

INDEKS *FILE* (TEKS) MENGGUNAKAN SWISH-E

by: [M. SALSABILA JAMIL \(mailto:jamilsalsabila@gmail.com\)](mailto:M.SALSABILA.JAMIL)

SWISH-e (<https://www.esa.org/tiee/search/html/index.html>) adalah sebuah *tool* yang digunakan untuk meng-indeks teks dalam berbagai format, seperti PDF, html, txt, XML, PostScript, dll. Selain meng-indeks, *tool* ini juga dapat mencari dokumen berdasarkan *query* yang di-input-kan. *Tool* ini memiliki banyak fitur/kemampuan yang dimilikinya, contohnya:

- Dapat menentukan *stopword list* yang hendak digunakan
- Dapat Menentukan direktori (berisi file) yang ingin di-indeks
- Dapat Menentukan jenis file yang di-indeks (PDF?, XML?, email?, txt?, html?, dll)
- Bisa melakukan *stemming* (dalam beberapa bahasa)
- *Fast*
- Dapat meng-indeks berita dari suatu *website* (*crawling*)
- Dapat Mencari dokumen yang mengandung *query* yang dimasukkan
- Dan masih banyak lagi -> [Baca Lebih Lanjut \(https://www.esa.org/tiee/search/html/readme.html\)](https://www.esa.org/tiee/search/html/readme.html)

Berikut dijelaskan cara menggunakan *tool* ini.

Langkah pertama yaitu **DOWNLOAD FILE BERISI TEKS** (disini penulis menggunakan TEKS BERITA)

1. Penulis membuat skrip untuk mengunduh berita pada website:

- **KOMPAS** (<https://www.kompas.com>): `src/spider/spider_kompas.go`
- **TEMPO** (<https://www.tempo.co>): `src/spider/spider_tempo.go`
- **ANTARA** (<https://www.antaranews.com>):
 - `src/spider/spider_antara.go`
 - `src/spider/spider_antara_oto.go`

2. Tema-tema berita yang di-download:

- *Entertainment*/Hiburan
- *Sport*/Olahraga
- Tekno
- Otomotif

3. Bagian-bagian *HTML* yang diambil:

- *Title*/Judul
- Tanggal Terbit
- *Tag* (Opsional)
- *Content*/Isi

4. Berita yang terkumpul Dapat dilihat pada *folder*:

- `kompas/`
- `tempo/`
- `antara/`

5. Selanjutnya dilakukan proses **PREPROCESSING**:

- Menghapus `\n\r`
- Menghapus tanda baca yang dianggap ***TIDAK PENTING**
- Mengubah *spasi* lebih dari satu menjadi satu *spasi*
- *Lowercase every letter*
- Hasil disimpan pada folder:
 - `ALL/`
 - Secara keseluruhan berjumlah: **~55K**

Langkah kedua, **STEMMING**:

1. Pengertian:

Mengubah kata ke bentuk dasarnya, contoh:

- berjalan -> jalan
- memancing -> pancing
- menggunakan -> guna

2. Tujuan:

Sebisa mungkin mengurangi kata yang nantinya akan di-indeks, sehingga jumlah memori yang digunakan semakin sedikit

- Proses *STEMMING* menggunakan pustaka/modul yang dikembangkan oleh [RadhiFadillah](https://github.com/RadhiFadillah/go-sastrawi) (<https://github.com/RadhiFadillah/go-sastrawi>) dalam bahasa [GO](https://golang.org/) (<https://golang.org/>), dimana RadhiFadillah mengambil referensi dari [andylibrian](https://github.com/sastrawi/sastrawi) (<https://github.com/sastrawi/sastrawi>) yang mengembangkannya dalam bahasa PHP.
- Pustaka (kopi sebagian dari [README.md](#) RadhiFadillah (<https://github.com/RadhiFadillah/go-sastrawi>))

Algoritma

- Algoritma Nazief dan Adriani
- Asian J. 2007. *Effective Techniques for Indonesian Text Retrieval*. PhD thesis School of Computer Science and Information Technology RMIT University Australia. (PDF (<http://researchbank.mmit.edu.au/eserv/mit:6312/Asian.pdf>) dan [Amazon](https://www.amazon.com/Effective-Techniques-Indonesian-Text-Retrieval/dp/3639021649) (<https://www.amazon.com/Effective-Techniques-Indonesian-Text-Retrieval/dp/3639021649>))
- Arifin, A.Z., I.P.A.K. Mahendra dan H.T. Ciptaningtyas. 2009. *Enhanced Confix Stripping Stemmer and Ants Algorithm for Classifying News Document in Indonesian Language*, Proceeding of International Conference on Information & Communication Technology and Systems (ICTS). (PDF (<http://personal.its.ac.id/files/pub/2623-agusza-baru%2021%20d%20VIP%20enhanced-confix-stripping-stem.pdf>))
- A. D. Tahitoe, D. Purwitasari. 2010. *Implementasi Modifikasi Enhanced Confix Stripping Stemmer Untuk Bahasa Indonesia dengan Metode Corpus Based Stemming*, Institut Teknologi Sepuluh Nopember (ITS) – Surabaya, 60111, Indonesia. (PDF (<http://digilib.its.ac.id/public/ITS-Undergraduate-14255-paperpdf.pdf>))
- Tambahan aturan *stemming* dari [kontributor Sastrawi](https://github.com/sastrawi/sastrawi/graphs/contributors) (<https://github.com/sastrawi/sastrawi/graphs/contributors>).

Kamus Kata Dasar

Proses stemming oleh Sastrawi sangat bergantung pada kamus kata dasar. Sastrawi menggunakan kamus kata dasar dari kateglo.com (<http://kateglo.com>) dengan sedikit perubahan.

Lisensi

Sebagaimana [Sastrawi](https://github.com/sastrawi/sastrawi) (<https://github.com/sastrawi/sastrawi>) untuk PHP, [Go-Sastrawi](https://github.com/RadhiFadillah/go-sastrawi) (<https://github.com/RadhiFadillah/go-sastrawi>) untuk GO, proyek ini juga disebar dengan lisensi [MIT](http://choosealicense.com/licenses/mit/) (<http://choosealicense.com/licenses/mit/>). Untuk lisensi kamus kata dasar dari Kateglo adalah [CC-BY-NC-SA 3.0](https://github.com/ivanlanin/kateglo#lisensi-isi) (<https://github.com/ivanlanin/kateglo#lisensi-isi>).

- Proses *STEMMING* dilakukan menggunakan skrip yang terdapat di folder `src/preprocess/stemming/main.go`
- Cuplikan ketika proses *STEMMING* berlangsung:

- Hasil *STEMMING* ditaruh di folder:

- ALL_2/kompas
- ALL_2/tempo
- ALL_2/antara

Langkah ketiga, *INDEXING*:

- Menggunakan tool *swish-e* ([download](https://github.com/swish-e/swish-e) (<https://github.com/swish-e/swish-e>))
 - Di-install pada OS Linux 16.04 LTS 64bit
 - Proses instalasi dapat dibaca [disini](https://www.esa.org/tiee/search/html/install.html) (<https://www.esa.org/tiee/search/html/install.html>)
- Download

- o Unzip
- o \$ cd
- o \$./configure

```

s@L: swish-e-naster
checking for vsnprintf... yes
checking for mkstemp... yes
checking for xml2-config... no
Not building with libxml2 - use --with-libxml2 to enable
checking for zlib.h usability... yes
checking for zlib.h presence... yes
checking for gzread in -lz... yes
Not building with perl compatible regex - use --with-pcre to enable
checking for special C compiler options needed for large files... no
checking for _FILE_OFFSET_BITS value needed for large files... no
configure: fileoffset-bits = no
configure: Setting libexecdir to $(exec_prefix)/lib/swish-e
checking config option mendebug for setting MEM_DEBUG... no
checking config option mnttrace for setting MEM_TRACE... no
checking config option menstats for setting MEM_STATISTICS... no
configure: creating ./config.status
config.status: creating html/Makefile
config.status: creating pod/Makefile
config.status: creating man/Makefile
config.status: creating src/Makefile
config.status: creating src/expat/Makefile
config.status: creating src/replace/Makefile
config.status: creating src/snowball/Makefile
config.status: creating rpm/swish-e.spec
config.status: creating tests/Makefile
config.status: creating example/Makefile
config.status: creating filters/Makefile
config.status: creating filters/swish/Makefile
config.status: creating conf/Makefile
config.status: creating filter-bin/Makefile
config.status: creating swish-e.pc
config.status: creating swish-config
config.status: creating src/acconfig.h
config.status: executing depfiles commands
07:47 PM Sun Oct 06
~/Desktop/KULIAH/RDM/tugas1/TEXT/swish-e-naster

```

- o \$ make

```

s@L: swish-e-naster
swish-e-naster -> make
Making all in filters
make[1]: Entering directory '/home/s/Desktop/KULIAH/RDM/tugas1/TEXT/swish-e-naster/filters'
make[2]: Entering directory '/home/s/Desktop/KULIAH/RDM/tugas1/TEXT/swish-e-naster/filters/SWISH'
make[2]: Leaving directory '/home/s/Desktop/KULIAH/RDM/tugas1/TEXT/swish-e-naster/filters/SWISH'
make[2]: Entering directory '/home/s/Desktop/KULIAH/RDM/tugas1/TEXT/swish-e-naster/filters'
make[2]: Leaving directory '/home/s/Desktop/KULIAH/RDM/tugas1/TEXT/swish-e-naster/filters'
make[1]: Leaving directory '/home/s/Desktop/KULIAH/RDM/tugas1/TEXT/swish-e-naster/filters'
Making all in prog-bin
make[1]: Entering directory '/home/s/Desktop/KULIAH/RDM/tugas1/TEXT/swish-e-naster/program'
make[1]: Leaving directory '/home/s/Desktop/KULIAH/RDM/tugas1/TEXT/swish-e-naster/program'
Making all in conf
make[1]: Entering directory '/home/s/Desktop/KULIAH/RDM/tugas1/TEXT/swish-e-naster/conf'
make[1]: Nothing to be done for 'all'.
make[1]: Leaving directory '/home/s/Desktop/KULIAH/RDM/tugas1/TEXT/swish-e-naster/conf'
Making all in filter-bin
make[1]: Entering directory '/home/s/Desktop/KULIAH/RDM/tugas1/TEXT/swish-e-naster/filter-bin'
make[1]: Leaving directory '/home/s/Desktop/KULIAH/RDM/tugas1/TEXT/swish-e-naster/filter-bin'
Making all in example
make[1]: Entering directory '/home/s/Desktop/KULIAH/RDM/tugas1/TEXT/swish-e-naster/example'
make[1]: Leaving directory '/home/s/Desktop/KULIAH/RDM/tugas1/TEXT/swish-e-naster/example'
Making all in html
make[1]: Entering directory '/home/s/Desktop/KULIAH/RDM/tugas1/TEXT/swish-e-naster/html'
make[1]: Nothing to be done for 'all'.
make[1]: Leaving directory '/home/s/Desktop/KULIAH/RDM/tugas1/TEXT/swish-e-naster/html'
Making all in man
make[1]: Entering directory '/home/s/Desktop/KULIAH/RDM/tugas1/TEXT/swish-e-naster/man'
make[1]: Nothing to be done for 'all'.
make[1]: Leaving directory '/home/s/Desktop/KULIAH/RDM/tugas1/TEXT/swish-e-naster/man'
Making all in src
make[1]: Entering directory '/home/s/Desktop/KULIAH/RDM/tugas1/TEXT/swish-e-naster/src'
make all-recursive
make[2]: Entering directory '/home/s/Desktop/KULIAH/RDM/tugas1/TEXT/swish-e-naster/src'
Making all in expat
make[3]: Entering directory '/home/s/Desktop/KULIAH/RDM/tugas1/TEXT/swish-e-naster/src/expat'
if /bin/bash ../libtool --tag=C --mode=compile gcc -DHAVE_CONFIG_H -I. -I../src -I../xmlparse -I../xsltproc -I../xsltproc -g -O2 -MT xsltproc.lo -MD -MP -MF .deps/xsltproc.Tpo -c -o xsltproc.lo xsltproc.c; \
then mv -f ".deps/xsltproc.Tpo" ".deps/xsltproc.Plo"; else rm -f ".deps/xsltproc.Tpo"; exit 1; fi
mkdir .libs
gcc -DHAVE_CONFIG_H -I. -I../src -I../xmlparse -I../xsltproc -g -O2 -MT xsltproc.lo -MD -MP -MF .deps/xsltproc.Tpo -c xsltproc.c -fpic -DPIC -o .libs/xsltproc.o

```

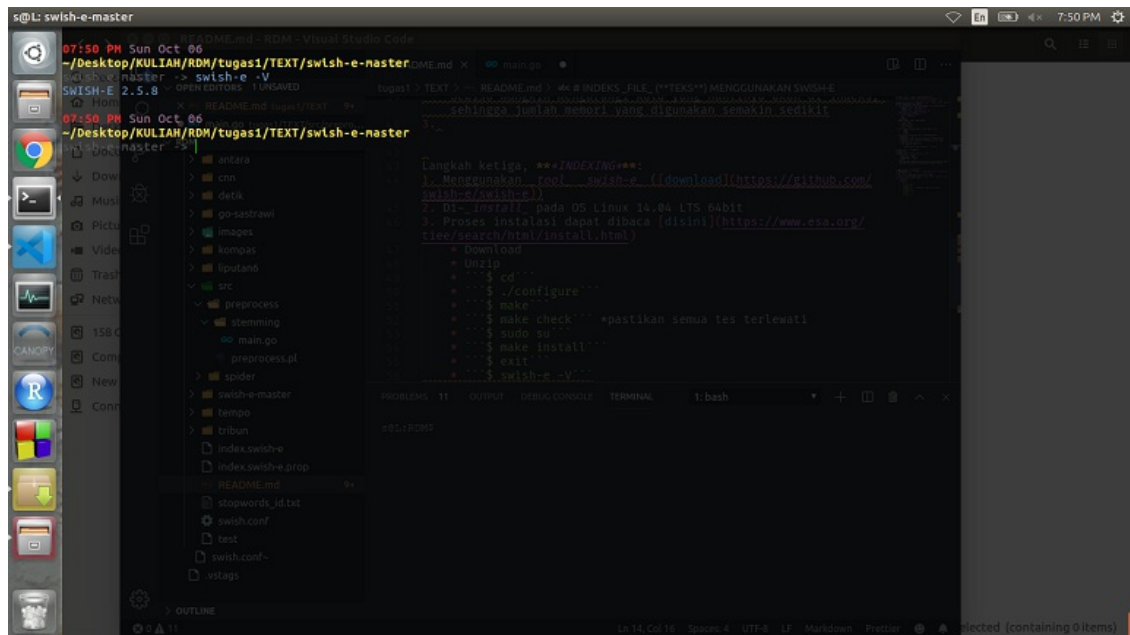
- o \$ make check *pastikan semua tes berhasil


```
s@L: swish-e-master
make[1]: Nothing to be done for 'check'.
make[1]: Leaving directory '/home/s/Desktop/KULIAH/RDM/tugas1/TEXT/swish-e-master/man'
Making check in src
make[1]: Entering directory '/home/s/Desktop/KULIAH/RDM/tugas1/TEXT/swish-e-master/src'
Making check in expat
make[2]: Entering directory '/home/s/Desktop/KULIAH/RDM/tugas1/TEXT/swish-e-master/src/expat'
make[2]: Nothing to be done for 'check'.
make[2]: Leaving directory '/home/s/Desktop/KULIAH/RDM/tugas1/TEXT/swish-e-master/src/expat'
Making check in replace
make[2]: Entering directory '/home/s/Desktop/KULIAH/RDM/tugas1/TEXT/swish-e-master/src/replace'
make[2]: Nothing to be done for 'check'.
make[2]: Leaving directory '/home/s/Desktop/KULIAH/RDM/tugas1/TEXT/swish-e-master/src/replace'
Making check in snowball
make[2]: Entering directory '/home/s/Desktop/KULIAH/RDM/tugas1/TEXT/swish-e-master/src/snowball'
make[2]: Nothing to be done for 'check'.
make[2]: Leaving directory '/home/s/Desktop/KULIAH/RDM/tugas1/TEXT/swish-e-master/src/snowball'
make[2]: Entering directory '/home/s/Desktop/KULIAH/RDM/tugas1/TEXT/swish-e-master/src'
make[1]: Leaving directory '/home/s/Desktop/KULIAH/RDM/tugas1/TEXT/swish-e-master/src'
Making check in tests
make[1]: Entering directory '/home/s/Desktop/KULIAH/RDM/tugas1/TEXT/swish-e-master/tests'
make check-TESTS
make[2]: Entering directory '/home/s/Desktop/KULIAH/RDM/tugas1/TEXT/swish-e-master/tests'
PASS: check_index
PASS: check_search
PASS: check_metasearch
PASS: check_fuzzy
=====
All 4 tests passed
=====
make[2]: Leaving directory '/home/s/Desktop/KULIAH/RDM/tugas1/TEXT/swish-e-master/tests'
make[1]: Leaving directory '/home/s/Desktop/KULIAH/RDM/tugas1/TEXT/swish-e-master/tests'
Making check in pod
make[1]: Entering directory '/home/s/Desktop/KULIAH/RDM/tugas1/TEXT/swish-e-master/pod'
make[1]: Nothing to be done for 'check'.
make[1]: Leaving directory '/home/s/Desktop/KULIAH/RDM/tugas1/TEXT/swish-e-master/pod'
make[1]: Entering directory '/home/s/Desktop/KULIAH/RDM/tugas1/TEXT/swish-e-master'
make[1]: Nothing to be done for 'check-an'.
make[1]: Leaving directory '/home/s/Desktop/KULIAH/RDM/tugas1/TEXT/swish-e-master'
root@L: /home/s/Desktop/KULIAH/RDM/tugas1/TEXT/swish-e-master
```

- o \$ sudo su
- o \$ make install

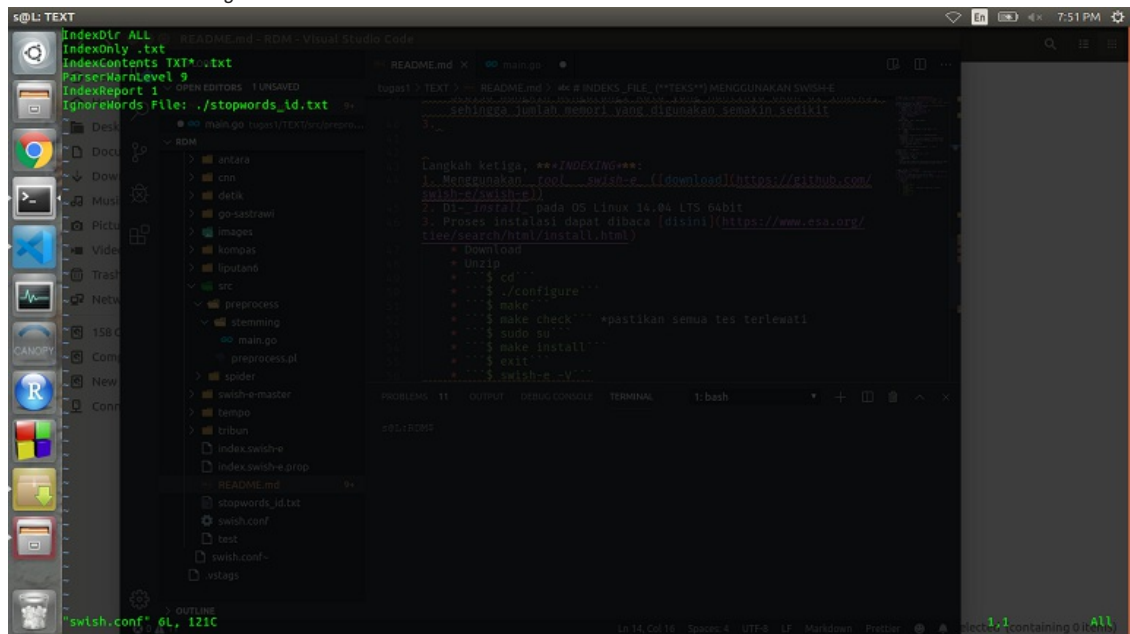
```
root@L: /home/s/Desktop/KULIAH/RDM/tugas1/TEXT/swish-e-master
test -z "/usr/local/include" || mkdir -p -- "/usr/local/include"
/usr/bin/install -c -m 644 'swish-e.h' '/usr/local/include/swish-e.h'
make[1]: Leaving directory '/home/s/Desktop/KULIAH/RDM/tugas1/TEXT/swish-e-master/src'
make[2]: Leaving directory '/home/s/Desktop/KULIAH/RDM/tugas1/TEXT/swish-e-master/src'
make[1]: Leaving directory '/home/s/Desktop/KULIAH/RDM/tugas1/TEXT/swish-e-master/src'
Making install in tests
make[1]: Entering directory '/home/s/Desktop/KULIAH/RDM/tugas1/TEXT/swish-e-master/tests'
make[2]: Entering directory '/home/s/Desktop/KULIAH/RDM/tugas1/TEXT/swish-e-master/tests'
make[2]: Nothing to be done for 'install-exec-am'.
make[2]: Nothing to be done for 'install-data-am'.
make[2]: Leaving directory '/home/s/Desktop/KULIAH/RDM/tugas1/TEXT/swish-e-master/tests'
make[1]: Leaving directory '/home/s/Desktop/KULIAH/RDM/tugas1/TEXT/swish-e-master/tests'
Making install in pod
make[1]: Entering directory '/home/s/Desktop/KULIAH/RDM/tugas1/TEXT/swish-e-master/pod'
make[2]: Entering directory '/home/s/Desktop/KULIAH/RDM/tugas1/TEXT/swish-e-master/pod'
make[2]: Nothing to be done for 'install-exec-am'.
make[2]: Leaving directory '/home/s/Desktop/KULIAH/RDM/tugas1/TEXT/swish-e-master/pod'
make[1]: Leaving directory '/home/s/Desktop/KULIAH/RDM/tugas1/TEXT/swish-e-master/pod'
test -z "/usr/local/share/doc/swish-e/pod" || mkdir -p -- "/usr/local/share/doc/swish-e/pod"
/usr/bin/install -c -m 644 'CHANGES.pod' '/usr/local/share/doc/swish-e/pod/CHANGES.pod'
/usr/bin/install -c -m 644 'INSTALL.pod' '/usr/local/share/doc/swish-e/pod/INSTALL.pod'
/usr/bin/install -c -m 644 'README.pod' '/usr/local/share/doc/swish-e/pod/README.pod'
/usr/bin/install -c -m 644 'SWISH-3.0.pod' '/usr/local/share/doc/swish-e/pod/SWISH-3.0.pod'
/usr/bin/install -c -m 644 'SWISH-BUGS.pod' '/usr/local/share/doc/swish-e/pod/SWISH-BUGS.pod'
/usr/bin/install -c -m 644 'SWISH-CONFIG.pod' '/usr/local/share/doc/swish-e/pod/SWISH-CONFIG.pod'
/usr/bin/install -c -m 644 'swish-e.pod' '/usr/local/share/doc/swish-e/pod/swish-e.pod'
/usr/bin/install -c -m 644 'SWISH-FAQ.pod' '/usr/local/share/doc/swish-e/pod/SWISH-FAQ.pod'
/usr/bin/install -c -m 644 'SWISH-LIBRARY.pod' '/usr/local/share/doc/swish-e/pod/SWISH-LIBRARY.pod'
/usr/bin/install -c -m 644 'SWISH-RUN.pod' '/usr/local/share/doc/swish-e/pod/SWISH-RUN.pod'
/usr/bin/install -c -m 644 'SWISH-SEARCH.pod' '/usr/local/share/doc/swish-e/pod/SWISH-SEARCH.pod'
make[2]: Leaving directory '/home/s/Desktop/KULIAH/RDM/tugas1/TEXT/swish-e-master/pod'
make[1]: Leaving directory '/home/s/Desktop/KULIAH/RDM/tugas1/TEXT/swish-e-master/pod'
make[1]: Entering directory '/home/s/Desktop/KULIAH/RDM/tugas1/TEXT/swish-e-master'
test -z "/usr/local/bin" || mkdir -p -- "/usr/local/bin"
/usr/bin/install -c 'swish-config' '/usr/local/bin/swish-config'
test -z "/usr/local/share/doc/swish-e" || mkdir -p -- "/usr/local/share/doc/swish-e"
/usr/bin/install -c -m 644 './INSTALL' '/usr/local/share/doc/swish-e/INSTALL'
/usr/bin/install -c -m 644 './README.cvs' '/usr/local/share/doc/swish-e/README.cvs'
test -z "/usr/local/lib/pkgconfig" || mkdir -p -- "/usr/local/lib/pkgconfig"
/usr/bin/install -c -m 644 'swish-e.pc' '/usr/local/lib/pkgconfig/swish-e.pc'
make[2]: Leaving directory '/home/s/Desktop/KULIAH/RDM/tugas1/TEXT/swish-e-master'
make[1]: Leaving directory '/home/s/Desktop/KULIAH/RDM/tugas1/TEXT/swish-e-master'
root@L: /home/s/Desktop/KULIAH/RDM/tugas1/TEXT/swish-e-master#
```

- o \$ exit
 - o \$ swish-e -V
- SWISH-E 2.5.8

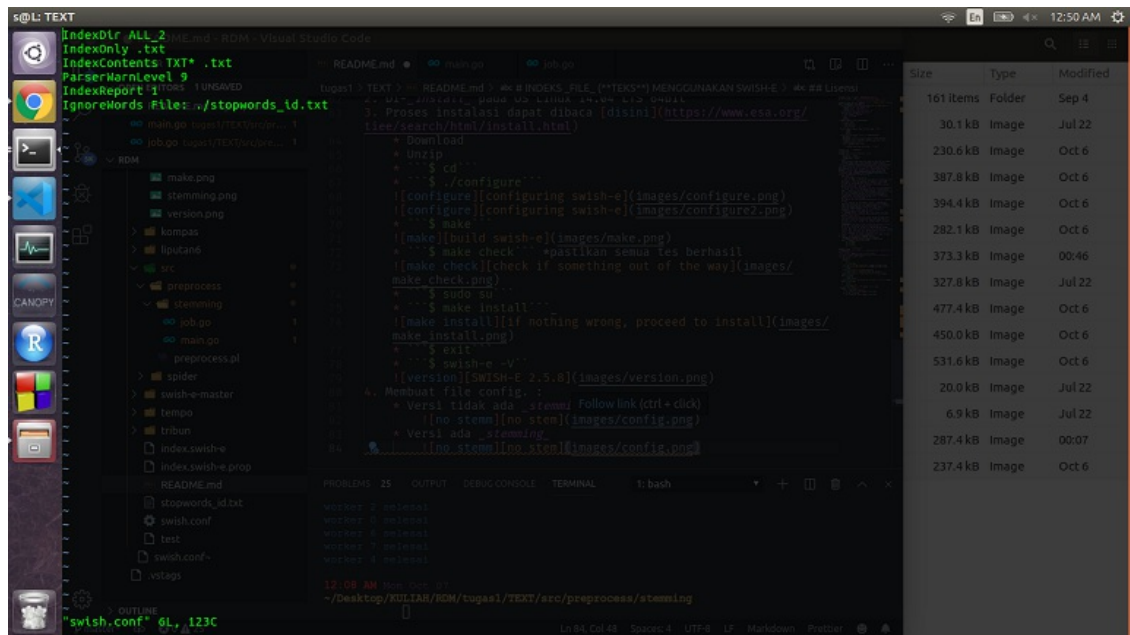


4. Membuat file config. : [BACA LEBIH LANJUT \(https://www.esa.org/tiee/search/html/swish-config.html\)](https://www.esa.org/tiee/search/html/swish-config.html)

- o Versi tidak ada *stemming*



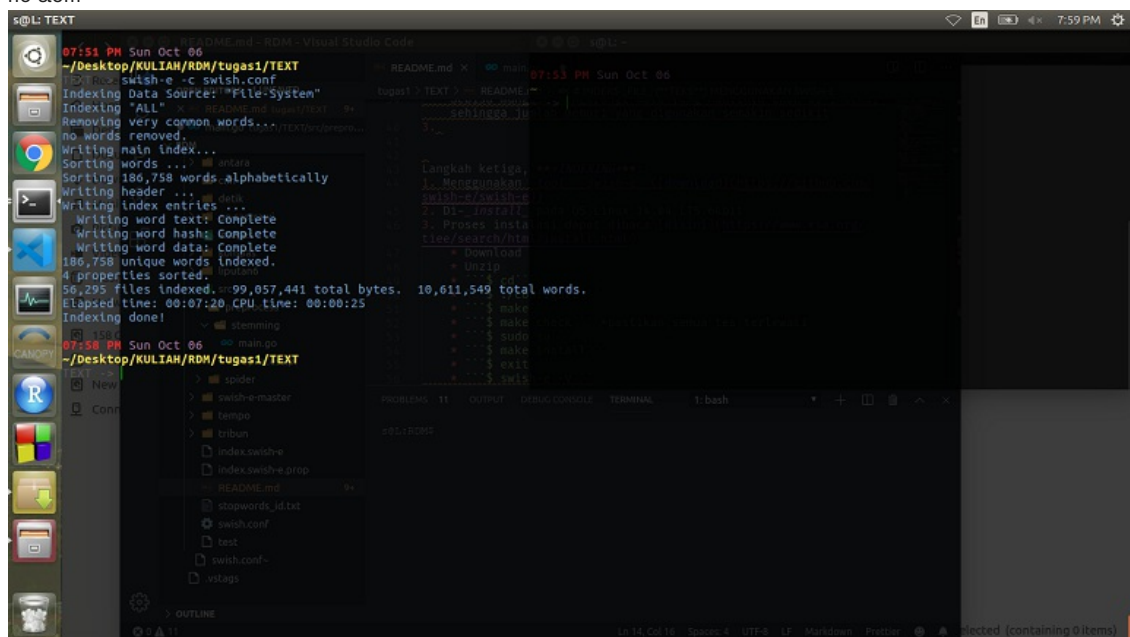
- o Versi ada *stemming*



5. Konfigurasi proses peng-indeks-an tercantum dalam *file* config yang telah dibuat: **swish.conf**.

Perlu diperhatikan adanya pemangkasan kata umum (common words) (stopwords) menggunakan data stopwords bahasa indonesia (stopwords_id.txt*) yang dapat di-download melalui link ini (<https://docs.google.com/viewer?u=v&pid=sites&scid=ZGVmYXVsdGRvbWFpbmRvZGd4OjI4ZWY0ZGVkMzJmM2FkYWU>). Data Stopwords ini didapat pada website sites.google.com/site/kevinbouge/stopwords-lists (<https://sites.google.com/site/kevinbouge/stopwords-lists>).

- o no stem



- o with stem


```

12:34 AM Mon Oct 07
~/Desktop/KULIAH/RDM/tugas1/TEXT
TEXT -> swish-e -c swish.conf -f index_after_stemmed.swish-e-ancing
Indexing Data Source: "File-System"
Indexing "ALL_2"
Removing very common words...
no words removed.
Writing main index...
Sorting words ...
Sorting 162,450 words alphabetically
Writing header...
Writing word text: Complete
Writing word hash: Complete
Writing word data: Complete
162,450 unique words indexed.
4 properties sorted.
56,295 files indexed. 88,941,833 total bytes
Elapsed time: 00:09:46 CPU time: 00:00:26
Indexing done! preprocess.pl

12:44 AM Mon Oct 07
~/Desktop/KULIAH/RDM/tugas1/TEXT
TEXT ->

```

6. Dapat dilihat *file* yang diindeks berjumlah **56295 files**

7. Waktu peng-indeks-an: **~10 menit**

8. Jumlah *Unique words*:

- no stem: **~186K**
- with stem: **~162K**

9. Terlihat perbedaan mencolok antara **no stem** dan **with stem** dalam hal jumlah *Unique words*

10. Uji Coba *search* sebuah kata:

- no stem

```

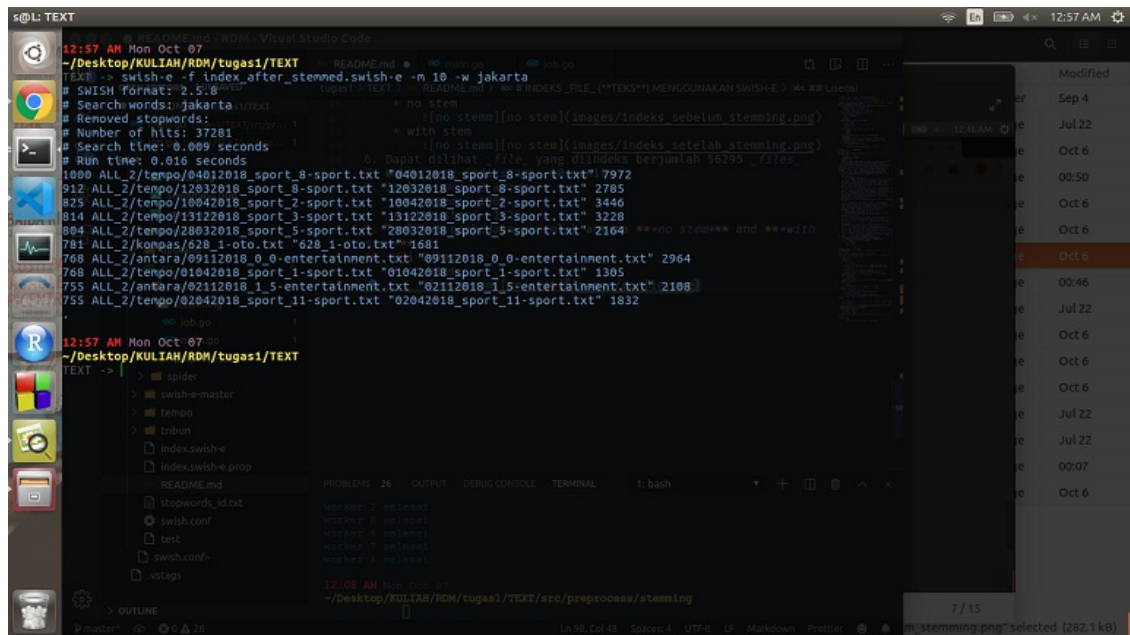
12:58 AM Mon Oct 07
~/Desktop/KULIAH/RDM/tugas1/TEXT
TEXT -> swish-e -f index.swish-e -m 10 -w jakarta
# SWISH format: 2.5.8
# Search words: jakarta
# Removed stopwords:
# Number of hits: 37282
# Search time: 0.007 seconds
# Run time: 0.016 seconds
1000 ALL/tempo/04012018_sport_8-sport.txt "04012018_sport_8-sport.txt" 8357
912 ALL/tempo/12032018_sport_8-sport.txt "12032018_sport_8-sport.txt" 2071
825 ALL/tempo/10042018_sport_2-sport.txt "10042018_sport_2-sport.txt" 3579
814 ALL/tempo/13122018_sport_3-sport.txt "13122018_sport_3-sport.txt" 3593
804 ALL/tempo/28032018_sport_5-sport.txt "28032018_sport_5-sport.txt" 2261
781 ALL/kompas/620_1-oto.txt "620_1-oto.txt" 1708
768 ALL/tempo/01042018_sport_1-sport.txt "01042018_sport_1-sport.txt" 1383
768 ALL/antara/09112018_0_0-entertainment.txt "09112018_0_0-entertainment.txt" 3291
755 ALL/antara/11032018_0_3-entertainment.txt "11032018_0_3-entertainment.txt" 3268
755 ALL/kompas/490_1-sport.txt "490_1-sport.txt" 3339

12:58 AM Mon Oct 07
~/Desktop/KULIAH/RDM/tugas1/TEXT
TEXT -> spider
# swish-e-master
# tempo
# tribun
# index.swish-e
# index.swish-e-prop
# README.md
# stopwords_id.txt
# swish.conf
# test
# swish.conf-
# vstags

PROBLEMS 26 OUTPUT DEBUG CONSOLE TERMINAL t: bash

```

- with stem



```
12:57 AM Mon Oct 07
~/Desktop/KULIAH/RDM/tugas1/TEXT
TEXT >> swish-e -f index_after_stemmed.swish-e -m 10 -w jakarta
# SWISH format: 2.5.8
# Search words: jakarta
# Removed stopwords:
# Number of hits: 37281
# Search time: 0.009 seconds
# Run times: 0.016 seconds
1000 ALL_2/tempo/04012018_sport_8-sport.txt "04012018_sport_8-sport.txt" 7972
912 ALL_2/tempo/12032018_sport_8-sport.txt "12032018_sport_8-sport.txt" 2785
825 ALL_2/tempo/10042018_sport_2-sport.txt "10042018_sport_2-sport.txt" 3446
814 ALL_2/tempo/13122018_sport_3-sport.txt "13122018_sport_3-sport.txt" 3228
804 ALL_2/tempo/28032018_sport_5-sport.txt "28032018_sport_5-sport.txt" 2164
781 ALL_2/kompas/628_1-oto.txt "628_1-oto.txt" 1681
768 ALL_2/antara/09112018_0_0-entertainment.txt "09112018_0_0-entertainment.txt" 2964
768 ALL_2/tempo/01042018_sport_1-sport.txt "01042018_sport_1-sport.txt" 1305
755 ALL_2/antara/02112018_1_5-entertainment.txt "02112018_1_5-entertainment.txt" 2108
755 ALL_2/tempo/02042018_sport_11-sport.txt "02042018_sport_11-sport.txt" 1832
```

11. Jika ingin mencari dokumen dengan jumlah kata lebih dari satu, maka gunakan tanda petik dua ("[query]"), contohnya `swish-e -f [index.swish-e] -m 10 -w "jakarta bandung surabaya"`. *Query* tersebut akan mencari dokumen yang mengandung kata **jakarta AND bandung AND surabaya**. Jika dalam *query* tersebut ada kata yang terdapat dalam daftar *stopword*, maka akan dihapus/dilewatkan/tidak di proses. Argument dari parameter `-f` adalah nama file hasil *indexing* (default: `index.swish-e`).
12. Arti dari Output (format):

- o Kolom-1: Ranking
- o Kolom-2: Lokasi dokumen/*file*
- o Kolom-3: Judul>Nama dokumen/*file*
- o Kolom-4: Size/Ukuran dokumen (dalam bytes)

[BACA LEBIH LANJUT \(https://www.esa.org/tiee/search/html/swish-config.html\)](https://www.esa.org/tiee/search/html/swish-config.html)

---Download---

Semua file yang digunakan dalam projek ini dapat di unduh pada link berikut

[UNDUH FILE \(https://drive.google.com/open?id=1e0Qtx1KgmzGm3IJGxHW3qBkp9F3BTbBb\)](https://drive.google.com/open?id=1e0Qtx1KgmzGm3IJGxHW3qBkp9F3BTbBb)