



※講義本編は別資料です

医療・創薬データサイエンスコンソーシアム  
入門編

# Linux入門 [ 補講 ]

神沼英里

東京医科歯科大学 医療データ科学推進室 特任講師  
(兼 理化学研究所 環境資源科学研究センター 客員研究員)  
(兼 国立遺伝学研究所 生命情報・DDBJセンター 外来研究員)



# ZOOM補講の進め方

## <Zoom補講の御案内>

- +講義はしませんので、[e-Learning映像を必ず先に視聴して下さい。](#)
- +本講義はe-Learning視聴終了時点で「講義参加済」として登録されます。
- +補講ではGoogle Colabを講師と一緒に動かします。

補講ですから、皆で会話しながら進めましょう。

+双方向の音声有り

(ハウリング防止のため発言以外はミュートでお願い致します。)

+途中でインタラプト有り

(どんどん発言してください。普通に進めば90分より相当早く終わるでしょう)。

+Zoom会議室には出入り自由、補講自体も受けなくてもOKです。

(補講では出席確認はありません)



# 深層学習でLinuxコマンドは必須

## ■Linuxコマンドは深層学習のどのステージで使うのか？

### 1.訓練Phase

\* データのダウンロード

\* データの前処理

( \* 学習済モデルのダウンロード )

\* モデルの訓練

### 2.評価Phase

\* モデルの評価

### 3.利用Phase

\* モデルの推論

wgetと unzip等の解凍コマンドは必須



# LinuxコマンドはGoogle検索で探そう

## ■Google検索でLinuxコマンドを探せばOK

Linux コマン

Linux コマン - Google 検索

linux コマンド

linux コマンド 一覧

linux コマンド コピー

linux コマンド ディレクトリ削除

linux コマンド 練習

linux コマンド grep

linux コマンド find

<http://www.redout.net/data/command.html>

## Linux コマンドリファレンス

[ ]は省略可能

...は複数の引数を指定可能

**man [オプション] [セクション]**  
オンラインマニュアル

**ls [オプション] [ファイル名...]**

ディレクトリの内容一覧表示

-l 垂直方向に1列で表示

-a ドットで始まる隠しファイルも表示

-C 最後にファイルのステータスを変更した時間でソートした時間になる。

-C 垂直方向に並べて表示

-d ディレクトリ自身を一覧表示

名前の後ろにファイル型を表示

通常のファイル

実行可能ファイル

-F ディレクトリ

シンボリックリンク



# Google Colabの使い方を覚えよう

Colab=ブラウザ上の機械学習プログラミング環境。Googleが無償提供。

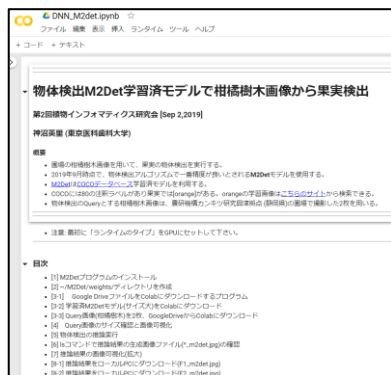
## ■今後のMDDSC実習でも使用

+ (瀬々先生、小寺先生、長谷先生)  
+ NGS解析入門  
+ 深層学習の基礎技術  
+ オープンデータ入門  
+ IoT時系列解析

## ■第3回MDDSC機械学習コンペや 応用編修了課題でも利用



## ■スマホブラウザから使える (デモに最適)



## ■実はR言語も使える！ (Python, Linuxだけではない)

### 2 Google ColaboratoryでR言語を使う

#### 2.1 Google Colaboratory について

Google Colaboratory (以下Colabと略称) は Google が提供している、ブラウザ上で機械学習などのプログラミング (言語は Python) が行える環境です。numpy や Tensorflow など機械学習によく使うライブラリがインストール済みであり、環境構築がほぼ不要で機械学習プログラミングが始められます。GPU を使うことも可能で、非力なパソコンでもディープラーニングができるのが学習の環境として非常に便利です。

Colab は基本的に Python 言語の環境ですが、R 言語を動かすことも可能です。R 本体 (RStudio) などをインストールしなくても R のプログラミングが始められるので、Colab は R の学習環境としても魅力的であると言えます。このページでは Colab で R を使う方法について解説します。

<https://htsuda.net/stats/colab.html>



# 本日補講の実習内容

	補講のステップ
1	投票1回目
2	スライド説明(補講用)
3	Classroomページへ移動
4	Colab起動
5	Linuxコマンドを実行(各セル)
6	Linuxコマンドを実行(全セル)
7	投票2回目
8	wget+unzip
9	質問回答(文献番号取得)
10	質問タイム



# 受講生Aさんからの質問

※難易度高（応用編レベル）  
の質問です

Linuxのご講義有難うございます。google colabの便利さを実感いたしました。

google colabで pubmed API より

- ・キーワードから配列情報を取得して、blastで解析する
- ・キーワード、著者名から文献情報を取得して、自然言語処理する

手順を教えていただきたいと思っています。

APIとXMLデータ処理ができますと、プログラミングが身近になり、モチベーションが上がると思いました。

受講生の皆様ともぜひ共有したいです。

受講生Aさんより

- 「キーワード」から配列情報を取得
- 「キーワード」、「著者名」から文献情報を取得

\* API=>ウェブからデータを取得する規格。オープンデータ入門で解説有り。

\* 自然言語処理する=>テキストマイニングを受講してください。



# キーワードからPubMed 文献情報を取得する

※難易度高（応用編レベル）  
の内容です

Colabで  
プログラムを  
実行

文献情報要求

文献情報応答

オープン  
データ  
入門で  
紹介

## APIs（米国NCBI）

NCBI provides several public APIs that allow programm

### Entrez Programming Utilities (E-utilities)

The E-utilities are the public API to the NCBI Entrez sys  
The E-utilities are a suite of eight server-side programs  
named Entrez Direct consists of several executables th

[Documentation](#) [Quick Start](#) [Examples](#) [Entrez](#)

### BLAST URL API

The BLAST API allows developers to submit BLAST se  
status of submitted searches and retrieve results when

[Overview](#) [Documentation](#)

今回はNCBI PubMedデータベースから「COVID-19」の  
文献番号を取得するAPIを実行してみよう