■報道実務家フォーラム：
データは整形が７割&付き合わせてなんぼ
【事前準備のお願い】

普段お使いのノートパソコンを持ってくるのが苦にならない方は、なるべくお持ち下さい。無謀かもしれないが、各自、ＰＣを実際に操作しながら、という進行を予定しています。ファイル名の末尾の「拡張子」(.txt とか.xlsx や.csv のたぐい）が表示される設定にしておいて下さい。ＰＣが得意なかたは、ぜひ、周囲のヘルプにご協力を。

表計算ソフトが入っていないかたは、あらかじめ Libre Office のインストールを（使うのは、そのうち Calc のみ)。正規表現が使えるテキストエディタが入っていないかたは、ATOM のインストールまたは、Regex101 の URL のブックマークを。

あと、入れておいて損はないので、Tabula と OpenRefine もインストールを。講義冒頭に簡単なデモを予定しています。

<Libre Office>
Excel など普段使いのものがあれば、それで十分。講師も Excel を使います。ＰＣに入っていない、という方のみ、下記の場所から「Libr Office」のインストールを。
https://ja.libreoffice.org/download/libreoffice-stable/
Excel は素のままでは正規表現は使えない。しかし、フィルター機能などを使って工夫次第でやれることはある。例えば、「長島」→「長嶋」の一括変換とか、表記の揺れを検出するとか。

<ATOM>
多機能なテキストエディタ（文書入力・編集のためのソフト)。「正規表現」を使って検索、置換ができる。コンピューターの基本文化として、あくまでヨコ方向（行）優位。お任せで文字化けしてしまったときは、「Shift+Ctrl+U」で文字コードを選び直すことが可能。欠点は動作が重たいこと。
https://atom.io/

<regex101>
正規表現の動作テストができるサイト。ATOM と同じ動きにするには、flavor を javascript に。optin は g(global)と m(multiline)を選ぶ。
https://regex101.com/

〈Tabula〉
PDF 形式のファイルを、表計算ソフトで読める形に変換する。必ず成功するわけではないが、かなり便利。
Tabula も OpenRefine も、自分のＰＣ上に作ったサーバーに、ブラウザを使って接続する、という仕組みで動く。ブラウザが起動するが、外部にデータを送っているわけではないので、そこは安心を。
http://tabula.technology/
下記の JRE が入っていないと動かない。

〈OpenRefine〉
データの下処理に威力を発揮。正規表現が使えて、タテ方向の一括変換も得意。とっつきにくいが、とても強力で役に立つ。Facet（自動フィルター）や名寄せ機能あり。
http://openrefine.org/download.html
これも下記の JRE が必要。

〈JRE〉
Tabula も OpenRefine も、Java Runtime Environment(JRE)というものがＰＣに入っていないと動作しないので、入っていないならインストールを。１度入ればＯＫ（Tabula と OpenRefine で同じものを使う）。ただし、会社貸与のＰＣでは、JRE のインストールに制限がかかっている場合もある。そのときは、Tabula と OpenRefine と JRE は「なし」で結構。
https://java.com/ja/download/

以上です