
정렬 HOW TO

출시 버전 3.8.12

Guido van Rossum
and the Python development team

9월 07, 2021

Contents

1	정렬 기초	2
2	키 함수	2
3	<code>operator</code> 모듈 함수	3
4	오름차순과 내림차순	3
5	정렬 안정성과 복잡한 정렬	4
6	장식-정렬-복원을 사용하는 낡은 방법	4
7	<code>cmp</code> 매개 변수를 사용하는 낡은 방법	5
8	잡동사니	6

저자 Andrew Dalke와 Raymond Hettinger

배포 0.1

파이썬 리스트에는 리스트를 제자리에서 (in-place) 수정하는 내장 `list.sort()` 메서드가 있습니다. 또한, 이터러블로부터 새로운 정렬된 리스트를 만드는 `sorted()` 내장 함수가 있습니다.

이 문서에서는, 파이썬을 사용하여 데이터를 정렬하는 다양한 기술을 살펴봅니다.

1 정렬 기초

간단한 오름차순 정렬은 매우 쉽습니다; 그저 `sorted()` 함수를 호출하면 됩니다. 새로운 정렬된 리스트를 반환합니다:

```
>>> sorted([5, 2, 3, 1, 4])
[1, 2, 3, 4, 5]
```

`list.sort()` 메서드를 사용할 수도 있습니다. 리스트를 제자리에서 수정합니다(그리고 혼동을 피하고자 `None`을 반환합니다). 일반적으로 `sorted()` 보다 덜 편리합니다 - 하지만 원래 목록이 필요하지 않다면, 이것이 약간 더 효율적입니다.

```
>>> a = [5, 2, 3, 1, 4]
>>> a.sort()
>>> a
[1, 2, 3, 4, 5]
```

또 다른 점은 `list.sort()` 메서드가 리스트에게만 정의된다는 것입니다. 이와 달리, `sorted()` 함수는 모든 이터러블을 받아들입니다.

```
>>> sorted({1: 'D', 2: 'B', 3: 'B', 4: 'E', 5: 'A'})
[1, 2, 3, 4, 5]
```

2 키 함수

`list.sort()`와 `sorted()`는 모두 비교하기 전에 각 리스트 요소에 대해 호출할 함수를 지정하는 `key` 매개 변수를 가지고 있습니다.

예를 들어, 다음은 대소 문자를 구분하지 않는 문자열 비교입니다:

```
>>> sorted("This is a test string from Andrew".split(), key=str.lower)
['a', 'Andrew', 'from', 'is', 'string', 'test', 'This']
```

`key` 매개 변수의 값은 단일 인자를 취하고 정렬 목적으로 사용할 키를 반환하는 함수여야 합니다. 키 함수가 각 입력 레코드에 대해 정확히 한 번 호출되기 때문에 이 기법은 빠릅니다.

일반적인 패턴은 객체의 인덱스 중 일부를 키로 사용하여 복잡한 객체를 정렬하는 것입니다. 예를 들어:

```
>>> student_tuples = [
...     ('john', 'A', 15),
...     ('jane', 'B', 12),
...     ('dave', 'B', 10),
... ]
>>> sorted(student_tuples, key=lambda student: student[2])    # sort by age
[('dave', 'B', 10), ('jane', 'B', 12), ('john', 'A', 15)]
```

같은 기법이 이름있는 어트리뷰트를 갖는 객체에서도 작동합니다. 예를 들어:

```
>>> class Student:
...     def __init__(self, name, grade, age):
...         self.name = name
...         self.grade = grade
...         self.age = age
...     def __repr__(self):
...         return repr((self.name, self.grade, self.age))
```

```
>>> student_objects = [
...     Student('john', 'A', 15),
...     Student('jane', 'B', 12),
...     Student('dave', 'B', 10),
... ]
>>> sorted(student_objects, key=lambda student: student.age)    # sort by age
[('dave', 'B', 10), ('jane', 'B', 12), ('john', 'A', 15)]
```

3 operator 모듈 함수

위에서 보여준 키 함수 패턴은 매우 일반적이므로, 파이썬은 액세스 함수를 더 쉽고 빠르게 만드는 편리 함수를 제공합니다. operator 모듈에는 itemgetter(), attrgetter() 및 methodcaller() 함수가 있습니다. 이러한 함수를 사용하면, 위의 예제가 더 간단 해지고 빨라집니다:

```
>>> from operator import itemgetter, attrgetter
```

```
>>> sorted(student_tuples, key=itemgetter(2))
[('dave', 'B', 10), ('jane', 'B', 12), ('john', 'A', 15)]
```

```
>>> sorted(student_objects, key=attrgetter('age'))
[('dave', 'B', 10), ('jane', 'B', 12), ('john', 'A', 15)]
```

operator 모듈 함수는 다중 수준의 정렬을 허용합니다. 예를 들어, 먼저 *grade*로 정렬한 다음 *age*로 정렬하려면, 이렇게 합니다:

```
>>> sorted(student_tuples, key=itemgetter(1,2))
[('john', 'A', 15), ('dave', 'B', 10), ('jane', 'B', 12)]
```

```
>>> sorted(student_objects, key=attrgetter('grade', 'age'))
[('john', 'A', 15), ('dave', 'B', 10), ('jane', 'B', 12)]
```

4 오름차순과 내림차순

list.sort()와 sorted()는 모두 불리언 값을 갖는 *reverse* 매개 변수를 받아들입니다. 내림차순 정렬을 지정하는 데 사용됩니다. 예를 들어, 학생 데이터를 역 *age* 순서로 얻으려면, 이렇게 합니다:

```
>>> sorted(student_tuples, key=itemgetter(2), reverse=True)
[('john', 'A', 15), ('jane', 'B', 12), ('dave', 'B', 10)]
```

```
>>> sorted(student_objects, key=attrgetter('age'), reverse=True)
[('john', 'A', 15), ('jane', 'B', 12), ('dave', 'B', 10)]
```

5 정렬 안정성과 복잡한 정렬

정렬은 안정적임이 보장됩니다. 즉, 여러 레코드가 같은 키를 가질 때, 원래의 순서가 유지됩니다.

```
>>> data = [('red', 1), ('blue', 1), ('red', 2), ('blue', 2)]
>>> sorted(data, key=itemgetter(0))
[('blue', 1), ('blue', 2), ('red', 1), ('red', 2)]
```

*blue*에 대한 두 레코드가 원래 순서를 유지해서 ('blue', 1)이 ('blue', 2)보다 앞에 나옴이 보장됨에 유의하십시오.

이 멋진 속성은 일련의 정렬 단계로 복잡한 정렬을 만들 수 있도록 합니다. 예를 들어, 학생 데이터를 *grade*의 내림차순으로 정렬한 다음, *age*의 오름차순으로 정렬하려면, 먼저 *age* 정렬을 수행한 다음 *grade*를 사용하여 다시 정렬합니다:

```
>>> s = sorted(student_objects, key=attrgetter('age'))      # sort on secondary key
>>> sorted(s, key=attrgetter('grade'), reverse=True)        # now sort on primary key, ↵
↵descending
[('dave', 'B', 10), ('jane', 'B', 12), ('john', 'A', 15)]
```

이것은 다중 패스로 정렬하기 위해 필드와 순서의 튜플 리스트를 받을 수 있는 래퍼 함수로 추상화할 수 있습니다.

```
>>> def multisort(xs, specs):
...     for key, reverse in reversed(specs):
...         xs.sort(key=attrgetter(key), reverse=reverse)
...     return xs
```

```
>>> multisort(list(student_objects), (('grade', True), ('age', False)))
[('dave', 'B', 10), ('jane', 'B', 12), ('john', 'A', 15)]
```

파이썬에서 사용된 **Timsort** 알고리즘은 데이터 집합에 이미 존재하는 순서를 활용할 수 있으므로 효율적으로 여러 번의 정렬을 수행합니다.

6 장식-정렬-복원을 사용하는 낯은 방법

이 관용구는 그것의 세 단계를 따라 장식-정렬-복원 (Decorate-Sort-Undecorate) 이라고 합니다:

- 첫째, 초기 리스트가 정렬 순서를 제어하는 새로운 값으로 장식됩니다.
- 둘째, 장식된 리스트를 정렬합니다.
- 마지막으로, 장식을 제거하여, 새 순서로 초깃값만 포함하는 리스트를 만듭니다.

예를 들어, DSU 방식을 사용하여 *grade*로 학생 데이터를 정렬하려면 다음과 같이 합니다:

```
>>> decorated = [(student.grade, i, student) for i, student in enumerate(student_
↵objects)]
>>> decorated.sort()
>>> [student for grade, i, student in decorated]          # undecorate
[('john', 'A', 15), ('jane', 'B', 12), ('dave', 'B', 10)]
```

이 관용구는 튜플이 사전식으로 비교되기 때문에 작동합니다; 첫 번째 항목이 비교됩니다; 그들이 같으면 두 번째 항목이 비교되고, 이런 식으로 계속됩니다.

모든 경우에 장식된 리스트에 인덱스 *i*를 포함할 필요는 없지만, 두 가지 이점이 있습니다:

- 정렬이 안정적입니다 – 두 항목이 같은 키를 가지면, 그들의 순서가 정렬된 리스트에 유지됩니다.
- 장식된 튜플의 순서는 최대 처음 두 항목에 의해 결정되므로 원래 항목은 비교 가능할 필요가 없습니다. 그래서 예를 들어, 원래 리스트에는 직접 정렬될 수 없는 복소수가 포함될 수 있습니다.

이 관용구의 또 다른 이름은 펄 프로그래머들 사이에서 이것을 대중화한 Randal L. Schwartz의 이름을 딴 **Schwartzian 변환**입니다.

이제 파이썬 정렬이 키 함수를 제공하기 때문에, 이 기법은 자주 필요하지 않습니다.

7 cmp 매개 변수를 사용하는 낡은 방법

이 HOWTO에서 제공하는 많은 구문은 파이썬 2.4 이상을 가정합니다. 그전에는, `sorted()` 내장 함수가 없었으며 `list.sort()` 는 키워드 인자를 취하지 않았습니다. 대신, 모든 Py2.x 버전은 사용자 지정 비교 함수를 처리하기 위해 `cmp` 매개 변수를 지원했습니다.

Py3.0에서는, `cmp` 매개 변수가 완전히 제거되었습니다 (언어를 단순화하고 통합하기 위한 노력의 일환으로, 풍부한 비교 (rich comparisons)와 `__cmp__()` 매직 메서드 간의 충돌을 제거했습니다).

Py2.x에서, `sort` 는 비교 작업을 위해 호출할 수 있는 선택적 함수를 허용했습니다. 이 함수는 비교할 두 개의 인자를 취한 다음, 첫 번째가 두 번째보다 작으면 (less-than) 음수 값을 반환하고, 같으면 0을 반환하고, 크면 (greater-than) 양수 값을 반환해야 합니다. 예를 들어, 다음과 같이 할 수 있습니다:

```
>>> def numeric_compare(x, y):
...     return x - y
>>> sorted([5, 2, 4, 1, 3], cmp=numeric_compare)
[1, 2, 3, 4, 5]
```

또는 비교 순서를 다음과 같이 뒤집을 수 있습니다:

```
>>> def reverse_numeric(x, y):
...     return y - x
>>> sorted([5, 2, 4, 1, 3], cmp=reverse_numeric)
[5, 4, 3, 2, 1]
```

파이썬 2.x에서 3.x로 코드를 이식할 때, 사용자가 비교 함수를 제공하고 이를 키 함수로 변환해야 하는 상황이 발생할 수 있습니다. 다음 래퍼를 사용하면 쉽게 할 수 있습니다:

```
def cmp_to_key(mycmp):
    'Convert a cmp= function into a key= function'
    class K:
        def __init__(self, obj, *args):
            self.obj = obj
        def __lt__(self, other):
            return mycmp(self.obj, other.obj) < 0
        def __gt__(self, other):
            return mycmp(self.obj, other.obj) > 0
        def __eq__(self, other):
            return mycmp(self.obj, other.obj) == 0
        def __le__(self, other):
            return mycmp(self.obj, other.obj) <= 0
        def __ge__(self, other):
            return mycmp(self.obj, other.obj) >= 0
        def __ne__(self, other):
            return mycmp(self.obj, other.obj) != 0
    return K
```

키 함수로 변환하려면, 단지 이전 비교 함수를 감싸면 됩니다:

```
>>> sorted([5, 2, 4, 1, 3], key=cmp_to_key(reverse_numeric))
[5, 4, 3, 2, 1]
```

파이썬 3.2에서, `functools.cmp_to_key()` 함수가 표준 라이브러리의 `functools` 모듈에 추가되었습니다.

8 잡동사니

- 로케일 인식 정렬의 경우, 키 함수로는 `locale.strxfrm()` 를, 비교 함수로는 `locale.strcoll()` 을 사용하십시오.
- `reverse` 매개 변수는 여전히 정렬 안정성을 유지합니다 (그래서 같은 키를 갖는 레코드는 원래 순서를 유지합니다). 흥미롭게도, 그 효과는 내장 `reversed()` 함수를 두 번 사용하여 매개 변수 없이 흉내 낼 수 있습니다:

```
>>> data = [('red', 1), ('blue', 1), ('red', 2), ('blue', 2)]
>>> standard_way = sorted(data, key=itemgetter(0), reverse=True)
>>> double_reversed = list(reversed(sorted(reversed(data), key=itemgetter(0))))
>>> assert standard_way == double_reversed
>>> standard_way
[('red', 1), ('red', 2), ('blue', 1), ('blue', 2)]
```

- 정렬 루틴은 두 객체를 비교할 때 `__lt__()` 를 사용하도록 보장됩니다. 따라서 `__lt__()` 메서드를 정의하여, 표준 정렬 순서를 클래스에 추가하기는 쉽습니다:

```
>>> Student.__lt__ = lambda self, other: self.age < other.age
>>> sorted(student_objects)
[('dave', 'B', 10), ('jane', 'B', 12), ('john', 'A', 15)]
```

- 키 함수는 정렬되는 객체에 직접 의존할 필요가 없습니다. 키 함수는 외부 자원에 액세스할 수도 있습니다. 예를 들어, 학생 성적이 디렉터리에 저장되어 있다면, 학생 이름의 별도 리스트를 정렬하는 데 사용할 수 있습니다:

```
>>> students = ['dave', 'john', 'jane']
>>> newgrades = {'john': 'F', 'jane': 'A', 'dave': 'C'}
>>> sorted(students, key=newgrades.__getitem__)
['jane', 'dave', 'john']
```