

Text Editor and Data Wrangling

A Basic CS Skill, ABC Winter School

박원

Objective

- Text Editor를 배울거예요
 - vim
 - VS Code
- Data Wrangling이 뭔지 배울거예요
 - pipe, regexp, grep
 - sort, uniq
 - head, tail

Vim

- 기본적으로 내장되어있는 텍스트 에디터입니다.
- 파일을 열어보거나, 수정할때 사용합니다.
 - 모든 텍스트 에디터의 역할이에요!
- 커스터마이징이 가능합니다.
- neo vim, space vim 같이 다양한 업그레이드(?) 버전이 있어요.

Mode of Vim

Key	Mode	Detail
Esc	Normal	
i	Insert	For inserting text
v	Visual	For select block of text
:	Command	For running command

Command of Vim

- `:q` -> quit
- `:w` -> save
- `:wq` -> save and quit
- `:e <file name>` -> open file
- `:/<string>` or `:?<string>` -> search
 - `n`: 다음 단어, `shift+n`: 이전 단어

VS Code

- google을 열어서 vs code를 검색하고, 설치할게요.
- 그리고 터미널을 열어서 `code` 명령어를 입력해보아요

VS Code Remote

Data Wrangling

Data Wrangling?

- 원래의 데이터를 보다 쉽게 접근하고 분석할 수 있도록 데이터를 정리, 통합하는 과정

Pipe

- 한 명령어의 출력을 다음 명령어나 셸에 전달해주는 기법
- `|`를 통해서 왼쪽 명령어의 표준출력을 오른쪽 명령어, 프로그램 또는 셸에 표준입력으로 전달한다.
- 중첩해서 사용이 가능하다.

```
$ echo ls -l | sh
```

Broken Pipe

- 파이프가 전달 도중에 실패하면, 끝까지 전달되지 않는다.
- 이 경우를 broken pipe라고 한다.

```
$ A | B | broken | D
```

- 여기서 broken까지만 실행되고 D는 실행이 안된다!

Regular Expression

- 문자열을 검색할 때, 특정 규칙을 만족하는 것을 찾고싶다.
- 해당 규칙들을 정규 표현식으로 나타낼 수 있어요!
- <https://regexr.com/>

Regex pattern

- `*` 바로 앞 문자열이나 식에서 0개 이상 반복되는 문자
- `.` 개행 문자가 아닌 문자 중 하나의 글자와 일치
- `+` 개행 문자가 아닌 문자 중 하나 이상의 글자와 일치
- `[...]` 표현하기 위한 문자
- `[a-z]` a~c 사이의 문자 하나
- `^` 는 줄의 시작, `$` 은 줄의 끝
- `(condition1|condition2)` condition1 조건 또는 condition2 조건

Grep

- 대상 파일이나 파이프에서 특정한 패턴의 문자열을 찾아준다
- 정규표현식을 사용...!

```
$ grep <pattern> <file>  
$ <program> | grep <pattern>
```

- Option
 - **-i** 대소문자 구분없이 찾게한다
 - **-n** 일치하는 줄과 그 줄번호를 같이 보여준다

sort

```
> cat some.txt
가나다라마바사
나다라마바사
가나다라마바사
나다라마바사
> cat some.txt | sort
가나다라마바사
가나다라마바사
나다라마바사
나다라마바사
> cat some.txt | sort -n
가나다라마바사
가나다라마바사
나다라마바사
나다라마바사
```

sort

- `-n` 숫자 순서대로 정렬한다
- `-r` 역순으로 정렬한다

uniq

```
> cat some.txt | uniq
가나다라마바사
나다라마바사
가나다라마바사
나다라마바사
> cat some.txt | sort | uniq -c
  2 가나다라마바사
  2 나다라마바사
```

- **-c** 중복된 횟수를 출력한다

head / tail

- head는 처음부터 10줄을 출력한다.
- tail은 마지막 10줄을 출력한다.
- `-n` 옵션으로 출력될 줄 수를 조정할 수 있다!

Questions?

Next