



# Kaldi 실습 환경 구축

# 목차

1. Docker 설치
2. Kaldi 환경 다운로드
3. 데이터 및 언어 모델 다운로드



# Docker 설치

# Docker 설치

## ❑ Docker toolbox 다운로드

- <https://github.com/docker/toolbox/releases> 에서 다운로드

v19.03.1

 guillaumerose released this on 1 Aug · 4 commits to master since this release

Please ensure that your system has all of the latest updates before attempting the installation. In some cases, this will require a reboot. If you run into issues creating VMs, you may need to uninstall VirtualBox before re-installing the Docker Toolbox.

The following list of components is included with this Toolbox release. If you have a previously installed version of Toolbox, these installers will update the components to these versions.

### Included Components

- docker 19.03.1
- docker-machine 0.16.1
- docker-compose 1.24.1
- Kitematic 0.17.7
- Boot2Docker ISO 19.03.1
- VirtualBox 5.2.20

### ▼ Assets 6

 <a href="#">DockerToolbox-19.03.1.exe</a>	다운로드	231 MB
 <a href="#">DockerToolbox-19.03.1.pkg</a>		235 MB
 <a href="#">md5sum.txt</a>		102 Bytes
 <a href="#">sha256sum.txt</a>		168 Bytes
 <a href="#">Source code (zip)</a>		
 <a href="#">Source code (tar.gz)</a>		

# Docker 설치

- ❑ Docker toolbox 설치
  - 다운로드 받은 파일 실행 후 설치 진행



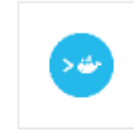


# Kaldi 환경 다운로드

# Kaldi 환경 다운로드

## ❑ Docker toolbox 실행

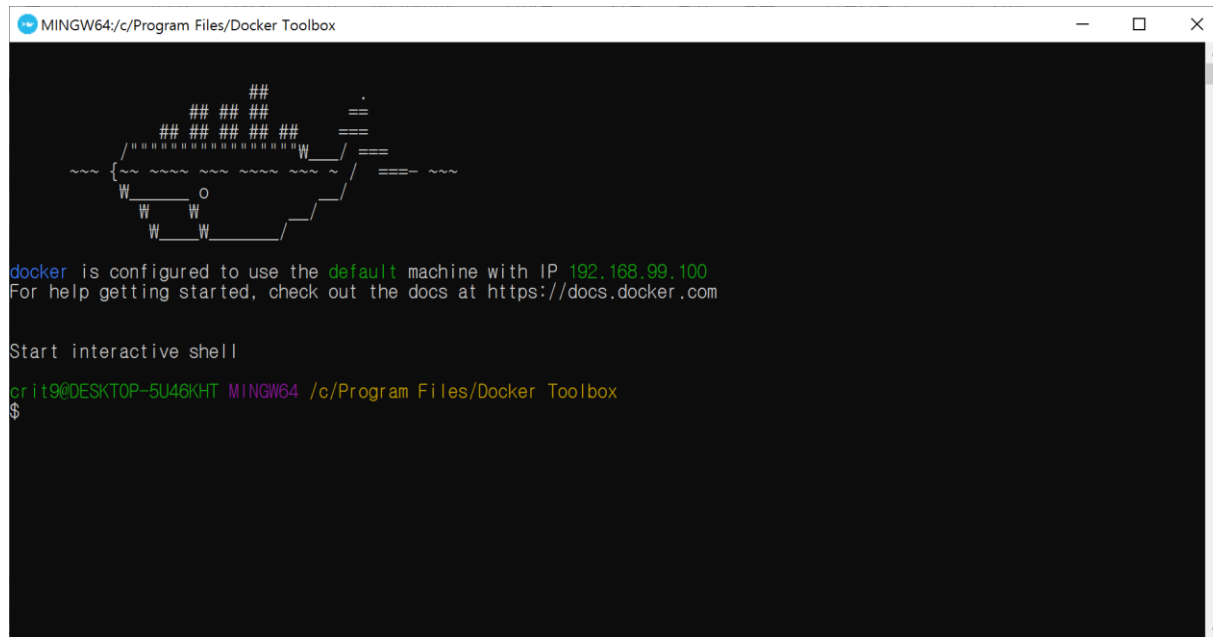
- Docker Quickstart Terminal 실행



Docker Quickstart Terminal

앱

- 실행 후 아래와 같은 창이 나올 때까지 약 1분 정도 대기

A screenshot of a terminal window titled "MINGW64/c/Program Files/Docker Toolbox". The terminal has a black background with white text. At the top, there is a large ASCII art cat face. Below the cat, the text reads: "docker is configured to use the default machine with IP 192.168.99.100", "For help getting started, check out the docs at https://docs.docker.com", and "Start interactive shell". At the bottom, the prompt "crit9@DESKTOP-5U46KHT MINGW64 /c/Program Files/Docker Toolbox" is followed by a dollar sign "\$".

```
MINGW64/c/Program Files/Docker Toolbox

      ##
    ## ##
  ## ## ## ##
{ ~~~~~ W ~~~~~
  W      O
  W      W
  W      W

docker is configured to use the default machine with IP 192.168.99.100
For help getting started, check out the docs at https://docs.docker.com

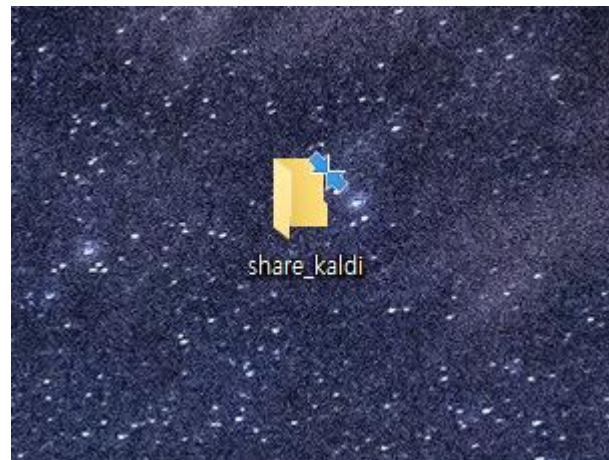
Start interactive shell

crit9@DESKTOP-5U46KHT MINGW64 /c/Program Files/Docker Toolbox
$
```



# Kaldi 환경 다운로드

- Docker 환경과 공유할 폴더 생성
  - 바탕화면에 share\_kaldi 폴더 생성



- 이 폴더를 통해 Docker 환경과 Host 환경 사이에 폴더 공유 가능



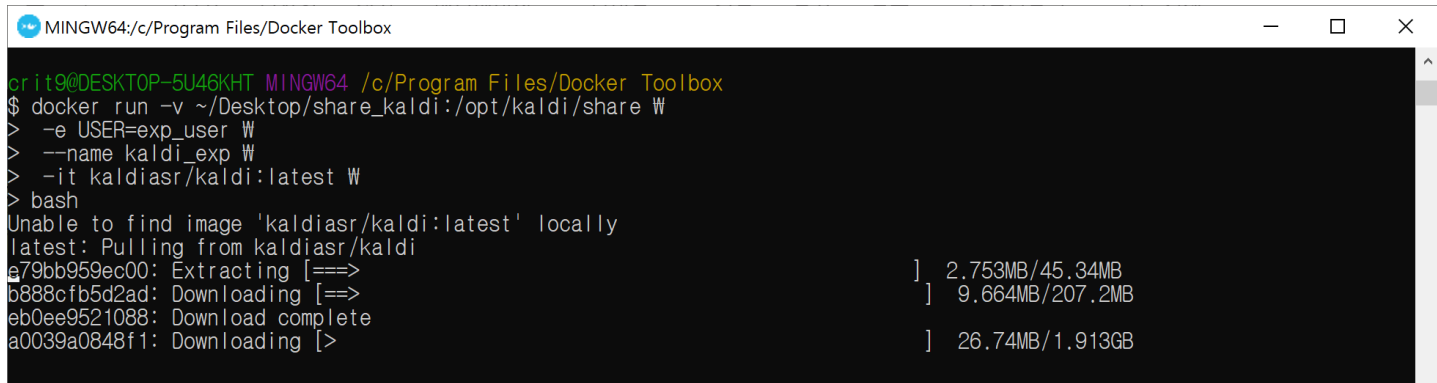
# Kaldi 환경 다운로드

## □ 공유 폴더 지정 및 kaldi 환경 다운로드

- Docker 커맨드창에 아래 명령어 입력 후 kaldi 환경 다운로드

```
docker run -v ~/Desktop/share_kaldi:/opt/kaldi/share \  
-e USER=exp_user \  
--name kaldi_exp \  
-it kaldiasr/kaldi:latest \  
bash
```

- 박스 안의 명령어를 복사한 후 커맨드창에 오른쪽 클릭하면 복사+붙여넣기 가능



```
crit9@DESKTOP-5U46KHT MINGW64 /c/Program Files/Docker Toolbox  
$ docker run -v ~/Desktop/share_kaldi:/opt/kaldi/share W  
> -e USER=exp_user W  
> --name kaldi_exp W  
> -it kaldiasr/kaldi:latest W  
> bash  
Unable to find image 'kaldiasr/kaldi:latest' locally  
latest: Pulling from kaldiasr/kaldi  
e79bb959ec00: Extracting [==>] 2.753MB/45.34MB  
b888cfb5d2ad: Downloading [==>] 9.664MB/207.2MB  
eb0ee9521088: Download complete  
a0039a0848f1: Downloading [>] 26.74MB/1.913GB
```

# Kaldi 환경 다운로드

## ❑ kaldi 환경 사용법

- 아래 그림과 같은 환경에서 Linux 명령어 사용가능
  - ls: 현재 디렉토리에 있는 파일 목록
  - cd [이동할 디렉토리]: 이동하고 싶은 디렉토리로 이동

현재 디렉토리

```
root@a50c15ffed3d: /opt/kaldi# ls
COPYING INSTALL README.md docker egs misc scripts share src tools windows
root@a50c15ffed3d:/opt/kaldi# cd misc
root@a50c15ffed3d:/opt/kaldi/misc# ls
README.txt htk_conversion htk_decode_example htk_graph_creation_example logo maintenance papers
root@a50c15ffed3d:/opt/kaldi/misc#
```

- Kaldi 환경 종료 명령어

exit

- Kaldi 환경 재진입 명령어

docker start kaldi\_exp && docker attach kaldi\_exp

# Kaldi 환경 다운로드

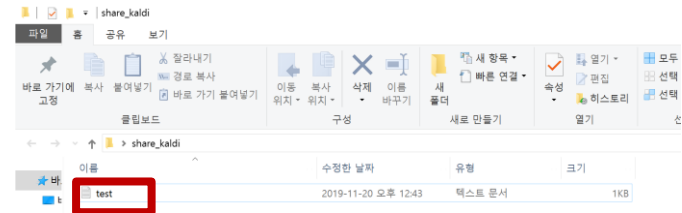
## □ 파일 공유 여부 확인

- /opt/kaldi/share 폴더와 바탕화면의 share\_kaldi 폴더 간 공유 가능
- 아래 명령어 실행 후 바탕화면의 share\_kaldi 폴더에 test.txt 파일이 있는지 확인

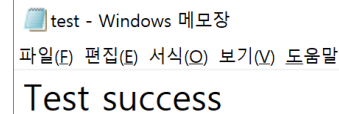
```
echo "Test success" > share/test.txt
```

```
root@a50c15ffed3d:/opt/kaldi# echo "Test success" > share/test.txt
root@a50c15ffed3d:/opt/kaldi# cat share/test.txt
Test success
root@a50c15ffed3d:/opt/kaldi#
```

kaldi 환경에서 입력



바탕화면에 생성한 share\_kaldi 폴더에 해당 파일이 있는지 확인



kaldi 환경에서 생성한 파일이 윈도우에서도 공유 가능

# Kaldi 환경 다운로드

- GPU 환경 원할 시 (Linux 환경에 GPU가 있을 경우)
  - 아래 명령어 실행 시 GPU 환경으로 이용가능

```
nvidia-docker run -v ~/Desktop/share_kaldi:/opt/kaldi/share \  
-e USER=exp_user \  
--name kaldi_exp \  
-it --runtime=nvidia kaldiasr/kaldi:gpu-latest \  
bash
```

# Kaldi 환경 다운로드

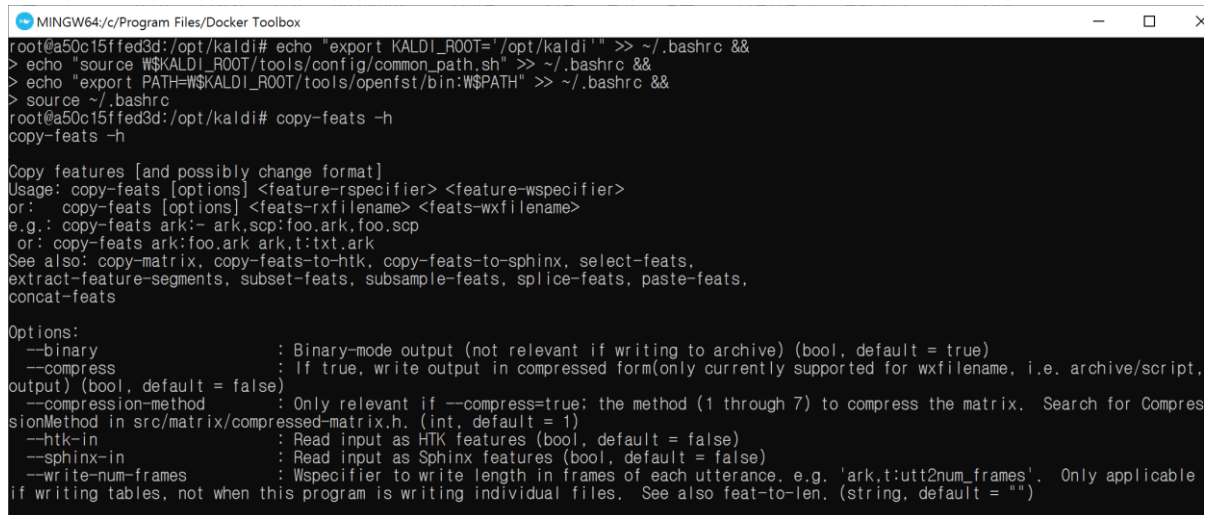
## □ Kaldi 명령어 path 지정

- Kaldi 명령어를 커맨드창에서 바로 사용할 수 있도록 환경 설정
- 아래 명령어를 실행

```
echo "export KALDI_ROOT='/opt/kaldi'" >> ~/.bashrc &&  
echo "source \"$KALDI_ROOT/tools/config/common_path.sh\" >> ~/.bashrc &&  
echo "export PATH=\"$KALDI_ROOT/tools/openfst/bin:$PATH\" >> ~/.bashrc &&  
source ~/.bashrc
```

- 아래 명령어 실행 후 그림과 같은 창이 나오는지 확인

```
copy-feats -h
```



```
MINGW64/c/Program Files/Docker Toolbox  
root@a50c15ffed3d:/opt/kaldi# echo "export KALDI_ROOT='/opt/kaldi'" >> ~/.bashrc &&  
> echo "source \"$KALDI_ROOT/tools/config/common_path.sh\" >> ~/.bashrc &&  
> echo "export PATH=\"$KALDI_ROOT/tools/openfst/bin:$PATH\" >> ~/.bashrc &&  
> source ~/.bashrc  
root@a50c15ffed3d:/opt/kaldi# copy-feats -h  
copy-feats -h  
  
Copy features [and possibly change format]  
Usage: copy-feats [options] <feature-rspecifier> <feature-wspecifier>  
or: copy-feats [options] <feats-rxfilename> <feats-wfilename>  
e.g.: copy-feats ark:- ark,scp:foo.ark,foo.scp  
or: copy-feats ark:foo.ark ark,t:txt.ark  
See also: copy-matrix, copy-feats-to-htk, copy-feats-to-sphinx, select-feats,  
extract-feature-segments, subset-feats, subsample-feats, splice-feats, paste-feats,  
concat-feats  
  
Options:  
  --binary                : Binary-mode output (not relevant if writing to archive) (bool, default = true)  
  --compress              : If true, write output in compressed form(only currently supported for wfilename, i.e. archive/script,  
output) (bool, default = false)  
  --compression-method    : Only relevant if --compress=true; the method (1 through 7) to compress the matrix. Search for Compress  
sionMethod in src/matrix/compressed-matrix.h. (int, default = 1)  
  --htk-in                : Read input as HTK features (bool, default = false)  
  --sphinx-in             : Read input as Sphinx features (bool, default = false)  
  --write-num-frames      : Wspecifier to write length in frames of each utterance. e.g. 'ark,t:utt2num_frames'. Only applicable  
if writing tables, not when this program is writing individual files. See also feat-to-len. (string, default = "")
```



# 데이터 및 언어 모델 다운로드

# 데이터 및 언어 모델 다운로드

## □ 아래 명령어 차례로 실행

- 실습 환경으로 이동

```
cd egs/mini_librispeech/s5
```

- 데이터 다운로드 받을 폴더 생성

```
mkdir -p corpus
```

- 학습 파일 다운로드

```
local/download_and_untar.sh ./corpus www.openslr.org/resources/31 train-clean-5
```

- 검증 파일 다운로드

```
local/download_and_untar.sh ./corpus www.openslr.org/resources/31 dev-clean-2
```

- 언어 모델 다운로드

```
local/download_lm.sh www.openslr.org/resources/11 ./corpus data/local/lm
```



# 데이터 및 언어 모델 다운로드

- ❑ Corpus 폴더 이동 후 아래창과 같이 나오면 다운로드 완료

```
cd corpus
```

```
root@a50c15ffed3d:/opt/kaldi/egs/mini_librispeech/s5# cd corpus/  
root@a50c15ffed3d:/opt/kaldi/egs/mini_librispeech/s5/corpus# ls  
3-gram.arpa.gz          3-gram.pruned.3e-7.arpa.gz  dev-clean-2.tar.gz      librispeech-vocab.txt  
3-gram.pruned.1e-7.arpa.gz  LibriSpeech                librispeech-lexicon.txt  train-clean-5.tar.gz  
root@a50c15ffed3d:/opt/kaldi/egs/mini_librispeech/s5/corpus#
```

```
cd LibriSpeech
```

```
root@a50c15ffed3d:/opt/kaldi/egs/mini_librispeech/s5/corpus/LibriSpeech# ls  
CHAPTERS.TXT  LICENSE.TXT  README.TXT  SPEAKERS.TXT  dev-clean-2  train-clean-5  
root@a50c15ffed3d:/opt/kaldi/egs/mini_librispeech/s5/corpus/LibriSpeech#
```

- ❑ 추가로 데이터 전처리에 필요한 라이브러리 다운로드

```
apt install flac
```