For more information on the str class and its methods, see Text Sequence Type — str and the String Methods section below. To output formatted strings, see the Formatted string literals and Format String Syntax sections. In addition, see the Text Processing Services section.

String Methods

Strings implement all of the common sequence operations, along with the additional methods described below.

Strings also support two styles of string formatting, one providing a large degree of flexibility and customization (see str.format(), Format String Syntax and Custom String Formatting) and the other based on C printf style formatting that handles a narrower range of types and is slightly harder to use correctly, but is often faster for the cases it can handle (printf-style String Formatting).

The Text Processing Services section of the standard library covers a number of other modules that provide various text related utilities (including regular expression support in the re module).

str.capitalize()

Return a copy of the string with its first character capitalized and the rest lowercased.

Changed in version 3.8: The first character is now put into titlecase rather than uppercase. This means that characters like digraphs will only have their first letter capitalized, instead of the full character.

str. casefold()

Return a casefolded copy of the string. Casefolded strings may be used for caseless matching.

Casefolding is similar to lowercasing but more aggressive because it is intended to remove all case distinctions in a string. For example, the German lowercase letter 'ß' is equivalent to "ss". Since it is already lowercase, lower() would do nothing to 'ß'; casefold() converts it to "ss".

The casefolding algorithm is described in section 3.13 of the Unicode Standard.

New in version 3.3.

str. center(width[, fillchar])

Return centered in a string of length *width*. Padding is done using the specified *fillchar* (default is an ASCII space). The original string is returned if *width* is less than or equal to len(s).

str. count(sub[, start[, end]])

Return the number of non-overlapping occurrences of substring sub in the range [start,

end]. Optional arguments start and end are interpreted as in slice notation.

str.encode(encoding="utf-8", errors="strict")

Return an encoded version of the string as a bytes object. Default encoding is 'utf-8'. errors may be given to set a different error handling scheme. The default for errors is 'strict', meaning that encoding errors raise a UnicodeError. Other 'ignore', 'replace', 'xmlcharrefreplace', possible values are 'backslashreplace' and via any other name registered codecs.register error(), see section Error Handlers. For a list of possible encodings, see section Standard Encodings.

By default, the *errors* argument is not checked for best performances, but only used at the first encoding error. Enable the Python Development Mode, or use a debug build to check *errors*.

Changed in version 3.1: Support for keyword arguments added.

Changed in version 3.9: The errors is now checked in development mode and in debug mode.

str.endswith(suffix[, start[, end]])

Return True if the string ends with the specified *suffix*, otherwise return False. *suffix* can also be a tuple of suffixes to look for. With optional *start*, test beginning at that position. With optional *end*, stop comparing at that position.

str. **expandtabs**(tabsize=8)

Return a copy of the string where all tab characters are replaced by one or more spaces, depending on the current column and the given tab size. Tab positions occur every *tabsize* characters (default is 8, giving tab positions at columns 0, 8, 16 and so on). To expand the string, the current column is set to zero and the string is examined character by character. If the character is a tab (\t), one or more space characters are inserted in the result until the current column is equal to the next tab position. (The tab character itself is not copied.) If the character is a newline (\n) or return (\r), it is copied and the current column is reset to zero. Any other character is copied unchanged and the current column is incremented by one regardless of how the character is represented when printed.

str. **find**(sub[, start[, end]])

Return the lowest index in the string where substring *sub* is found within the slice s[start:end]. Optional arguments *start* and *end* are interpreted as in slice notation. Return -1 if *sub* is not found.

```
Note: The find() method should be used only if you need to know the position of sub. To check if sub is a substring or not, use the in operator:

>>> 'Py' in 'Python'
True
```

str. format(*args, **kwargs)

Perform a string formatting operation. The string on which this method is called can contain literal text or replacement fields delimited by braces {}. Each replacement field contains either the numeric index of a positional argument, or the name of a keyword argument. Returns a copy of the string where each replacement field is replaced with the string value of the corresponding argument.

```
>>> "The sum of 1 + 2 is {0}".format(1+2)
'The sum of 1 + 2 is 3'
```

See Format String Syntax for a description of the various formatting options that can be specified in format strings.

Note: When formatting a number (int, float, complex, decimal.Decimal and subclasses) with the n type (ex: '{:n}'.format(1234)), the function temporarily sets the LC_CTYPE locale to the LC_NUMERIC locale to decode decimal_point and thousands_sep fields of localeconv() if they are non-ASCII or longer than 1 byte, and the LC_NUMERIC locale is different than the LC_CTYPE locale. This temporary change affects other threads.

Changed in version 3.7: When formatting a number with the n type, the function sets temporarily the LC_CTYPE locale to the LC_NUMERIC locale in some cases.

str. format_map(mapping)

Similar to str.format(**mapping), except that mapping is used directly and not copied to a dict. This is useful if for example mapping is a dict subclass:

```
>>> class Default(dict):
...     def __missing__(self, key):
...         return key
...
>>> '{name} was born in {country}'.format_map(Default(name='Guido'')
'Guido was born in country'
```

New in version 3.2.

str.index(sub[, start[, end]])

Like find(), but raise ValueError when the substring is not found.

str. **isalnum**()

Return True if all characters in the string are alphanumeric and there is at least one

character, False otherwise. A character c is alphanumeric if one of the following returns True: c.isalpha(), c.isdecimal(), c.isdigit(), or c.isnumeric().

str. isalpha()

Return True if all characters in the string are alphabetic and there is at least one character, False otherwise. Alphabetic characters are those characters defined in the Unicode character database as "Letter", i.e., those with general category property being one of "Lm", "Lt", "Lu", "Ll", or "Lo". Note that this is different from the "Alphabetic" property defined in the Unicode Standard.

str. isascii()

Return True if the string is empty or all characters in the string are ASCII, False otherwise. ASCII characters have code points in the range U+0000-U+007F.

New in version 3.7.

str. isdecimal()

Return True if all characters in the string are decimal characters and there is at least one character, False otherwise. Decimal characters are those that can be used to form numbers in base 10, e.g. U+0660, ARABIC-INDIC DIGIT ZERO. Formally a decimal character is a character in the Unicode General Category "Nd".

str. **isdigit**()

Return True if all characters in the string are digits and there is at least one character, False otherwise. Digits include decimal characters and digits that need special handling, such as the compatibility superscript digits. This covers digits which cannot be used to form numbers in base 10, like the Kharosthi numbers. Formally, a digit is a character that has the property value Numeric_Type=Digit or Numeric_Type=Decimal.

str.isidentifier()

Return True if the string is a valid identifier according to the language definition, section Identifiers and keywords.

Call keyword.iskeyword() to test whether string s is a reserved identifier, such as def and class.

Example:

```
>>> from keyword import iskeyword
>>> 'hello'.isidentifier(), iskeyword('hello')
True, False
>>> 'def'.isidentifier(), iskeyword('def')
True, True
```

str. **islower**()

Return True if all cased characters [4] in the string are lowercase and there is at least

one cased character, False otherwise.

str.isnumeric()

Return True if all characters in the string are numeric characters, and there is at least one character, False otherwise. Numeric characters include digit characters, and all characters that have the Unicode numeric value property, e.g. U+2155, VULGAR FRACTION ONE FIFTH. Formally, numeric characters are those with the property value Numeric_Type=Digit, Numeric_Type=Decimal or Numeric_Type=Numeric.

str.isprintable()

Return True if all characters in the string are printable or the string is empty, False otherwise. Nonprintable characters are those characters defined in the Unicode character database as "Other" or "Separator", excepting the ASCII space (0x20) which is considered printable. (Note that printable characters in this context are those which should not be escaped when repr() is invoked on a string. It has no bearing on the handling of strings written to sys.stdout or sys.stderr.)

str. isspace()

Return True if there are only whitespace characters in the string and there is at least one character, False otherwise.

A character is *whitespace* if in the Unicode character database (see unicodedata), either its general category is Zs ("Separator, space"), or its bidirectional class is one of WS, B, or S.

str. **istitle**()

Return True if the string is a titlecased string and there is at least one character, for example uppercase characters may only follow uncased characters and lowercase characters only cased ones. Return False otherwise.

str.**isupper**()

Return True if all cased characters [4] in the string are uppercase and there is at least one cased character, False otherwise.

```
>>> 'BANANA'.isupper()
True
>>> 'banana'.isupper()
False
>>> 'baNana'.isupper()
False
>>> ' '.isupper()
False
```

str. join(iterable)

Return a string which is the concatenation of the strings in *iterable*. A TypeError will be raised if there are any non-string values in *iterable*, including bytes objects. The separator between elements is the string providing this method.

str.ljust(width[, fillchar])

Return the string left justified in a string of length *width*. Padding is done using the specified *fillchar* (default is an ASCII space). The original string is returned if *width* is less than or equal to len(s).

str. lower()

Return a copy of the string with all the cased characters [4] converted to lowercase.

The lowercasing algorithm used is described in section 3.13 of the Unicode Standard.

str.lstrip([chars])

Return a copy of the string with leading characters removed. The *chars* argument is a string specifying the set of characters to be removed. If omitted or None, the *chars* argument defaults to removing whitespace. The *chars* argument is not a prefix; rather, all combinations of its values are stripped:

```
>>> ' spacious '.lstrip()
'spacious '
>>> 'www.example.com'.lstrip('cmowz.')
'example.com'
```

See str.removeprefix() for a method that will remove a single prefix string rather than all of a set of characters. For example:

```
>>> 'Arthur: three!'.lstrip('Arthur: ')
'ee!'
>>> 'Arthur: three!'.removeprefix('Arthur: ')
'three!'
```

static str. maketrans(x[, y[, z]])

This static method returns a translation table usable for str.translate().

If there is only one argument, it must be a dictionary mapping Unicode ordinals (integers) or characters (strings of length 1) to Unicode ordinals, strings (of arbitrary lengths) or None. Character keys will then be converted to ordinals.

If there are two arguments, they must be strings of equal length, and in the resulting dictionary, each character in x will be mapped to the character at the same position in y. If there is a third argument, it must be a string, whose characters will be mapped to None in the result.

str.partition(sep)

Split the string at the first occurrence of *sep*, and return a 3-tuple containing the part before the separator, the separator itself, and the part after the separator. If the separator is not found, return a 3-tuple containing the string itself, followed by two empty strings.

str. removeprefix(prefix,/)

If the string starts with the *prefix* string, return string[len(prefix):]. Otherwise, return a copy of the original string:

```
>>> 'TestHook'.removeprefix('Test')
'Hook'
>>> 'BaseTestCase'.removeprefix('Test')
'BaseTestCase'
```

New in version 3.9.

str. removesuffix(suffix,/)

If the string ends with the *suffix* string and that *suffix* is not empty, return string[:-len(suffix)]. Otherwise, return a copy of the original string:

```
>>> 'MiscTests'.removesuffix('Tests')
'Misc'
>>> 'TmpDirMixin'.removesuffix('Tests')
'TmpDirMixin'
```

New in version 3.9.

str. replace(old, new[, count])

Return a copy of the string with all occurrences of substring *old* replaced by *new*. If the optional argument *count* is given, only the first *count* occurrences are replaced.

str. rfind(sub[, start[, end]])

Return the highest index in the string where substring *sub* is found, such that *sub* is contained within s[start:end]. Optional arguments *start* and *end* are interpreted as in slice notation. Return -1 on failure.

str. rindex(sub[, start[, end]])

Like rfind() but raises ValueError when the substring *sub* is not found.

str. rjust(width[, fillchar])

Return the string right justified in a string of length *width*. Padding is done using the specified *fillchar* (default is an ASCII space). The original string is returned if *width* is less than or equal to len(s).

str. rpartition(sep)

Split the string at the last occurrence of *sep*, and return a 3-tuple containing the part before the separator, the separator itself, and the part after the separator. If the separator is not found, return a 3-tuple containing two empty strings, followed by the string itself.

str. rsplit(sep=None, maxsplit=-1)

Return a list of the words in the string, using *sep* as the delimiter string. If *maxsplit* is given, at most *maxsplit* splits are done, the *rightmost* ones. If *sep* is not specified or

None, any whitespace string is a separator. Except for splitting from the right, rsplit() behaves like split() which is described in detail below.

str. rstrip([chars])

Return a copy of the string with trailing characters removed. The *chars* argument is a string specifying the set of characters to be removed. If omitted or None, the *chars* argument defaults to removing whitespace. The *chars* argument is not a suffix; rather, all combinations of its values are stripped:

```
>>> ' spacious '.rstrip()
' spacious'
>>> 'mississippi'.rstrip('ipz')
'mississ'
```

See str.removesuffix() for a method that will remove a single suffix string rather than all of a set of characters. For example:

```
>>> 'Monty Python'.rstrip(' Python')
'M'
>>> 'Monty Python'.removesuffix(' Python')
'Monty'
```

str. split(sep=None, maxsplit=-1)

Return a list of the words in the string, using *sep* as the delimiter string. If *maxsplit* is given, at most *maxsplit* splits are done (thus, the list will have at most maxsplit+1 elements). If *maxsplit* is not specified or -1, then there is no limit on the number of splits (all possible splits are made).

If sep is given, consecutive delimiters are not grouped together and are deemed to delimit empty strings (for example, '1,,2'.split(',') returns ['1', '', '2']). The sep argument may consist of multiple characters (for example, '1<>2<>3'.split('<>') returns ['1', '2', '3']). Splitting an empty string with a specified separator returns [''].

For example:

```
>>> '1,2,3'.split(',')
['1', '2', '3']
>>> '1,2,3'.split(',', maxsplit=1)
['1', '2,3']
>>> '1,2,,3,'.split(',')
['1', '2', '', '3', '']
```

If *sep* is not specified or is None, a different splitting algorithm is applied: runs of consecutive whitespace are regarded as a single separator, and the result will contain no empty strings at the start or end if the string has leading or trailing whitespace. Consequently, splitting an empty string or a string consisting of just whitespace with a None separator returns [].

For example:

```
>>> '1 2 3'.split()
['1', '2', '3']
>>> '1 2 3'.split(maxsplit=1)
['1', '2 3']
>>> ' 1 2 3 '.split()
['1', '2', '3']
```

str. splitlines([keepends])

Return a list of the lines in the string, breaking at line boundaries. Line breaks are not included in the resulting list unless *keepends* is given and true.

This method splits on the following line boundaries. In particular, the boundaries are a superset of universal newlines.

Representation	Description
\n	Line Feed
\r	Carriage Return
\r\n	Carriage Return + Line Feed
\v or \x0b	Line Tabulation
\f or \x0c	Form Feed
\x1c	File Separator
\x1d	Group Separator
\x1e	Record Separator
\x85	Next Line (C1 Control Code)
\u2028	Line Separator
\u2029	Paragraph Separator

Changed in version 3.2: \v and \f added to list of line boundaries.

For example:

```
>>> 'ab c\n\nde fg\rkl\r\n'.splitlines()
['ab c', '', 'de fg', 'kl']
>>> 'ab c\n\nde fg\rkl\r\n'.splitlines(keepends=True)
['ab c\n', '\n', 'de fg\r', 'kl\r\n']
```

Unlike split() when a delimiter string *sep* is given, this method returns an empty list for the empty string, and a terminal line break does not result in an extra line:

```
>>> "".splitlines()
[]
```

```
>>> "One line\n".splitlines()
['One line']
```

For comparison, split('\n') gives:

```
>>> ''.split('\n')
['']
>>> 'Two lines\n'.split('\n')
['Two lines', '']
```

str. startswith(prefix[, start[, end]])

Return True if string starts with the *prefix*, otherwise return False. *prefix* can also be a tuple of prefixes to look for. With optional *start*, test string beginning at that position. With optional *end*, stop comparing string at that position.

str. **strip**([chars])

Return a copy of the string with the leading and trailing characters removed. The *chars* argument is a string specifying the set of characters to be removed. If omitted or None, the *chars* argument defaults to removing whitespace. The *chars* argument is not a prefix or suffix; rather, all combinations of its values are stripped:

```
>>> ' spacious '.strip()
'spacious'
>>> 'www.example.com'.strip('cmowz.')
'example'
```

The outermost leading and trailing *chars* argument values are stripped from the string. Characters are removed from the leading end until reaching a string character that is not contained in the set of characters in *chars*. A similar action takes place on the trailing end. For example:

```
>>> comment_string = '#..... Section 3.2.1 Issue #32 .....'
>>> comment_string.strip('.#! ')
'Section 3.2.1 Issue #32'
```

str. swapcase()

Return a copy of the string with uppercase characters converted to lowercase and vice versa. Note that it is not necessarily true that s.swapcase().swapcase() == s.

str.**title**()

Return a titlecased version of the string where words start with an uppercase character and the remaining characters are lowercase.

For example:

```
>>> 'Hello world'.title()
'Hello World'
```

The algorithm uses a simple language-independent definition of a word as groups of consecutive letters. The definition works in many contexts but it means that apostrophes in contractions and possessives form word boundaries, which may not be the desired result:

```
>>> "they're bill's friends from the UK".title()
"They'Re Bill'S Friends From The Uk"
```

A workaround for apostrophes can be constructed using regular expressions:

str. translate(table)

Return a copy of the string in which each character has been mapped through the given translation table. The table must be an object that implements indexing via __getitem__(), typically a mapping or sequence. When indexed by a Unicode ordinal (an integer), the table object can do any of the following: return a Unicode ordinal or a string, to map the character to one or more other characters; return None, to delete the character from the return string; or raise a LookupError exception, to map the character to itself.

You can use str.maketrans() to create a translation map from character-tocharacter mappings in different formats.

See also the codecs module for a more flexible approach to custom character mappings.

str.upper()

Return a copy of the string with all the cased characters [4] converted to uppercase. Note that s.upper().isupper() might be False if s contains uncased characters or if the Unicode category of the resulting character(s) is not "Lu" (Letter, uppercase), but e.g. "Lt" (Letter, titlecase).

The uppercasing algorithm used is described in section 3.13 of the Unicode Standard.

str.zfill(width)

Return a copy of the string left filled with ASCII '0' digits to make a string of length width. A leading sign prefix ('+'/'-') is handled by inserting the padding after the sign character rather than before. The original string is returned if width is less than or equal to len(s).

For example:

```
>>> "42".zfill(5)
'00042'
>>> "-42".zfill(5)
'-0042'
```

printf-style String Formatting

Note: The formatting operations described here exhibit a variety of quirks that lead to a number of common errors (such as failing to display tuples and dictionaries correctly). Using the newer formatted string literals, the str.format() interface, or template strings may help avoid these errors. Each of these alternatives provides their own tradeoffs and benefits of simplicity, flexibility, and/or extensibility.

String objects have one unique built-in operation: the % operator (modulo). This is also known as the string *formatting* or *interpolation* operator. Given format % values (where *format* is a string), % conversion specifications in *format* are replaced with zero or more elements of *values*. The effect is similar to using the sprintf() in the C language.

If *format* requires a single argument, *values* may be a single non-tuple object. [5] Otherwise, *values* must be a tuple with exactly the number of items specified by the format string, or a single mapping object (for example, a dictionary).

A conversion specifier contains two or more characters and has the following components, which must occur in this order:

- 1. The '%' character, which marks the start of the specifier.
- 2. Mapping key (optional), consisting of a parenthesised sequence of characters (for example, (somename)).
- 3. Conversion flags (optional), which affect the result of some conversion types.
- 4. Minimum field width (optional). If specified as an '*' (asterisk), the actual width is read from the next element of the tuple in *values*, and the object to convert comes after the minimum field width and optional precision.
- 5. Precision (optional), given as a '.' (dot) followed by the precision. If specified as '*' (an asterisk), the actual precision is read from the next element of the tuple in *values*, and the value to convert comes after the precision.
- 6. Length modifier (optional).
- 7. Conversion type.

When the right argument is a dictionary (or other mapping type), then the formats in the string *must* include a parenthesised mapping key into that dictionary inserted immediately after the '%' character. The mapping key selects the value to be formatted from the mapping. For example:

Python has 002 quote types.

In this case no * specifiers may occur in a format (since they require a sequential parameter list).

The conversion flag characters are:

Flag	Meaning
'#'	The value conversion will use the "alternate form" (where defined below).
'0'	The conversion will be zero padded for numeric values.
1_1	The converted value is left adjusted (overrides the '0' conversion if both are given).
1 1	(a space) A blank should be left before a positive number (or empty string) produced by a signed conversion.
'+'	A sign character ('+' or '-') will precede the conversion (overrides a "space" flag).

A length modifier (h, l, or L) may be present, but is ignored as it is not necessary for Python - so e.g. %ld is identical to %d.

The conversion types are:

Conversion	Meaning	Notes
' d '	Signed integer decimal.	
'i'	Signed integer decimal.	
'0'	Signed octal value.	(1)
'u'	Obsolete type – it is identical to 'd'.	(6)
'X'	Signed hexadecimal (lowercase).	(2)
'X'	Signed hexadecimal (uppercase).	(2)
'e'	Floating point exponential format (lowercase).	(3)
'E'	Floating point exponential format (uppercase).	(3)
'f'	Floating point decimal format.	(3)
'F'	Floating point decimal format.	(3)
'g'	Floating point format. Uses lowercase exponential format if exponent is less than -4 or not less than precision, decimal format otherwise.	(4)
'G'	Floating point format. Uses uppercase exponential format if exponent is less than -4 or not less than precision, decimal	(4)

Conversion	Meaning	Notes
	format otherwise.	
'c'	Single character (accepts integer or single character string).	
'r'	String (converts any Python object using repr()).	(5)
's'	String (converts any Python object using str()).	(5)
'a'	String (converts any Python object using ascii()).	(5)
1%1	No argument is converted, results in a '%' character in the result.	

Notes:

- 1. The alternate form causes a leading octal specifier ('0o') to be inserted before the first digit.
- 2. The alternate form causes a leading '0x' or '0X' (depending on whether the 'x' or 'X' format was used) to be inserted before the first digit.
- 3. The alternate form causes the result to always contain a decimal point, even if no digits follow it.

The precision determines the number of digits after the decimal point and defaults to 6.

4. The alternate form causes the result to always contain a decimal point, and trailing zeroes are not removed as they would otherwise be.

The precision determines the number of significant digits before and after the decimal point and defaults to 6.

- 5. If precision is N, the output is truncated to N characters.
- 6. See PEP 237.

Since Python strings have an explicit length, %s conversions do not assume that '\0' is the end of the string.

Changed in version 3.1: %f conversions for numbers whose absolute value is over 1e50 are no longer replaced by %g conversions.

Binary Sequence Types — bytes, bytearray, memoryview

The core built-in types for manipulating binary data are bytes and bytearray. They are supported by memoryview which uses the buffer protocol to access the memory of other binary objects without needing to make a copy.

The array module supports efficient storage of basic data types like 32-bit integers and IEEE754 double-precision floating values.