



- » [HowTo](#)
- » [Sorting](#)
- » [HowTo/Sorting](#)
- » [FrontPage](#)
- » [RecentChanges](#)
- » [FindPage](#)
- » [HelpContents](#)
- » [HowTo/Sorting](#)

Page

- » Immutable Page
- » [Info](#)
- » [Attachments](#)
- »

More Actions:

User

- » [Login](#)

Sorting Mini-HOW TO

Original version by Andrew Dalke with a major update by Raymond Hettinger

Contents
<div><div>1. Sorting Mini-HOW TO</div><div><div>1. Sorting Basics</div><div>2. Key Functions</div><div>3. Operator Module Functions</div><div>4. Ascending and Descending</div><div>5. Sort Stability and Complex Sorts</div><div>6. The Old Way Using Decorate-Sort-Undecorate</div><div>7. The Old Way Using the cmp Parameter</div><div>8. Maintaining Sort Order</div><div>9. Odd and Ends</div></div></div>

Python lists have a built-in `sort()` method that modifies the list in-place and a `sorted()` built-in function that builds a new sorted list from an iterable.

There are many ways to use them to sort data and there doesn't appear to be a single, central place in the various manuals describing them, so I'll do so here.

Sorting Basics

A simple ascending sort is very easy -- just call the `sorted()` function. It returns a new sorted list:

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

You can also use the `list.sort()` method of a list. It modifies the list in-place (and returns `None` to avoid confusion). Usually it's less convenient than `sorted()` - but if you don't need the original list, it's slightly more efficient.

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

Another difference is that the `list.sort()` method is only defined for lists. In contrast, the `sorted()` function accepts any iterable.

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

Key Functions

Starting with Python 2.4, both `list.sort()` and `sorted()` added a `key` parameter to specify a function to be called on each list element prior to making comparisons.

For example, here's a case-insensitive string comparison:

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

The value of the `key` parameter should be a function that takes a single argument and returns a key to use for sorting purposes. This technique is fast because the key function is called exactly once for each input record.

A common pattern is to sort complex objects using some of the object's indices as a key. For example:


```
>>> 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)]
```

The same technique works for objects with named attributes. For example:

```
>>> 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))
    def weighted_grade(self):
        return 'CBA'.index(self.grade) / float(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)]
```

Operator Module Functions

The key-function patterns shown above are very common, so Python provides convenience functions to make accessor functions easier and faster. The  [operator module](#) has `itemgetter`, `attrgetter`, and starting in Python 2.6 a `methodcaller` function.

Using those functions, the above examples become simpler and faster.

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

>>> 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)]
```

The operator module functions allow multiple levels of sorting. For example, to sort by grade then by age:

```
>>> 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)]
```

The third function from the operator module, `methodcaller` is used in the following example in which the weighted grade of each student is shown before sorting on it:

```
>>> [(student.name, student.weighted_grade()) for student in student_objects]
[('john', 0.13333333333333333), ('jane', 0.08333333333333333), ('dave', 0.1)]
>>> sorted(student_objects, key=methodcaller('weighted_grade'))
[('jane', 'B', 12), ('dave', 'B', 10), ('john', 'A', 15)]
```

Ascending and Descending

Both `list.sort()` and `sorted()` accept a `reverse` parameter with a boolean value. This is used to flag descending sorts. For example, to get the student data in reverse age order:

```
>>> 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)]
```

Sort Stability and Complex Sorts

Starting with Python 2.2, sorts are guaranteed to be  [stable](#). That means that when multiple records have the same key, their original order is preserved.

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

Notice how the two records for 'blue' retain their original order so that ('blue', 1) is guaranteed to precede ('blue', 2).

This wonderful property lets you build complex sorts in a series of sorting steps. For example, to sort the student data by descending grade and then ascending age, do the age sort first and then sort again using grade:

```
>>> 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)]
```

The [!\[\]\(21199eb166cc97331a0c54c649195dcc_img.jpg\) Timsort](#) algorithm used in Python does multiple sorts efficiently because it can take advantage of any ordering already present in a dataset.

The Old Way Using Decorate-Sort-Undecorate

This idiom is called Decorate-Sort-Undecorate after its three steps:

- » First, the initial list is decorated with new values that control the sort order.
- » Second, the decorated list is sorted.
- » Finally, the decorations are removed, creating a list that contains only the initial values in the new order.

For example, to sort the student data by grade using the DSU approach:

```
>>> 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)]
```

This idiom works because tuples are compared lexicographically; the first items are compared; if they are the same then the second items are compared, and so on.

It is not strictly necessary in all cases to include the index `i` in the decorated list. Including it gives two benefits:

- » The sort is stable - if two items have the same key, their order will be preserved in the sorted list.
- » The original items do not have to be comparable because the ordering of the decorated tuples will be determined by at most the first two items. So for example the original list could contain complex numbers which cannot be sorted directly.

Another name for this idiom is [!\[\]\(dd161862f9164df98f62b726e9846241_img.jpg\) Schwartzian transform](#), after Randal L. Schwartz, who popularized it among Perl programmers.

For large lists and lists where the comparison information is expensive to calculate, and Python versions before 2.4, DSU is likely to be the fastest way to sort the list. For 2.4 and later, key functions provide the same functionality.

The Old Way Using the cmp Parameter

Many constructs given in this HOWTO assume Python 2.4 or later. Before that, there was no `sorted()` builtin and `list.sort()` took no keyword arguments. Instead, all of the Py2.x versions supported a `cmp` parameter to handle user specified comparison functions.

In Py3.0, the `cmp` parameter was removed entirely (as part of a larger effort to simplify and unify the language, eliminating the conflict between rich comparisons and the `__cmp__` methods).

In Py2.x, `sort` allowed an optional function which can be called for doing the comparisons. That function should take two arguments to be compared and then return a negative value for less-than, return zero if they are equal, or return a positive value for greater-than. For example, we can do:

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

Or you can reverse the order of comparison with:

```
>>> def reverse_numeric(x, y):
    return y - x
```

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

When porting code from Python 2.x to 3.x, the situation can arise when you have the user supplying a comparison function and you need to convert that to a key function. The following wrapper makes that easy to do:

```
def cmp_to_key(mycmp):
    'Convert a cmp= function into a key= function'
    class K(object):
        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
```

To convert to a key function, just wrap the old comparison function:

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

In Python 2.7, the `cmp_to_key()` tool was added to the `functools` module.

Maintaining Sort Order

Python does not provide modules like C++'s `set` and `map` data types as part of its standard library. This is a conscious decision on the part of Guido, et al to preserve "one obvious way to do it." Instead Python delegates this task to third-party libraries that are available on the [Python Package Index](#). These libraries use various techniques to maintain list, dict, and set types in sorted order. Maintaining order using a specialized data structure can avoid very slow behavior (quadratic run-time) in the naive approach of editing and constantly re-sorting. Several implementations are described here.

- » [Python SortedContainers Module](#) - Pure-Python implementation that is fast-as-C implementations. Implements sorted list, dict, and set data types. Testing includes 100% code coverage and hours of stress. Documentation includes full API reference, [performance comparison](#), and contributing/development guidelines.
- » [Python rbtree Module](#) - Provides a fast, C-implementation for dict and set data types. Based on a red-black tree implementation.
- » [Python treap Module](#) - Provides a sorted dict data type. Uses a treap for implementation and improves performance using Cython.
- » [Python bintrees Module](#) - Provides several tree-based implementations for dict and set data types. Fastest implementations are based on AVL and Red-Black trees. Implemented in C. Extends the conventional API to provide set operations for dict data types.
- » [Python banyan Module](#) - Provides a fast, C-implementation for dict and set data types.
- » [Python skiplistcollections Module](#) - Pure-Python implementation based on skip-lists providing a limited API for dict and set data types.
- » [Python blist Module](#) - Provides sorted list, dict and set data types based on the "blist" data type, a B-tree implementation. Implemented in Python and C.

Odd and Ends

- » For locale aware sorting, use `locale.strxfrm()` for a key function or `locale.strcoll()` for a comparison function.
- » The `reverse` parameter still maintains sort stability (i.e. records with equal keys retain the original order). Interestingly, that effect can be simulated without the parameter by using the builtin `reversed` function twice:

```
» >>> data = [('red', 1), ('blue', 1), ('red', 2), ('blue', 2)]
>>> assert sorted(data, reverse=True) == list(reversed(sorted(reversed(data))))
```

» To create a standard sort order for a class, just add the appropriate rich comparison methods:

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

For general purpose comparisons, the recommended approach is to define all six rich comparison operators. The `functools.total_ordering` class decorator makes this easy to implement.

» Key functions need not access data internal to objects being sorted. A key function can also access external resources. For instance, if the student grades are stored in a dictionary, they can be used to sort a separate list of student names:

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

HowTo/Sorting (last edited 2014-10-12 06:26:39 by [Paddy3118](#))

» [MoinMoin Powered](#)

» [Python Powered](#)

» [GPL licensed](#)

» [Valid HTML 4.01](#)

[Unable to edit the page? See the FrontPage for instructions.](#)