

# Ch5. 데이터의 이해: 데이터를 처리하라!

- summarized by 전승일(2014/10/12)
- source code: <http://www.headfirstlabs.com/books/hfpython/>
- support site: <http://python.itcarlow.ie>

## 켈리 감독에게 여러분의 도움이 필요합니다

---

- 선수들의 달리기 기록: james, julie, mikey, sarah
- 목적: 각 선수마다 상위 3개의 기록을 빨리 찾아내기
- 실습(data strip split): 선수들의 파일에서 데이터를 읽어서 각 선수별 리스트로 저장하고, 리스트를 화면에 출력하기

## 두 가지 정렬 방법

---

- 원본정렬(in-place sorting): 원래의 데이터를 정렬 (180쪽 그림 참조): 원본 데이터는 사라짐. 원본 데이터 복구 불가.

```
list.sort()
```

- 사본정렬(copied sorting): 원래의 데이터는 그대로 있고 정렬된 사본을 return

```
sorted(list)
```

- 실습(data sorted): 사본정렬로 정렬된 결과를 출력하기

## 시간 데이터가 가진 문제

---

- 데이터의 형태가 일정하지 않아 제대로 정렬되지 않음

- 문자열을 정렬할 때 하이픈이 점보다 앞에 올
- 실습(def sanitized): sanitize() 함수 만들기: 문자열에 있는 하이픈과 콜론을 점으로 바꾼 후에 반환
- 실습(cleaned data): sanitize() 함수를 사용하여 이전 프로그램 수정하기
- sorting 순서를 내림차순으로 하기 위해서는 *reverse=True* 인자 추가하면 됨

## 리스트의 지능화 (list comprehension)

- 하나의 리스트를 변형하여 다른 리스트를 만들 때 코드를 간결하게 하기

1. 변환된 데이터를 보관할 리스트를 새로 만들어야 합니다.
2. 원래 리스트의 모든 데이터 항목을 나열해야 합니다.
3. 나열하면서 각 데이터 항목을 변환해야 합니다.
4. 변환된 데이터를 새 리스트에 추가해야 합니다.

- 실습(list comprehension exercise): 연습하기
- 실습(list comprehension): list comprehension을 사용하여 이전 프로그램 간략화 하기
- 실습(list comprehension2): 좀 더 간략화 하기

## 중복된 데이터를 제거하기 위해 나열하기

- 중복 데이터 제거는 변환이 아니고 필터에 가깝기 때문에 list comprehension을 이용할 수 없음
- 실습(list comprehension unique): 리스트 나열코드를 사용하여 중복 데이터 제거하기, list slicing으로 앞의 3개만 출력하기

## 집합을 사용해서 중복된 데이터 제거하기

- set : 중복 데이터를 자동으로 제거해 주고, 순서가 없는 데이터 유형

```
set()
```

- Factory Function: A factory function is used to make new data items of a particular

type.

```
set() : set()
list() : []
str() : ''
float() : 0.0
int() : 0
dict() : {}
tuple() : ()
```

- 집합과 리스트의 비교: 203쪽 참고
- 실습(list comprehension with set): 집합을 사용하여 이전 프로그램 수정
- 실습(list comprehension perfect): 중복된 코드를 함수를 이용하여 간략화 하기, with open 사용 시 try/except로 보험 들기(예외 처리)

## Additional Information

---