontent.com)|151.101.0.133|:443...connected. HTTP

request sent, awaiting response... 200 OK

Length: 4022585 (3.8M)[application/octet-stream] Saving to:

'fastText.tar.gz'

fastText.tar.gz 100%=====>] 3.84M 12.6MB/s in 0.3s

2020-07-08 10:36:25 (12.6 MB/s) - 'fastText.tar.gz' saved [4022585/4022585]

fastText/ fastText/.circleci/

fastText/.circleci/cmake\_test.sh fastText/.circleci/config.yml

fastText/.circleci/gcc\_test.sh fastText/.circleci/pip\_test.sh

fastText/.circleci/pull\_data.sh fastText/.circleci/python\_test.sh

fastText/.circleci/run\_locally.sh  
fastText/.circleci/setup\_circleimg.sh  
fastText/.circleci/setup\_debian.sh  
fastText/.gitignore fastText/alignment/  
fastText/alignment/align.py  
fastText/alignment/eval.py  
fastText/alignment/example.sh  
fastText/alignment/README.md  
fastText/alignment/unsup\_align.py  
fastText/alignment/unsup\_multialign.py  
fastText/alignment/utils.py  
fastText/classification-example.sh  
fastText/classification-results.sh  
fastText/CMakeLists.txt  
fastText/CODE\_OF\_CONDUCT.md  
fastText/CONTRIBUTING.md fastText/crawl/  
fastText/crawl/dedup.cc  
fastText/crawl/download\_crawl.sh  
fastText/crawl/filter\_dedup.sh  
fastText/crawl/filter\_utf8.cc  
fastText/crawl/process\_wet\_file.sh  
fastText/crawl/README.md fastText/docs/  
fastText/docs/aligned-vectors.md  
fastText/docs/api.md fastText/docs/autotune.md  
fastText/docs/cheatsheet.md fastText/docs/crawl-  
vectors.md fastText/docs/dataset.md  
fastText/docs/english-vectors.md  
fastText/docs/faqs.md fastText/docs/language-  
identification.md fastText/docs/options.md  
fastText/docs/pretrained-vectors.md  
fastText/docs/python-module.md  
fastText/docs/references.md  
fastText/docs/supervised-models.md  
fastText/docs/supervised-tutorial.md  
fastText/docs/support.md  
fastText/docs/unsupervised-tutorials.md  
fastText/download\_model.py fastText/eval.py  
fastText/get-wikimedia.sh  
fastText/LICENSE  
fastText/Makefile  
fastText/MANIFEST.in

fastText/python/  
fastText/python/benchmarks/  
fastText/python/benchmarks/get\_word\_vector.py  
fastText/python/benchmarks/README.rst  
fastText/python/doc/ fastText/python/doc/examples/  
fastText/python/doc/examples/bin\_to\_vec.py  
fastText/python/doc/examples/compute\_accuracy.py  
fastText/python/doc/examples/FastTextEmbeddingBag.py  
fastText/python/doc/examples/get\_vocab.py  
fastText/python/doc/examples/train\_supervised.py  
fastText/python/doc/examples/train\_unsupervised.py  
fastText/python/fasttext\_module/ fastText/python/fasttext\_module/fasttext/  
fastText/python/fasttext\_module/fasttext/FastText.py  
fastText/python/fasttext\_module/fasttext/pybind/  
fastText/python/fasttext\_module/fasttext/pybind/fasttext\_pybind.cc  
fastText/python/fasttext\_module/fasttext/tests/  
fastText/python/fasttext\_module/fasttext/tests/test\_configurations.py  
fastText/python/fasttext\_module/fasttext/tests/test\_script.py  
fastText/python/fasttext\_module/fasttext/tests/\_init\_.py  
fastText/python/fasttext\_module/fasttext/util/  
fastText/python/fasttext\_module/fasttext/util/util.py  
fastText/python/fasttext\_module/fasttext/util/\_init\_.py  
fastText/python/fasttext\_module/fasttext/\_init\_.py fastText/python/README.md  
fastText/python/README.rst  
fastText/quantization-example.sh  
fastText/README.md  
fastText/reduce\_model.py fastText/runtests.py  
fastText/scripts/  
fastText/scripts/kbcompletion/  
fastText/scripts/kbcompletion/data.sh  
fastText/scripts/kbcompletion/eval.cpp  
fastText/scripts/kbcompletion/fb15k.sh  
fastText/scripts/kbcompletion/fb15k237.sh  
fastText/scripts/kbcompletion/README.md  
fastText/scripts/kbcompletion/svo.sh  
fastText/scripts/kbcompletion/wn18.sh  
fastText/scripts/quantization/  
fastText/scripts/quantization/quantization-results.sh  
fastText/setup.cfg  
fastText/setup.py fastText/src/  
fastText/src/args.cc  
fastText/src/args.h  
fastText/src/autotune.cc

fastText/src/autotune.h  
fastText/src/densematrix.cc  
fastText/src/densematrix.h  
fastText/src/dictionary.cc  
fastText/src/dictionary.h  
fastText/src/fasttext.cc  
fastText/src/fasttext.h fastText/src/loss.cc  
fastText/src/loss.h fastText/src/main.cc  
fastText/src/matrix.cc  
fastText/src/matrix.h fastText/src/meter.cc  
fastText/src/meter.h fastText/src/model.cc  
fastText/src/model.h  
fastText/src/productquantizer.cc  
fastText/src/productquantizer.h  
fastText/src/quantmatrix.cc  
fastText/src/quantmatrix.h  
fastText/src/real.h fastText/src/utils.cc  
fastText/src/utils.h fastText/src/vector.cc  
fastText/src/vector.h fastText/tests/  
fastText/tests/fetch\_test\_data.sh  
fastText/website/ fastText/website/blog/  
fastText/website/blog/2016-08-18-blog-post.md  
fastText/website/blog/2017-05-02-blog-post.md  
fastText/website/blog/2017-10-02-blog-post.md  
fastText/website/blog/2019-06-25-blog-post.md  
fastText/website/core/ fastText/website/core/Footer.js  
fastText/website/package.json fastText/website/pages/  
fastText/website/pages/en/  
fastText/website/pages/en/index.js  
fastText/website/README.md  
fastText/website/sidebars.json  
fastText/website/siteConfig.js fastText/website/static/  
fastText/website/static/docs/  
fastText/website/static/docs/en/  
fastText/website/static/docs/en/html/  
fastText/website/static/docs/en/html/.classfasttext\_1\_1QMatrix- members.html.i4eKqy

fastText/website/static/docs/en/html/annotated.html  
fastText/website/static/docs/en/html/annotated\_dup.js  
fastText/website/static/docs/en/html/args\_8cc.html  
fastText/website/static/docs/en/html/args\_8h.html  
fastText/website/static/docs/en/html/args\_8h.js  
fastText/website/static/docs/en/html/args\_8h\_source.html  
fastText/website/static/docs/en/html/bc\_s.png fastText/website/static/docs/en/html/bdwn.png  
fastText/website/static/docs/en/html/classes.html  
fastText/website/static/docs/en/html/classfasttext\_1\_1Args-members.html  
fastText/website/static/docs/en/html/classfasttext\_1\_1Args.html  
fastText/website/static/docs/en/html/classfasttext\_1\_1Args.js  
fastText/website/static/docs/en/html/classfasttext\_1\_1Dictionary-members.html  
fastText/website/static/docs/en/html/classfasttext\_1\_1Dictionary.html  
fastText/website/static/docs/en/html/classfasttext\_1\_1Dictionary.js  
fastText/website/static/docs/en/html/classfasttext\_1\_1FastText-members.html  
fastText/website/static/docs/en/html/classfasttext\_1\_1FastText.html  
fastText/website/static/docs/en/html/classfasttext\_1\_1FastText.js  
fastText/website/static/docs/en/html/classfasttext\_1\_1Matrix-members.html  
fastText/website/static/docs/en/html/classfasttext\_1\_1Matrix.html  
fastText/website/static/docs/en/html/classfasttext\_1\_1Matrix.js  
fastText/website/static/docs/en/html/classfasttext\_1\_1Model-members.html  
fastText/website/static/docs/en/html/classfasttext\_1\_1Model.html  
fastText/website/static/docs/en/html/classfasttext\_1\_1Model.js  
fastText/website/static/docs/en/html/classfasttext\_1\_1ProductQuantizer-members.html  
fastText/website/static/docs/en/html/classfasttext\_1\_1ProductQuantizer.html  
fastText/website/static/docs/en/html/classfasttext\_1\_1ProductQuantizer.js  
fastText/website/static/docs/en/html/classfasttext\_1\_1QMatrix-members.html  
fastText/website/static/docs/en/html/classfasttext\_1\_1QMatrix.html  
fastText/website/static/docs/en/html/classfasttext\_1\_1QMatrix.js  
fastText/website/static/docs/en/html/classfasttext\_1\_1Vector-members.html  
fastText/website/static/docs/en/html/classfasttext\_1\_1Vector.html  
fastText/website/static/docs/en/html/classfasttext\_1\_1Vector.js  
fastText/website/static/docs/en/html/closed.png  
fastText/website/static/docs/en/html/dictionary\_8cc.html  
fastText/website/static/docs/en/html/dictionary\_8h.html  
fastText/website/static/docs/en/html/dictionary\_8h.js  
fastText/website/static/docs/en/html/dictionary\_8h\_source.html  
fastText/website/static/docs/en/html/dir\_68267d1309a1af8e8297ef4c3efbcdba.html  
fastText/website/static/docs/en/html/dir\_68267d1309a1af8e8297ef4c3efbcdba.js  
fastText/website/static/docs/en/html/doc.png fastText/website/static/docs/en/html/doxygen.css  
fastText/website/static/docs/en/html/doxygen.png  
fastText/website/static/docs/en/html/dynsections.js  
fastText/website/static/docs/en/html/fasttext\_8cc.html  
fastText/website/static/docs/en/html/fasttext\_8h.html  
fastText/website/static/docs/en/html/fasttext\_8h.js

fastText/website/static/docs/en/html/fasttext\_8h\_source.html  
fastText/website/static/docs/en/html/favicon.png fastText/website/static/docs/en/html/files.html  
fastText/website/static/docs/en/html/files.js fastText/website/static/docs/en/html/folderclosed.png  
fastText/website/static/docs/en/html/folderopen.png  
fastText/website/static/docs/en/html/functions.html  
fastText/website/static/docs/en/html/functions\_0x7e.html  
fastText/website/static/docs/en/html/functions\_b.html  
fastText/website/static/docs/en/html/functions\_c.html  
fastText/website/static/docs/en/html/functions\_d.html  
fastText/website/static/docs/en/html/functions\_dup.js  
fastText/website/static/docs/en/html/functions\_e.html  
fastText/website/static/docs/en/html/functions\_f.html  
fastText/website/static/docs/en/html/functions\_func.html  
fastText/website/static/docs/en/html/functions\_g.html  
fastText/website/static/docs/en/html/functions\_h.html  
fastText/website/static/docs/en/html/functions\_i.html  
fastText/website/static/docs/en/html/functions\_k.html  
fastText/website/static/docs/en/html/functions\_l.html  
fastText/website/static/docs/en/html/functions\_m.html  
fastText/website/static/docs/en/html/functions\_n.html  
fastText/website/static/docs/en/html/functions\_o.html  
fastText/website/static/docs/en/html/functions\_p.html  
fastText/website/static/docs/en/html/functions\_q.html  
fastText/website/static/docs/en/html/functions\_r.html  
fastText/website/static/docs/en/html/functions\_s.html  
fastText/website/static/docs/en/html/functions\_t.html  
fastText/website/static/docs/en/html/functions\_u.html  
fastText/website/static/docs/en/html/functions\_v.html  
fastText/website/static/docs/en/html/functions\_vars.html  
fastText/website/static/docs/en/html/functions\_w.html  
fastText/website/static/docs/en/html/functions\_z.html  
fastText/website/static/docs/en/html/globals.html  
fastText/website/static/docs/en/html/globals\_defs.html  
fastText/website/static/docs/en/html/globals\_func.html  
fastText/website/static/docs/en/html/index.html fastText/website/static/docs/en/html/jquery.js  
fastText/website/static/docs/en/html/main\_8cc.html fastText/website/static/docs/en/html/main\_8cc.js  
fastText/website/static/docs/en/html/matrix\_8cc.html  
fastText/website/static/docs/en/html/matrix\_8h.html  
fastText/website/static/docs/en/html/matrix\_8h\_source.html  
fastText/website/static/docs/en/html/menu.js fastText/website/static/docs/en/html/menudata.js  
fastText/website/static/docs/en/html/model\_8cc.html  
fastText/website/static/docs/en/html/model\_8h.html fastText/website/static/docs/en/html/model\_8h.js

fastText/website/static/docs/en/html/model\_8h\_source.html  
fastText/website/static/docs/en/html/namespacefasttext.html  
fastText/website/static/docs/en/html/namespacefasttext.js  
fastText/website/static/docs/en/html/namespacefasttext\_1\_1utils.html  
fastText/website/static/docs/en/html/namespacemembers.html  
fastText/website/static/docs/en/html/namespacemembers\_enum.html  
fastText/website/static/docs/en/html/namespacemembers\_func.html  
fastText/website/static/docs/en/html/namespacemembers\_type.html  
fastText/website/static/docs/en/html/namespaces.html  
fastText/website/static/docs/en/html/namespaces.js fastText/website/static/docs/en/html/navtree.css  
fastText/website/static/docs/en/html/navtree.js fastText/website/static/docs/en/html/navtreedata.js  
fastText/website/static/docs/en/html/navtreeindex0.js  
fastText/website/static/docs/en/html/navtreeindex1.js fastText/website/static/docs/en/html/nav\_f.png  
fastText/website/static/docs/en/html/nav\_g.png fastText/website/static/docs/en/html/nav\_h.png  
fastText/website/static/docs/en/html/open.png  
fastText/website/static/docs/en/html/productquantizer\_8cc.html  
fastText/website/static/docs/en/html/productquantizer\_8cc.js  
fastText/website/static/docs/en/html/productquantizer\_8h.html  
fastText/website/static/docs/en/html/productquantizer\_8h\_source.html  
fastText/website/static/docs/en/html/qmatrix\_8cc.html  
fastText/website/static/docs/en/html/qmatrix\_8h.html  
fastText/website/static/docs/en/html/qmatrix\_8h\_source.html  
fastText/website/static/docs/en/html/real\_8h.html fastText/website/static/docs/en/html/real\_8h.js  
fastText/website/static/docs/en/html/real\_8h\_source.html  
fastText/website/static/docs/en/html/resize.js fastText/website/static/docs/en/html/search/  
fastText/website/static/docs/en/html/search/.files\_7.html.StRRNc  
fastText/website/static/docs/en/html/search/.variables\_a.html.1MGQ27  
fastText/website/static/docs/en/html/search/all\_0.html  
fastText/website/static/docs/en/html/search/all\_0.js  
fastText/website/static/docs/en/html/search/all\_1.html  
fastText/website/static/docs/en/html/search/all\_1.js  
fastText/website/static/docs/en/html/search/all\_10.html  
fastText/website/static/docs/en/html/search/all\_10.js  
fastText/website/static/docs/en/html/search/all\_11.html  
fastText/website/static/docs/en/html/search/all\_11.js  
fastText/website/static/docs/en/html/search/all\_12.html  
fastText/website/static/docs/en/html/search/all\_12.js  
fastText/website/static/docs/en/html/search/all\_13.html  
fastText/website/static/docs/en/html/search/all\_13.js  
fastText/website/static/docs/en/html/search/all\_14.html  
fastText/website/static/docs/en/html/search/all\_14.js  
fastText/website/static/docs/en/html/search/all\_15.html

[illegible]



[illegible]

[illegible]

fastText/website/static/docs/en/html/search/mag\_sel.png  
fastText/website/static/docs/en/html/search/namespaces\_0.html  
fastText/website/static/docs/en/html/search/namespaces\_0.js  
fastText/website/static/docs/en/html/search/nomatches.html  
fastText/website/static/docs/en/html/search/search.css  
fastText/website/static/docs/en/html/search/search.js  
fastText/website/static/docs/en/html/search/searchdata.js  
fastText/website/static/docs/en/html/search/search\_l.png  
fastText/website/static/docs/en/html/search/search\_m.png  
fastText/website/static/docs/en/html/search/search\_r.png  
fastText/website/static/docs/en/html/search/typedefs\_0.html  
fastText/website/static/docs/en/html/search/typedefs\_0.js  
fastText/website/static/docs/en/html/search/typedefs\_1.html  
fastText/website/static/docs/en/html/search/typedefs\_1.js  
fastText/website/static/docs/en/html/search/variables\_0.html  
fastText/website/static/docs/en/html/search/variables\_0.js  
fastText/website/static/docs/en/html/search/variables\_1.html  
fastText/website/static/docs/en/html/search/variables\_1.js  
fastText/website/static/docs/en/html/search/variables\_10.html  
fastText/website/static/docs/en/html/search/variables\_10.js  
fastText/website/static/docs/en/html/search/variables\_11.html  
fastText/website/static/docs/en/html/search/variables\_11.js  
fastText/website/static/docs/en/html/search/variables\_12.html  
fastText/website/static/docs/en/html/search/variables\_12.js  
fastText/website/static/docs/en/html/search/variables\_13.html  
fastText/website/static/docs/en/html/search/variables\_13.js  
fastText/website/static/docs/en/html/search/variables\_2.html  
fastText/website/static/docs/en/html/search/variables\_2.js  
fastText/website/static/docs/en/html/search/variables\_3.html  
fastText/website/static/docs/en/html/search/variables\_3.js  
fastText/website/static/docs/en/html/search/variables\_4.html  
fastText/website/static/docs/en/html/search/variables\_4.js  
fastText/website/static/docs/en/html/search/variables\_5.html  
fastText/website/static/docs/en/html/search/variables\_5.js  
fastText/website/static/docs/en/html/search/variables\_6.html  
fastText/website/static/docs/en/html/search/variables\_6.js  
fastText/website/static/docs/en/html/search/variables\_7.html  
fastText/website/static/docs/en/html/search/variables\_7.js  
fastText/website/static/docs/en/html/search/variables\_8.html  
fastText/website/static/docs/en/html/search/variables\_8.js  
fastText/website/static/docs/en/html/search/variables\_9.html  
fastText/website/static/docs/en/html/search/variables\_9.js  
fastText/website/static/docs/en/html/search/variables\_a.html  
fastText/website/static/docs/en/html/search/variables\_a.js  
fastText/website/static/docs/en/html/search/variables\_b.html  
fastText/website/static/docs/en/html/search/variables\_b.js  
fastText/website/static/docs/en/html/search/variables\_c.html  
fastText/website/static/docs/en/html/search/variables\_c.js

fastText/website/static/docs/en/html/search/variables\_d.html  
fastText/website/static/docs/en/html/search/variables\_d.js  
fastText/website/static/docs/en/html/search/variables\_e.html  
fastText/website/static/docs/en/html/search/variables\_e.js  
fastText/website/static/docs/en/html/search/variables\_f.html  
fastText/website/static/docs/en/html/search/variables\_f.js  
fastText/website/static/docs/en/html/splitbar.png  
fastText/website/static/docs/en/html/structfasttext\_1\_1entry-members.html  
fastText/website/static/docs/en/html/structfasttext\_1\_1entry.html  
fastText/website/static/docs/en/html/structfasttext\_1\_1entry.js  
fastText/website/static/docs/en/html/structfasttext\_1\_1Node-members.html  
fastText/website/static/docs/en/html/structfasttext\_1\_1Node.html  
fastText/website/static/docs/en/html/structfasttext\_1\_1Node.js  
fastText/website/static/docs/en/html/sync\_off.png  
fastText/website/static/docs/en/html/sync\_on.png  
fastText/website/static/docs/en/html/tabs.css  
fastText/website/static/docs/en/html/tab\_a.png  
fastText/website/static/docs/en/html/tab\_b.png  
fastText/website/static/docs/en/html/tab\_h.png  
fastText/website/static/docs/en/html/tab\_s.png  
fastText/website/static/docs/en/html/utils\_8cc.html  
fastText/website/static/docs/en/html/utils\_8cc.js  
fastText/website/static/docs/en/html/utils\_8h.html  
fastText/website/static/docs/en/html/utils\_8h.js  
fastText/website/static/docs/en/html/utils\_8h\_source.html  
fastText/website/static/docs/en/html/vector\_8cc.html  
fastText/website/static/docs/en/html/vector\_8cc.js  
fastText/website/static/docs/en/html/vector\_8h.html  
fastText/website/static/docs/en/html/vector\_8h.js  
fastText/website/static/docs/en/html/vector\_8h\_source.html  
fastText/website/static/fasttext.css  
fastText/website/static/img/ fastText/website/static/img/authors/  
fastText/website/static/img/authors/armand\_joulin.jpg  
fastText/website/static/img/authors/christian\_puhrsch.png  
fastText/website/static/img/authors/edouard\_grave.jpeg  
fastText/website/static/img/authors/piotr\_bojanowski.jpg  
fastText/website/static/img/authors/tomas\_mikolov.jpg  
fastText/website/static/img/blog/ fastText/website/static/img/blog/2016-08-18-  
blog-post-img1.png fastText/website/static/img/blog/2016-08-18-blog-post-  
img2.png fastText/website/static/img/blog/2017-05-02-blog-post-img1.jpg  
fastText/website/static/img/blog/2017-05-02-blog-post-img2.jpg  
fastText/website/static/img/blog/2017-10-02-blog-post-img1.png  
fastText/website/static/img/cbo\_vs\_skipgram.png  
fastText/website/static/img/fasttext-icon-api.png  
fastText/website/static/img/fasttext-icon-bg-web.png  
fastText/website/static/img/fasttext-icon-color-square.png

```

fastText/website/static/img/fasttext-icon-color-web.png
fastText/website/static/img/fasttext-icon-faq.png
fastText/website/static/img/fasttext-icon-tutorial.png
fastText/website/static/img/fasttext-icon-white-web.png
fastText/website/static/img/fasttext-logo-color-web.png
fastText/website/static/img/fasttext-logo-white-web.png
fastText/website/static/img/logo-color.png
fastText/website/static/img/model-black.png
fastText/website/static/img/model-blue.png
fastText/website/static/img/model-red.png
fastText/website/static/img/ogimage.png
fastText/website/static/img/oss_logo.png
fastText/website/static/tabber.js
fastText/wikifil.pl fastText/word-vector-
example.sh
-- The C compiler identification is GNU 7.5.0
-- The CXX compiler identification is GNU 7.5.0
-- Check for working C compiler: /usr/bin/cc
-- Check for working C compiler: /usr/bin/cc -- works
-- Detecting C compiler ABI info
-- Detecting C compiler ABI info - done
-- Detecting C compile features
-- Detecting C compile features - done
-- Check for working CXX compiler: /usr/bin/c++
-- Check for working CXX compiler: /usr/bin/c++ -- works
-- Detecting CXX compiler ABI info
-- Detecting CXX compiler ABI info - done
-- Detecting CXX compile features
-- Detecting CXX compile features - done
-- Configuring done
-- Generating done
-- Build files have been written to: //root/Tools/fastText/fastText/build
Scanning dependencies of target fasttext-static_pic Scanning
dependencies of target fasttext-shared
[ 2%] Building CXX object CMakeFiles/fasttext-
static_pic.dir/src/args.cc.o
Scanning dependencies of target fasttext-static
[ 6%] Building CXX object CMakeFiles/fasttext-
static_pic.dir/src/densematrix.cc.o
[ 6%] Building CXX object CMakeFiles/fasttext-
static_pic.dir/src/autotune.cc.o
[ 10%] Building CXX object CMakeFiles/fasttext-
static_pic.dir/src/fasttext.cc.o
[ 12%] Building CXX object CMakeFiles/fasttext-
static_pic.dir/src/loss.cc.o

```

```

[ 14 ]%Building CXX object CMakeFiles/fasttext-
static_pic.dir/src/main.cc.o
[ 17 ]%Building CXX object CMakeFiles/fasttext-
static_pic.dir/src/productquantizer.cc.o
[ 18 ]%Building CXX object CMakeFiles/fasttext-
static_pic.dir/src/meter.cc.o
[ 19 ]%Building CXX object CMakeFiles/fasttext-
static_pic.dir/src/utils.cc.o
[ 21 ]%Building CXX object CMakeFiles/fasttext-
static_pic.dir/src/matrix.cc.o
[ 23 ]%Building CXX object CMakeFiles/fasttext-
static_pic.dir/src/quantmatrix.cc.o
[ 25 ]%Building CXX object CMakeFiles/fasttext-
static_pic.dir/src/dictionary.cc.o
[ 29 ]%Building CXX object CMakeFiles/fasttext-
static_pic.dir/src/model.cc.o
[ 31 ]%Building CXX object CMakeFiles/fasttext-
static_pic.dir/src/vector.cc.o
[ 27 ]%Building CXX object CMakeFiles/fasttext-shared.dir/src/args.cc.o [ 34 ] Building CXX
object CMakeFiles/fasttext-
shared.dir/src/dictionary.cc.o
[ 36 ]%Building CXX object CMakeFiles/fasttext-
shared.dir/src/fasttext.cc.o
[ 40 ]%Building CXX object CMakeFiles/fasttext-
shared.dir/src/densematrix.cc.o
[ 42 ]%Building CXX object CMakeFiles/fasttext-
shared.dir/src/autotune.cc.o
[ 44 ]%Building CXX object CMakeFiles/fasttext-shared.dir/src/loss.cc.o [ 46 ] Building CXX
object CMakeFiles/fasttext-shared.dir/src/main.cc.o [ 48 ] Building CXX object
CMakeFiles/fasttext-
shared.dir/src/model.cc.o
[ 51 ]%Building CXX object CMakeFiles/fasttext-
shared.dir/src/matrix.cc.o
[ 53 ]%Building CXX object CMakeFiles/fasttext-
shared.dir/src/vector.cc.o
[ 55 ]%Building CXX object CMakeFiles/fasttext-
shared.dir/src/utils.cc.o

```

```

[ 55 ] %Building CXX object CMakeFiles/fasttext-
shared.dir/src/productquantizer.cc.o
[ 59 ] %Building CXX object CMakeFiles/fasttext-
shared.dir/src/quantmatrix.cc.o
[ 57 ] %Building CXX object CMakeFiles/fasttext-
shared.dir/src/meter.cc.o
[ 63 ] %Building CXX object CMakeFiles/fasttext-static.dir/src/args.cc.o [ 61 ] Building CXX
object CMakeFiles/fasttext-
static.dir/src/autotune.cc.o
[ 65 ] %Building CXX object CMakeFiles/fasttext-
static.dir/src/densematrix.cc.o
[ 68 ] %Building CXX object CMakeFiles/fasttext-
static.dir/src/dictionary.cc.o
[ 70 ] %Building CXX object CMakeFiles/fasttext-static.dir/src/loss.cc.o [ 72 ] Building CXX
object CMakeFiles/fasttext-
static.dir/src/fasttext.cc.o
[ 74 ] %Building CXX object CMakeFiles/fasttext-
static.dir/src/matrix.cc.o
[ 76 ] %Building CXX object CMakeFiles/fasttext-static.dir/src/main.cc.o [ 78 ] Building CXX
object CMakeFiles/fasttext-
static.dir/src/model.cc.o
[ 80 ] %Building CXX object CMakeFiles/fasttext-
static.dir/src/productquantizer.cc.o
[ 85 ] %Building CXX object CMakeFiles/fasttext-
static.dir/src/meter.cc.o
[ 82 ] %Building CXX object CMakeFiles/fasttext-
static.dir/src/quantmatrix.cc.o
[ 87 ] %Building CXX object CMakeFiles/fasttext-
static.dir/src/vector.cc.o
[ 89 ] %Building CXX object CMakeFiles/fasttext-
static.dir/src/utils.cc.o
[ 91 ] %Linking CXX shared library libfasttext.so
[ 93 ] %Linking CXX static library libfasttext_pic.a
[ 95 ] %Linking CXX static library libfasttext.a
[ 95 ] %Built target fasttext-static_pic [ 95 ] Built
target fasttext-static
Scanning dependencies of target fasttext-bin
[ 97 ] %Building CXX object CMakeFiles/fasttext-bin.dir/src/main.cc.o [ 97 ] Built target
fasttext-shared

```

```

[100 %]Linking CXX executable fasttext
[100 %]Built target fasttext-bin
[ 31 %]Built target fasttext-static_pic [ 63 %] Built
target fasttext-static
[ 68 %]Built target fasttext-bin [100 %] Built
target fasttext-shared Install the project...
-- Install configuration: ""
-- Installing: //root/Tools/fastText/lib/libfasttext.so
-- Installing: //root/Tools/fastText/lib/libfasttext.a
-- Installing: //root/Tools/fastText/lib/libfasttext_pic.a
-- Installing: //root/Tools/fastText/bin/fasttext
-- Installing: //root/Tools/fastText/include/fasttext/args.h
-- Installing: //root/Tools/fastText/include/fasttext/autotune.h
-- Installing: //root/Tools/fastText/include/fasttext/densematrix.h
-- Installing: //root/Tools/fastText/include/fasttext/dictionary.h
-- Installing: //root/Tools/fastText/include/fasttext/fasttext.h
-- Installing: //root/Tools/fastText/include/fasttext/loss.h
-- Installing: //root/Tools/fastText/include/fasttext/matrix.h
-- Installing: //root/Tools/fastText/include/fasttext/meter.h
-- Installing: //root/Tools/fastText/include/fasttext/model.h
-- Installing: //root/Tools/fastText/include/fasttext/productquantizer.h
-- Installing: //root/Tools/fastText/include/fasttext/quantmatrix.h
-- Installing: //root/Tools/fastText/include/fasttext/real.h
-- Installing: //root/Tools/fastText/include/fasttext/utils.h
-- Installing: //root/Tools/fastText/include/fasttext/vector.h
/content

```

## 1 Exercise 1

- Q1: Why should we use different texts for learning and testing? we need to use different data for testing and training to know whether our model has properly learn from the given data

### 1.1 split to an equal number of line for train.fr and test.fr

[89]:

```
f = open("/content/fr.txt", encoding = 'utf-8')
array_lines = f.readlines()
```

#### 1.1.1 the file contain the following lines

train , test = Total number of lines / 2

[92]:

```
len(array_lines)
```

[92]: 20000



```
[102]: train = array_lines[0          : int(len(array_lines)/2 ) ]test
      = array_lines[ int(len(array_lines)/2 ) :          ]
```

### 1.1.2 creating the train.fr and test.fr

```
[108]: with open('/content/train.fr','w') as f:
      for i in range(len(train)): f.writelines(train[i])
```

```
[109]: with open('/content/test.fr','w') as f:
      for i in range(len(test)): f.writelines(test[i])
```

### 1.1.3 writting it back in the necessairy file

## 2 Exercise 2:

Construct a trigram language model for each training file previously created using the following commands:

```
ngram-count -order 3 -text train.fr -write fr3g.bo -write-vocabfr.voc
```

```
ngram-count -read fr3g.bo -lm fr3g.BO -gt2min 1 -gt3min 2
```

```
[21]: pwd
```

```
[21]: '/root/Tools'
```

### 2.1 fasttext and SRILM are installed proopertly

```
[23]: ls
```

```
fastText/    SRILM-1.6.0/
```

```
[30]: cd i686-m64/
```

```
/root/Tools/SRILM-1.6.0/bin/i686-m64
```

```
[31]: ls
```

add-classes-to-pfsg*	htklat-vocab*
pfsg-to-dot*	
add-dummy-bows*	lattice-tool*
pfsg-to-fsm*	
add-pauses-to-pfsg*	log10-to-bytelog*
pfsg-vocab*	
add-ppls*	make-abs-discount*
ppl-from-log*	

anti-ngram*	make-diacritic-map*
prettify*	make-google-ngrams*
bytelog-to-log10*	make-gt-discounts*
remove-lowprob-ngrams*	make-hiddens-lm*
classes-to-fsm*	make-kn-counts*
replace-words-with-classes*	make-kn-discounts*
combine-acoustic-scores*	make-lm-subset*
reverse-lm*	make-nbest-pfsg*
combine-rover-controls*	make-ngram-pfsg*
reverse-ngram-counts*	make-sub-lm*
compare-ppls*	maxalloc*
reverse-text*	merge-nbest*
compute-best-mix*	metadb*
segment*	multi-ngram*
compute-best-rover-mix*	nbest2-to-nbest1*
segment-nbest*	nbest-lattice*
compute-best-sentence-mix*	nbest-mix*
select-vocab*	nbest-optimize*
compute-oov-rate*	nbest-posteriors*
sentid-to-ctm*	nbest-pron-score*
context-ngrams*	nbest-vocab*
sentid-to-sclite*	ngram*
continuous-ngram-count*	ngram-class*
sort-lm*	ngram-count*
cumbin*	ngram-merge*
split-tagged-ngrams*	
disambig*	
subset-context-ngrams*	
extract-skip-probs*	
subtract-ppls*	
filter-event-counts*	
tolower-ngram-counts*	
find-reference-posteriors*	
uniform-classes*	
fix-ctm*	
uniq-ngram-counts*	
fngram*	
vp2text*	
fngram-count*	
wlat-stats*	
fsm-to-pfsg*	
wlat-to-dot*	
get-gt-counts*	
wlat-to-pfsg*	
get-unigram-probs*	
wordlat-to-lisp*	
hidden-ngram*	
hits-from-log*	

```
[159]: !./ngram-count -order 3 -text /content/train.fr -write /content/fr3g.bo  
↪-write-vocab /content/fr.voc
```

```
[160]: !./ngram-count -read /content/fr3g.bo -lm /content/fr3g.BO -gt2min 1 -gt3min2
```

warning: discount coeff 1 is out of range: 0

### 2.1.1 Q1: Observe files .voc , .bo and .BO ; what do they contain?

```
[117]: BO = open("/content/fr3g.BO", encoding='utf-8')  
BO.readlines()[0:10]
```

```
[117]: ['\n',  
'\data\\n', 'ngram  
1=13997\n', 'ngram  
2=69997\n', 'ngram  
3=18186\n', '\n',  
'\1-grams:\n',  
'-3.156491\t!\t-0.8377973\n', '-  
2.767266\t"\t-0.2054707\n', '-  
3.66337\t\t-0.5037488\n']
```

```
[119]: bo = open("/content/fr3g.bo", encoding='utf-8')  
bo.readlines()[0:10]
```

```
[119]: ['professionnels\t5\n', 'professionnels .\t1\n',  
'professionnels . </s>\t1\n', 'professionnels  
"\t1\n', 'professionnels " .\t1\n',  
'professionnels de\t2\n', 'professionnels de  
la\t1\n', 'professionnels de santé\t1\n',  
'professionnels qui\t1\n',  
'professionnels qui commercialisent\t1\n']
```

```
[121]: voc_ = open("/content/fr.voc", encoding='utf-8')  
voc_.readlines()[0:10]
```

```
[121]: ['!\n', '"\n', '\n', '&\n', '%\n', '(\n', ')\n', '+\n', ',\n', '-\n']
```

### 2.1.2 Q2: In your opinion, what do the symbols " <'s'> and </s'> " correspond to?

```
[ ]:
```

### 2.1.3 What are the -gt2min and -gt3min options for ? In what cases are they useful / necessary?

- it use to modify the count of the model

### 2.1.4 Q4: What is the most common trigram in each language?

## 3 Exercice 3

### 3.0.1 Present the results in the form of a matrix.

[142]:

```
!./ngram -lm /content/fr3g.BO -ppl/content/test.fr
```

file /content/test.fr: 10000 sentences, 208169 words, 7456 OOVs  
0 zeroprobs, logprob= -366964 ppl= 55.1489 ppl1= 67.3448

### 3.0.2 Q2: How can we build a language recognition system for a text using a stochastic language model?

- The primary purpose of specifying a stochastic grammar format is to support large vocabulary and open vocabulary applications. In addition, stochastic grammars can be used to represent concepts or semantics. This specification defines the mechanism for combining stochastic and structured (in this case Context-Free) grammars as well as methods for combined semantic definitions. Since some structured grammars are also stochastic, we will avoid confusion from here on by only referring to these grammars as N-Gram grammars, or in some cases simply N-Grams.
- An N-Gram grammar is a representation of an N-th order Markov language model in which the probability of occurrence of a symbol is conditioned upon the prior occurrence of N-1 other symbols. N-Gram grammars are typically constructed from statistics obtained from a large corpus of text using the co-occurrences of words in the corpus to determine word sequence probabilities. N-Gram grammars have the advantage of be able to cover a much larger language than would normally be derived directly from a corpus. Open vocabulary applications are easily supported with N-Gram grammars.

## 4 Exercice 4

Redo the same system with n-gram CHARACTER models.

### 4.0.1 Q1: What are the perplexities for this character model? What might be the advantages / disadvantages of the character model compared to the word model?

```
!./ngram-count -lm option
```

[156]: warning: no singleton counts GT  
discounting disabled warning: no  
singleton counts GT discounting  
disabled

warning: no singleton counts GT  
discounting disabled

[163]: `!./ngram-count -read /content/fr3g.bo -lm character /content/fr3g.BO -gt2min1  
↪-gt3min 2`

warning: discount coeff 1 is out of range: 0

[164]: `!./ngram -lm /content/fr3g.BO -ppl/content/test.fr`

file /content/test.fr: 10000 sentences, 208169 words, 7456 OOVs  
0 zero probs, logprob= -366964 ppl= 55.1489 ppl1= 67.3448

[165]: `pwd`

[165]: `'/root/Tools/SRILM-1.6.0/bin/i686-m64'`

[178]: `ls`

<code>alignment/</code>	<code>eval.py*</code>
<code>scripts/</code>	
<code>build/</code>	<code>get-wikimedia.sh*</code>
<code>setup.cfg*</code>	
<code>classification-example.sh*</code>	<code>LICENSE*</code>
<code>setup.py*</code>	
<code>classification-results.sh*</code>	<code>Makefile*</code>
<code>src/</code>	
<code>CMakeLists.txt*</code>	<code>MANIFEST.in*</code>
<code>tests/</code>	
<code>CODE_OF_CONDUCT.md*</code>	<code>python/</code>
<code>website/</code>	
<code>CONTRIBUTING.md*</code>	<code>quantization-example.sh*</code>
<code>wikifil.pl*</code>	
<code>crawl/</code>	<code>README.md*</code>
<code>word-vector-example.sh*</code>	
<code>docs/</code>	<code>reduce_model.py*</code>
<code>download_model.py*</code>	<code>runtests.py*</code>

[189]: `cd fastText/`

`/root/Tools/fastText/fastText`

[190]: `pip install fasttext`

Collecting fasttext

Downloading <https://files.pythonhosted.org/packages/f8/85/e2b368ab6d3528827b147fdb814f8189acc981a4bc2f99ab894650e05c40/fasttext-0.9.2.tar.gz>(68kB)  
| | 71kB 2.0MB/s

Requirement already satisfied: pybind11>=2.2 in  
/usr/local/lib/python3.6/dist-packages (from fasttext)(2.5.0) Requirement  
already satisfied: setuptools>=0.7.0 in  
/usr/local/lib/python3.6/dist-packages (from fasttext) (47.3.1)  
Requirement already satisfied: numpy in /usr/local/lib/python3.6/dist-packages (from fasttext)  
(1.18.5)  
Building wheels for collected packages:fasttext Building  
wheel for fasttext (setup.py) ... done Created wheel for  
fasttext:  
filename=fasttext-0.9.2-cp36-cp36m-linux\_x86\_64.whl size=3016213  
sha256=ee64ad995646ad6819e536f141160154a0289828c59d222aa5b04e3d33b70ce3  
Stored in directory:/root/.cache/pip/wheels/98/ba/7f/b154944a1cf5a8cee91c154b  
75231136cc3a3321ab0e30f592  
Successfully built fasttext  
Installing collected packages: fasttext Successfully  
installed fasttext-0.9.2

[218]: `pwd`

[218]: `'/root/Tools/fastText/include/fasttext'`

[220]: `!./fasttext skipgram -input data.txt -outputmodel`

/bin/bash: ./fasttext: No such file or directory

[221]: `import fasttext`

[228]: `from gensim.test.utils import common_texts`

[226]: `model =fasttext.train_supervised('/content/train.fr.txt')`

[230]: `model.get_dimension()`

[230]: 100

[ ]: