DateParser Documentation

Release 0.7.0

Scrapinghub

Feb 25, 2018

Contents

dateparser provides modules to easily parse localized dates in almost any string formats commonly found on web pages.

CHAPTER 1

Documentation

Documentation is built automatically and can be found on Read the Docs.

CHAPTER 2

Features

- Generic parsing of dates in over 200 language locales plus numerous formats in a language agnostic fashion.
- Generic parsing of relative dates like: `'1 min ago'`, `'2 weeks ago'`, `'3 months, 1 week and 1 day ago'`, `'in 2 days'`, `'tomorrow'`.
- Generic parsing of dates with time zones abbreviations or UTC offsets like: `'August 14, 2015 EST'`, `'July 4, 2013 PST'`, `'21 July 2013 10:15 pm +0500'`.
- Date lookup in longer texts.
- Support for non-Gregorian calendar systems. See *Supported Calendars*.
- Extensive test coverage.

CHAPTER 3

Usage

The most straightforward way is to use the *dateparser.parse* function, that wraps around most of the functionality in the module.

dateparser.**parse**(**args*, ***kwargs*)

Parse date and time from given date string.

Parameters

- **date_string** (*str|unicode*) – A string representing date and/or time in a recognizably valid format.
- **date_formats** (*list*) – A list of format strings using directives as given here. The parser applies formats one by one, taking into account the detected languages/locales.
- **languages** (*list*) – A list of language codes, e.g. [‘en’, ‘es’, ‘zh-Hant’]. If locales are not given, languages and region are used to construct locales for translation.
- **locales** (*list*) – A list of locale codes, e.g. [‘fr-PF’, ‘qu-EC’, ‘af-NA’]. The parser uses locales to translate date string.
- **region** (*str|unicode*) – A region code, e.g. ‘IN’, ‘001’, ‘NE’. If locales are not given, languages and region are used to construct locales for translation.
- **settings** (*dict*) – Configure customized behavior using settings defined in dateparser.conf.Settings.

Returns Returns datetime representing parsed date if successful, else returns None

Return type datetime.

Raises ValueError - Unknown Language

3.1 Popular Formats

```
>>> import dateparser
>>> dateparser.parse('12/12/12')
datetime.datetime(2012, 12, 12, 0, 0)
>>> dateparser.parse(u'Fri, 12 Dec 2014 10:55:50')
datetime.datetime(2014, 12, 12, 10, 55, 50)
>>> dateparser.parse(u'Martes 21 de Octubre de 2014')  # Spanish (Tuesday 21 October␣
↪2014)
datetime.datetime(2014, 10, 21, 0, 0)
>>> dateparser.parse(u'Le 11 Décembre 2014 à 09:00')  # French (11 December 2014 at␣
↪09:00)
datetime.datetime(2014, 12, 11, 9, 0)
>>> dateparser.parse(u'13  2015 .  13:34')  # Russian (13 January 2015 at 13:34)
datetime.datetime(2015, 1, 13, 13, 34)
>>> dateparser.parse(u'1  2005, 1:00 AM')  # Thai (1 October 2005, 1:00 AM)
datetime.datetime(2005, 10, 1, 1, 0)
```

This will try to parse a date from the given string, attempting to detect the language each time.

You can specify the language(s), if known, using `languages` argument. In this case, given languages are used and language detection is skipped:

```
>>> dateparser.parse('2015, Ago 15, 1:08 pm', languages=['pt', 'es'])
datetime.datetime(2015, 8, 15, 13, 8)
```

If you know the possible formats of the dates, you can use the `date_formats` argument:

```
>>> dateparser.parse(u'22 Décembre 2010', date_formats=['%d %B %Y'])
datetime.datetime(2010, 12, 22, 0, 0)
```

3.2 Relative Dates

```
>>> parse('1 hour ago')
datetime.datetime(2015, 5, 31, 23, 0)
>>> parse(u'Il ya 2 heures')  # French (2 hours ago)
datetime.datetime(2015, 5, 31, 22, 0)
>>> parse(u'1 anno 2 mesi')  # Italian (1 year 2 months)
datetime.datetime(2014, 4, 1, 0, 0)
>>> parse(u'yaklaşık 23 saat önce')  # Turkish (23 hours ago)
datetime.datetime(2015, 5, 31, 1, 0)
>>> parse(u'Hace una semana')  # Spanish (a week ago)
datetime.datetime(2015, 5, 25, 0, 0)
>>> parse(u'2')  # Chinese (2 hours ago)
datetime.datetime(2015, 5, 31, 22, 0)
```

Note: Testing above code might return different values for you depending on your environment's current date and time.

Note: Support for relative dates in future needs a lot of improvement, we look forward to community's contribution to get better on that part. See **'Contributing'_**.

3.3 OOTB Language Based Date Order Preference

```
>>> # parsing ambiguous date
>>> parse('02-03-2016')  # assumes english language, uses MDY date order
datetime.datetime(2016, 3, 2, 0, 0)
>>> parse('le 02-03-2016')  # detects french, uses DMY date order
datetime.datetime(2016, 3, 2, 0, 0)
```

Note: Ordering is not locale based, that's why do not expect *DMY* order for UK/Australia English. You can specify date order in that case as follows usings *Settings*:

```
>>> parse('18-12-15 06:00', settings={'DATE_ORDER': 'DMY'})
datetime.datetime(2015, 12, 18, 6, 0)
```

For more on date order, please look at *Settings*.

3.4 Timezone and UTC Offset

By default, *dateparser* returns tzaware *datetime* if timezone is present in date string. Otherwise, it returns a naive *datetime* object.

```
>>> parse('January 12, 2012 10:00 PM EST')
datetime.datetime(2012, 1, 12, 22, 0, tzinfo=<StaticTzInfo 'EST'>)
```

```
>>> parse('January 12, 2012 10:00 PM -0500')
datetime.datetime(2012, 1, 12, 22, 0, tzinfo=<StaticTzInfo 'UTC\-05:00'>)
```

```
>>> parse('2 hours ago EST')
datetime.datetime(2017, 3, 10, 15, 55, 39, 579667, tzinfo=<StaticTzInfo 'EST
↪'>)
```

```
>>> parse('2 hours ago -0500')
datetime.datetime(2017, 3, 10, 15, 59, 30, 193431, tzinfo=<StaticTzInfo
↪'UTC\-05:00'>)
```

If date has no timezone name/abbreviation or offset, you can specify it using *TIMEZONE* setting.

```
>>> parse('January 12, 2012 10:00 PM', settings={'TIMEZONE': 'US/Eastern'})
datetime.datetime(2012, 1, 12, 22, 0)
```

```
>>> parse('January 12, 2012 10:00 PM', settings={'TIMEZONE': '+0500'})
datetime.datetime(2012, 1, 12, 22, 0)
```

TIMEZONE option may not be useful alone as it only attaches given timezone to resultant *datetime* object. But can be useful in cases where you want conversions from and to different timezones or when simply want a tzaware date with given timezone info attached.

```
>>> parse('January 12, 2012 10:00 PM', settings={'TIMEZONE': 'US/Eastern', 'RETURN_AS_
↪TIMEZONE_AWARE': True})
datetime.datetime(2012, 1, 12, 22, 0, tzinfo=<DstTzInfo 'US/Eastern' EST-1 day,␣
↪19:00:00 STD>)
```

```
>>> parse('10:00 am', settings={'TIMEZONE': 'EST', 'TO_TIMEZONE': 'EDT'})
datetime.datetime(2016, 9, 25, 11, 0)
```

Some more use cases for conversion of timezones.

```
>>> parse('10:00 am EST', settings={'TO_TIMEZONE': 'EDT'})  # date string has␣
↪timezone info
datetime.datetime(2017, 3, 12, 11, 0, tzinfo=<StaticTzInfo 'EDT'>)
```

```
>>> parse('now EST', settings={'TO_TIMEZONE': 'UTC'})  # relative dates
datetime.datetime(2017, 3, 10, 23, 24, 47, 371823, tzinfo=<StaticTzInfo 'UTC'>)
```

In case, no timezone is present in date string or defined in *settings*. You can still return tzaware *datetime*. It is especially useful in case of relative dates when uncertain what timezone is relative base.

```
>>> parse('2 minutes ago', settings={'RETURN_AS_TIMEZONE_AWARE': True})
datetime.datetime(2017, 3, 11, 4, 25, 24, 152670, tzinfo=<DstTzInfo 'Asia/Karachi'␣
↪PKT+5:00:00 STD>)
```

In case, you want to compute relative dates in UTC instead of default system's local timezone, you can use *TIMEZONE* setting.

```
>>> parse('4 minutes ago', settings={'TIMEZONE': 'UTC'})
datetime.datetime(2017, 3, 10, 23, 27, 59, 647248, tzinfo=<StaticTzInfo 'UTC'>)
```

Note: In case, when timezone is present both in string and also specified using *settings*, string is parsed into tzaware representation and then converted to timezone specified in *settings*.

```
>>> parse('10:40 pm PKT', settings={'TIMEZONE': 'UTC'})
datetime.datetime(2017, 3, 12, 17, 40, tzinfo=<StaticTzInfo 'UTC'>)
```

```
>>> parse('20 mins ago EST', settings={'TIMEZONE': 'UTC'})
datetime.datetime(2017, 3, 12, 21, 16, 0, 885091, tzinfo=<StaticTzInfo 'UTC'>)
```

For more on timezones, please look at *Settings*.

3.5 Incomplete Dates

```
>>> from dateparser import parse
>>> parse(u'December 2015')  # default behavior
datetime.datetime(2015, 12, 16, 0, 0)
>>> parse(u'December 2015', settings={'PREFER_DAY_OF_MONTH': 'last'})
datetime.datetime(2015, 12, 31, 0, 0)
>>> parse(u'December 2015', settings={'PREFER_DAY_OF_MONTH': 'first'})
datetime.datetime(2015, 12, 1, 0, 0)
```

```
>>> parse(u'March')
datetime.datetime(2015, 3, 16, 0, 0)
>>> parse(u'March', settings={'PREFER_DATES_FROM': 'future'})
datetime.datetime(2016, 3, 16, 0, 0)
>>> # parsing with preference set for 'past'
```

```
>>> parse('August', settings={'PREFER_DATES_FROM': 'past'})
datetime.datetime(2015, 8, 15, 0, 0)
```

You can also ignore parsing incomplete dates altogether by setting *STRICT_PARSING* flag as follows:

```
>>> parse(u'December 2015', settings={'STRICT_PARSING': True})
None
```

For more on handling incomplete dates, please look at *Settings*.

3.6 Search for Dates in Longer Chunks of Text

You can extract dates from longer strings of text. They are returned as list of tuples with text chunk containing the date and parsed datetime object.

```
>>> from dateparser.search import search_dates
>>> search_dates("The client arrived to the office for the first time in March 3rd,␣
↪2004 and got serviced, after a couple of months, on May 6th 2004, the customer␣
↪returned indicating a defect on the part")
[(u'in March 3rd, 2004 and', datetime.datetime(2004, 3, 3, 0, 0)),
 (u'on May 6th 2004', datetime.datetime(2004, 5, 6, 0, 0))]
```

CHAPTER 4

Dependencies

dateparser relies on following libraries in some ways:

- dateutil's module `relativedelta` for its freshness parser.
- jdatetime to convert *Jalali* dates to *Gregorian*.
- umalqurra to convert *Hijri* dates to *Gregorian*.
- tzlocal to reliably get local timezone.
- ruamel.yaml (optional) for operations on language files.

CHAPTER 5

Supported languages and locales

Language	Locales
en	‘en-001’, ‘en-150’, ‘en-AG’, ‘en-AI’, ‘en-AS’, ‘en-AT’, ‘en-AU’, ‘en-BB’, ‘en-BE’, ‘en-BI’, ‘en-BM’, ‘en-BS’, ‘en-BW
zh	
zh-Hans	‘zh-Hans-HK’, ‘zh-Hans-MO’, ‘zh-Hans-SG’
hi	
es	‘es-419’, ‘es-AR’, ‘es-BO’, ‘es-BR’, ‘es-BZ’, ‘es-CL’, ‘es-CO’, ‘es-CR’, ‘es-CU’, ‘es-DO’, ‘es-EA’, ‘es-EC’, ‘es-GQ’,
ar	‘ar-AE’, ‘ar-BH’, ‘ar-DJ’, ‘ar-DZ’, ‘ar-EG’, ‘ar-EH’, ‘ar-ER’, ‘ar-IL’, ‘ar-IQ’, ‘ar-JO’, ‘ar-KM’, ‘ar-KW’, ‘ar-LB’, ‘ar-L
bn	‘bn-IN’
fr	‘fr-BE’, ‘fr-BF’, ‘fr-BI’, ‘fr-BJ’, ‘fr-BL’, ‘fr-CA’, ‘fr-CD’, ‘fr-CF’, ‘fr-CG’, ‘fr-CH’, ‘fr-CI’, ‘fr-CM’, ‘fr-DJ’, ‘fr-DZ’, ‘
ur	‘ur-IN’
pt	‘pt-AO’, ‘pt-CH’, ‘pt-CV’, ‘pt-GQ’, ‘pt-GW’, ‘pt-LU’, ‘pt-MO’, ‘pt-MZ’, ‘pt-PT’, ‘pt-ST’, ‘pt-TL’
ru	‘ru-BY’, ‘ru-KG’, ‘ru-KZ’, ‘ru-MD’, ‘ru-UA’
id	
sw	‘sw-CD’, ‘sw-KE’, ‘sw-UG’
pa-Arab	
de	‘de-AT’, ‘de-BE’, ‘de-CH’, ‘de-IT’, ‘de-LI’, ‘de-LU’
ja	
te	
mr	
vi	
fa	‘fa-AF’
ta	‘ta-LK’, ‘ta-MY’, ‘ta-SG’
tr	‘tr-CY’
yue	
ko	‘ko-KP’
it	‘it-CH’, ‘it-SM’, ‘it-VA’
fil	
gu	
th	

Language	Locales
kn	
ps	
zh-Hant	‘zh-Hant-HK’, ‘zh-Hant-MO’
ml	
or	
pl	
my	
pa	
pa-Guru	
am	
om	‘om-KE’
ha	‘ha-GH’, ‘ha-NE’
nl	‘nl-AW’, ‘nl-BE’, ‘nl-BQ’, ‘nl-CW’, ‘nl-SR’, ‘nl-SX’
uk	
uz	
uz-Latn	
yo	‘yo-BJ’
ms	‘ms-BN’, ‘ms-SG’
ig	
ro	‘ro-MD’
mg	
ne	‘ne-IN’
as	
so	‘so-DJ’, ‘so-ET’, ‘so-KE’
si	
km	
zu	
cs	
sv	‘sv-AX’, ‘sv-FI’
hu	
el	‘el-CY’
sn	
kk	
rw	
ckb	‘ckb-IR’
qu	‘qu-BO’, ‘qu-EC’
ak	
be	
ti	‘ti-ER’
az	
az-Latn	
af	‘af-NA’
ca	‘ca-AD’, ‘ca-FR’, ‘ca-IT’
sr-Latn	‘sr-Latn-BA’, ‘sr-Latn-ME’, ‘sr-Latn-XK’
ii	
he	
bg	
bm	
ki	

Language	Locales
gsw	‘gsw-FR’, ‘gsw-LI’
sr	
sr-Cyrl	‘sr-Cyrl-BA’, ‘sr-Cyrl-ME’, ‘sr-Cyrl-XK’
ug	
zgh	
ff	‘ff-CM’, ‘ff-GN’, ‘ff-MR’
rn	
da	‘da-GL’
hr	‘hr-BA’
sq	‘sq-MK’, ‘sq-XK’
sk	
fi	
ks	
hy	
nb	‘nb-SJ’
luy	
lg	
lo	
bem	
kok	
luo	
uz-Cyrl	
ka	
ee	‘ee-TG’
mzn	
bs-Cyrl	
bs	
bs-Latn	
kln	
kam	
gl	
tzm	
dje	
kab	
bo	‘bo-IN’
shi-Latn	
shi	
shi-Tfng	
mn	
ln	‘ln-AO’, ‘ln-CF’, ‘ln-CG’
ky	
sg	
lt	
nyn	
guz	
cgg	
xog	
lrc	‘lrc-IQ’
mer	

Language	Locales
lu	
sl	
teo	‘teo-KE’
brx	
nd	
mk	
uz-Arab	
mas	‘mas-TZ’
nn	
kde	
mfe	
lv	
seh	
mgh	
az-Cyrl	
ga	
eu	
yi	
ce	
et	
ksb	
bez	
ewo	
fy	
ebu	
nus	
ast	
asa	
ses	
os	‘os-RU’
br	
cy	
kea	
lag	
sah	
mt	
vun	
rof	
jmc	
lb	
dav	
dyo	
dz	
nnh	
is	
khq	
bas	
naq	
mua	

Language	Locales
ksh	
saq	
se	‘se-FI’, ‘se-SE’
dua	
rwk	
mgo	
sbp	
to	
jgo	
ksf	
fo	‘fo-DK’
gd	
kl	
rm	
fur	
agq	
haw	
chr	
hsb	
wae	
nmg	
lkt	
twq	
dsb	
yav	
kw	
gv	
smn	
eo	
tl	

CHAPTER 6

Supported Calendars

- Gregorian calendar.
- Persian Jalali calendar. For more information, refer to Persian Jalali Calendar.
- Hijri/Islamic Calendar. For more information, refer to Hijri Calendar.

```
>>> from dateparser.calendars.jalali import JalaliCalendar
>>> JalaliCalendar(u'    ').get_date()
{'date_obj': datetime.datetime(2009, 3, 20, 0, 0), 'period': 'day'}
```

```
>>> from dateparser.calendars.hijri import HijriCalendar
>>> HijriCalendar(u'17-01-1437  08:30 ').get_date()
{'date_obj': datetime.datetime(2015, 10, 30, 20, 30), 'period': 'day'}
```

Note: *HijriCalendar* has some limitations with Python 3.

Note: For *Finnish* language, please specify *settings={'SKIP_TOKENS': []}* to correctly parse freshness dates.

Install using following command to use calendars.

Tip: pip install dateparser[calendars]

6.1 Using DateDataParser

`dateparser.parse()` uses a default parser which tries to detect language every time it is called and is not the most efficient way while parsing dates from the same source.

`DateDataParser` provides an alternate and efficient way to control language detection behavior.

The instance of `DateDataParser` reduces the number of applicable languages, until only one or no language is left. It assumes the previously detected language for all the subsequent dates supplied.

This class wraps around the core *dateparser* functionality, and by default assumes that all of the dates fed to it are in the same language.

class dateparser.date.**DateDataParser**(**args*, ***kwargs*)

Class which handles language detection, translation and subsequent generic parsing of string representing date and/or time.

Parameters

- **languages** (*list*) – A list of language codes, e.g. ['en', 'es', 'zh-Hant']. If locales are not given, languages and region are used to construct locales for translation.
- **locales** (*list*) – A list of locale codes, e.g. ['fr-PF', 'qu-EC', 'af-NA']. The parser uses locales to translate date string.
- **region** (*str|unicode*) – A region code, e.g. 'IN', '001', 'NE'. If locales are not given, languages and region are used to construct locales for translation.
- **try_previous_locales** – If True, locales previously used to translate date are tried first.
- **use_given_order** – If True, locales are tried for translation of date string in the order in which they are given.
- **settings** (*dict*) – Configure customized behavior using settings defined in `dateparser.conf.Settings`.

Returns A parser instance

Raises ValueError - Unknown Language, TypeError - Languages argument must be a list

get_date_data(*date_string*, *date_formats=None*)

Parse string representing date and/or time in recognizable localized formats. Supports parsing multiple languages and timezones.

Parameters

- **date_string** (*str|unicode*) – A string representing date and/or time in a recognizably valid format.
- **date_formats** (*list*) – A list of format strings using directives as given here. The parser applies formats one by one, taking into account the detected languages.

Returns a dict mapping keys to `datetime.datetime` object and *period*. For example: {'date_obj': datetime.datetime(2015, 6, 1, 0, 0), 'period': u'day'}

Raises ValueError - Unknown Language

Note: *Period* values can be a 'day' (default), 'week', 'month', 'year'.

Period represents the granularity of date parsed from the given string.

In the example below, since no day information is present, the day is assumed to be current day `16` from *current date* (which is June 16, 2015, at the moment of writing this). Hence, the level of precision is `month`:

```
>>> DateDataParser().get_date_data(u'March 2015')
{'date_obj': datetime.datetime(2015, 3, 16, 0, 0), 'period': u'month'}
```

Similarly, for date strings with no day and month information present, level of precision is `year` and day `16` and month `6` are from *current_date*.

```
>>> DateDataParser().get_date_data(u'2014')
{'date_obj': datetime.datetime(2014, 6, 16, 0, 0), 'period': u'year'}
```

Dates with time zone indications or UTC offsets are returned in UTC time unless specified using *Settings*.

```
>>> DateDataParser().get_date_data(u'23 March 2000, 1:21 PM CET')
{'date_obj': datetime.datetime(2000, 3, 23, 14, 21), 'period': 'day'}
```

Warning: It fails to parse *English* dates in the example below, because *Spanish* was detected and stored with the `ddp` instance:

```
>>> ddp.get_date_data('11 August 2012')
{'date_obj': None, 'period': 'day'}
```

`dateparser.date.DateDataParser` can also be initialized with known languages:

```
>>> ddp = DateDataParser(languages=['de', 'nl'])
>>> ddp.get_date_data(u'vr jan 24, 2014 12:49')
{'date_obj': datetime.datetime(2014, 1, 24, 12, 49), 'period': u'day'}
>>> ddp.get_date_data(u'18.10.14 um 22:56 Uhr')
{'date_obj': datetime.datetime(2014, 10, 18, 22, 56), 'period': u'day'}
```

CHAPTER 7

Settings

dateparser's parsing behavior can be configured by supplying settings as a dictionary to *settings* argument in *dateparser.parse* or `DateDataParser` constructor.

All supported *settings* with their usage examples are given below:

`DATE_ORDER` specifies the order in which date components *year*, *month* and *day* are expected while parsing ambiguous dates. It defaults to *MDY* which translates to *month* first, *day* second and *year* last order. Characters *M*, *D* or *Y* can be shuffled to meet required order. For example, *DMY* specifies *day* first, *month* second and *year* last order:

```
>>> parse('15-12-18 06:00')  # assumes default order: MDY
datetime.datetime(2018, 12, 15, 6, 0)  # since 15 is not a valid value for Month, it␣
↪is swapped with Day's
>>> parse('15-12-18 06:00', settings={'DATE_ORDER': 'YMD'})
datetime.datetime(2015, 12, 18, 6, 0)
```

`PREFER_LANGUAGE_DATE_ORDER` defaults to *True*. Most languages have a default *DATE_ORDER* specified for them. For example, for French it is *DMY*:

```
>>> # parsing ambiguous date
>>> parse('02-03-2016')  # assumes english language, uses MDY date order
datetime.datetime(2016, 3, 2, 0, 0)
>>> parse('le 02-03-2016')  # detects french, hence, uses DMY date order
datetime.datetime(2016, 3, 2, 0, 0)
```

Note: There's no language level default *DATE_ORDER* associated with *en* language. That's why it assumes *MDY* which is :obj:`settings <dateparser.conf.settings>` default. If the language has a default *DATE_ORDER* associated, supplying custom date order will not be applied unless we set *PREFER_LANGUAGE_DATE_ORDER* to *False*:

```
>>> parse('le 02-03-2016', settings={'DATE_ORDER': 'MDY'})
datetime.datetime(2016, 3, 2, 0, 0)  # MDY didn't apply
```

```
>>> parse('le 02-03-2016', settings={'DATE_ORDER': 'MDY', 'PREFER_LANGUAGE_DATE_ORDER
↪': False})
datetime.datetime(2016, 2, 3, 0, 0)  # MDY worked!
```

TIMEZONE defaults to local timezone. When specified, resultant datetime is localized with the given timezone.

```
>>> parse('January 12, 2012 10:00 PM', settings={'TIMEZONE': 'US/Eastern'})
datetime.datetime(2012, 1, 12, 22, 0)
```

TO_TIMEZONE defaults to None. When specified, resultant datetime converts according to the supplied timezone:

```
>>> settings = {'TIMEZONE': 'UTC', 'TO_TIMEZONE': 'US/Eastern'}
>>> parse('January 12, 2012 10:00 PM', settings=settings)
datetime.datetime(2012, 1, 12, 17, 0)
```

RETURN_AS_TIMEZONE_AWARE is a flag to toggle between timezone aware/naive dates:

```
>>> parse('30 mins ago', settings={'RETURN_AS_TIMEZONE_AWARE': True})
datetime.datetime(2017, 3, 13, 1, 43, 10, 243565, tzinfo=<DstTzInfo 'Asia/Karachi'␣
↪PKT+5:00:00 STD>)
```

```
>>> parse('12 Feb 2015 10:56 PM EST', settings={'RETURN_AS_TIMEZONE_AWARE': False})
datetime.datetime(2015, 2, 12, 22, 56)
```

PREFER_DAY_OF_MONTH This option comes handy when the date string is missing the day part. It defaults to current and can have first and last denoting first and last day of months respectively as values:

```
>>> from dateparser import parse
>>> parse(u'December 2015')  # default behavior
datetime.datetime(2015, 12, 16, 0, 0)
>>> parse(u'December 2015', settings={'PREFER_DAY_OF_MONTH': 'last'})
datetime.datetime(2015, 12, 31, 0, 0)
>>> parse(u'December 2015', settings={'PREFER_DAY_OF_MONTH': 'first'})
datetime.datetime(2015, 12, 1, 0, 0)
```

PREFER_DATES_FROM defaults to *current_period* and can have *past* and *future* as values.

If date string is missing some part, this option ensures consistent results depending on the *past* or *future* preference, for example, assuming current date is *June 16, 2015*:

```
>>> from dateparser import parse
>>> parse(u'March')
datetime.datetime(2015, 3, 16, 0, 0)
>>> parse(u'March', settings={'PREFER_DATES_FROM': 'future'})
datetime.datetime(2016, 3, 16, 0, 0)
>>> # parsing with preference set for 'past'
>>> parse('August', settings={'PREFER_DATES_FROM': 'past'})
datetime.datetime(2015, 8, 15, 0, 0)
```

RELATIVE_BASE allows setting the base datetime to use for interpreting partial or relative date strings. Defaults to the current date and time.

For example, assuming current date is *June 16, 2015*:

```
>>> from dateparser import parse
>>> parse(u'14:30')
datetime.datetime(2015, 3, 16, 14, 30)
```

```
>>> parse(u'14:30', settings={'RELATIVE_BASE': datetime.datetime(2020, 1, 1)})
datetime.datetime(2020, 1, 1, 14, 30)
>>> parse(u'tomorrow', settings={'RELATIVE_BASE': datetime.datetime(2020, 1, 1)})
datetime.datetime(2020, 1, 2, 0, 0)
```

`STRICT_PARSING` defaults to *False*.

When set to *True* if missing any of *day*, *month* or *year* parts, it does not return any result altogether.:

```
>>> parse(u'March', settings={'STRICT_PARSING': True})
None
```

`SKIP_TOKENS` is a `list` of tokens to discard while detecting language. Defaults to `['t']` which skips T in iso format datetime string .e.g. `2015-05-02T10:20:19+0000`.:

```
>>> from dateparser.date import DateDataParser
>>> DateDataParser(settings={'SKIP_TOKENS': ['de']}).get_date_data(u'27 Haziran 1981
↪de')  # Turkish (at 27 June 1981)
{'date_obj': datetime.datetime(1981, 6, 27, 0, 0), 'period': 'day'}
```

CHAPTER 8

Documentation

Contents:

8.1 Installation

At the command line:

```
$ pip install dateparser
```

Or, if you don't have pip installed:

```
$ easy_install dateparser
```

If you want to install from the latest sources, you can do:

```
$ git clone https://github.com/scrapinghub/dateparser.git
$ cd dateparser
$ python setup.py install
```

8.1.1 Deploying dateparser in a Scrapy Cloud project

The initial use cases for *dateparser* were for Scrapy projects doing web scraping that needed to parse dates from websites. These instructions show how you can deploy it in a Scrapy project running in Scrapy Cloud.

Deploying with shub

The most straightforward way to do that is to use the latest version of the shub command line tool.

First, install `shub`, if you haven't already:

```
pip install shub
```

Then, you can choose between deploying a stable release or the latest from development.

Deploying a stable dateparser release:

1. Then, use `shub` to install python-dateutil (we require at least 2.3 version), jdatetime and PyYAML dependencies from PyPI:

```
shub deploy-egg --from-pypi python-dateutil YOUR_PROJECT_ID
shub deploy-egg --from-pypi jdatetime YOUR_PROJECT_ID
shub deploy-egg --from-pypi PyYAML YOUR_PROJECT_ID
```

2. Finally, deploy dateparser from PyPI:

```
shub deploy-egg --from-pypi dateparser YOUR_PROJECT_ID
```

Deploying from latest sources

Optionally, you can deploy it from the latest sources:

Inside the `dateparser` root directory:

1. Run the command to deploy the dependencies:

```
shub deploy-reqs YOUR_PROJECT_ID requirements.txt
```

2. Then, either deploy from the latest sources on GitHub:

```
shub deploy-egg --from-url git@github.com:scrapinghub/dateparser.git YOUR_PROJECT_
↪ID
```

Or, just deploy from the local sources (useful if you have local modifications):

```
shub deploy-egg
```

Deploying the egg manually

In case you run into trouble with the above procedure, you can deploy the egg manually. First clone the `dateparser`'s repo, then inside its directory run the command:

```
python setup.py bdist_egg
```

After that, you can upload the egg using Scrapy Cloud's Dashboard interface under Settings > Eggs section.

Dependencies

Similarly, you can download source and package PyYAML, jdatetime and dateutil (version >= 2.3) as *eggs* and deploy them like above.

8.2 Contributing

Contributions are welcome, and they are greatly appreciated! Every little bit helps, and credit will always be given.

You can contribute in many ways:

8.2.1 Types of Contributions

Report Bugs

Report bugs at https://github.com/scrapinghub/dateparser/issues.

If you are reporting a bug, please include:

- Your operating system name and version.
- Any details about your local setup that might be helpful in troubleshooting.
- Detailed steps to reproduce the bug.

Fix Bugs

Look through the GitHub issues for bugs. Anything tagged with "bug" is open to whoever wants to implement it.

Implement Features

Look through the GitHub issues for features. Anything tagged with "feature" is open to whoever wants to implement it. We encourage you to add new languages to existing stack.

Write Documentation

DateParser could always use more documentation, whether as part of the official DateParser docs, in docstrings, or even on the web in blog posts, articles, and such.

Submit Feedback

The best way to send feedback is to file an issue at https://github.com/scrapinghub/dateparser/issues.

If you are proposing a feature:

- Explain in detail how it would work.
- Keep the scope as narrow as possible, to make it easier to implement.
- Remember that contributions are welcome :)

8.2.2 Get Started!

Ready to contribute? Here's how to set up *dateparser* for local development.

1. Fork the *dateparser* repo on GitHub.
2. Clone your fork locally:

```
$ git clone git@github.com:your_name_here/dateparser.git
```

3. Install your local copy into a virtualenv. Assuming you have virtualenvwrapper installed, this is how you set up your fork for local development:

```
$ mkvirtualenv dateparser
$ cd dateparser/
$ python setup.py develop
```

4. Create a branch for local development:

```
$ git checkout -b name-of-your-bugfix-or-feature
```

   Now you can make your changes locally.

5. When you're done making changes, check that your changes pass flake8 and the tests, including testing other Python versions with tox:

```
$ pip install -r tests/requirements.txt # install test dependencies
$ pip install -r scripts/requirements.txt # install script dependencies
$ flake8 dateparser tests
$ nosetests
$ tox
```

   To get flake8 and tox, just pip install them into your virtualenv. (Note that we use `max-line-length = 100` for flake8, this is configured in `setup.cfg` file.)

6. Commit your changes and push your branch to GitHub:

```
$ git add .
$ git commit -m "Your detailed description of your changes."
$ git push origin name-of-your-bugfix-or-feature
```

7. Submit a pull request through the GitHub website.

8.2.3 Pull Request Guidelines

Before you submit a pull request, check that it meets these guidelines:

1. The pull request should include tests.
2. If the pull request adds functionality, the docs should be updated. Put your new functionality into a function with a docstring, and add the feature to the list in *README.rst*.
3. Check https://travis-ci.org/scrapinghub/dateparser/pull_requests and make sure that the tests pass for all supported Python versions.
4. Follow the core developers' advice which aim to ensure code's consistency regardless of variety of approaches used by many contributors.
5. In case you are unable to continue working on a PR, please leave a short comment to notify us. We will be pleased to make any changes required to get it done.

8.2.4 Guidelines for Editing Translation Data

English is the primary language of the dateparser. Dates in all other languages are translated into English equivalents before they are parsed. The language data required for parsing dates is contained in

dateparser/data/date_translation_data. It contains variable parts that can be used in dates, language by language: month and week names - and their abbreviations, prepositions, conjunctions and frequently used descriptive words and phrases (like "today"). The data in *dateparser/data/date_translation_data* is formed by supplementing data retrieved from unicode CLDR, contained in *data/cldr_language_data/date_translation_data*, with supplementary data contributed by the community, contained in *data/supplementary_language_data/date_translation_data*. Additional data to supplement existing data or translation data for a new language should be added to *dateparser_data/supplementary_language_data/date_translation_data*. The chosen data format is YAML because it is readable and simple to edit. After adding or changing any data in YAML files we need to move them to internal data files with *scripts/write_complete_data.py*. Otherwise the changes to YAML files will not have any effect.

Refer to language-data-template for details about its structure and take a look at already implemented languages for examples. As we deal with the delicate fabric of interwoven languages, tests are essential to keep the functionality across them. Therefore any addition or change should be reflected in tests. However, there is nothing to be afraid of: our tests are highly parameterized and in most cases a test fits in one declarative line of data. Alternatively, you can provide required information and ask the maintainers to create the tests for you.

8.3 Credits

8.3.1 Committers

- Adam LeVasseur
- Ahmad Musaffa
- Alec Koumjian
- Alexis Svinartchouk
- Ammar Azif
- Andrés Portillo
- Andrey Zhelnin
- Artur Sadurski
- Artur Gaspar
- atchoum31
- Benjamin Bach
- Cesar Flores
- CJStuart
- Claudio Salazar
- conanca
- David Beitey
- Dawid Wolski
- demelziraptor
- Edwin Zhang
- Elena Zakharova
- Elias Dorneles
- Eugene Amirov

- Faisal Anees
- Fernando Tricas García
- Georgi Valkov
- Hristo Vrigazov
- ishirav
- Ismael Carnales
- James M. Allen
- Ján Jančár
- Jolo Balbin
- Joseph Kahn
- Mark Baas
- Marko Horvatić
- Mateusz Golewski
- Mats Gustafsson
- Michael Palumbo
- nanolab
- Opp Lieamsiriwong
- Paul Tremberth
- Pengyu Chen
- phuslu
- Rajat Goyal
- Raul Gallegos
- Robert Schütz
- Roman
- Sakari Vaelma
- samoylovfp
- Sarthak Madaan
- Shuai Lin
- Sigit Dewanto
- Sinan Nalkaya
- Sviatoslav Sydorenko
- Taito Horiuchi
- Takahiro Kamatani
- Thomas Steinacher
- Timothy Allen
- tkisme

- Tom Russell
- Umair Ashraf
- Waqas Shabir
- Xavier Barbosa
- Yongwen Zhuang

8.4 History

8.4.1 0.7.0 (2018-02-08)

Features added during Google Summer of Code 2017: * Harvesting language data from Unicode CLDR database (https://github.com/unicode-cldr/cldr-json), which includes over 200 locales (#321) - authored by Sarthak Maddan. See full currently supported locale list in README. * Extracting dates from longer strings of text (#324) - authored by Elena Zakharova. Special thanks for their awesome contributions!

New features: * Added (independently from CLDR) Georgian (#308) and Swedish (#305)

Improvements: * Improved support of Chinese (#359), Thai (#345), French (#301, #304), Russian (#302) * Removed ruamel.yaml from dependencies (#374). This should reduce the number of installation issues and improve performance as the result of moving away from YAML as basic data storage format. Note that YAML is still used as format for support language files. * Improved performance through using pre-compiling frequent regexes and lazy loading of data (#293, #294, #295, #315) * Extended tests (#316, #317, #318, #323) * Updated nose_parameterized to its current package, parameterized (#381)

Planned for next release: * Full language and locale names * Performance and stability improvements * Documentation improvements

8.4.2 0.6.0 (2017-03-13)

New features: * Consistent parsing in terms of true python representation of date string. See #281 * Added support for Bangla, Bulgarian and Hindi languages.

Improvements:

- Major bug fixes related to parser and system's locale. See #277, #282
- Type check for timezone arguments in settings. see #267
- Pinned dependencies' versions in requirements. See #265
- Improved support for cn, es, dutch languages. See #274, #272, #285

Packaging: * Make calendars extras to be used at the time of installation if need to use calendars feature.

8.4.3 0.5.1 (2016-12-18)

New features:

- Added support for Hebrew

Improvements:

- Safer loading of YAML. See #251
- Better timezone parsing for freshness dates. See #256

- Pinned dependencies' versions in requirements. See #265
- Improved support for zh, fi languages. See #249, #250, #248, #244

8.4.4 0.5.0 (2016-09-26)

New features:

- *DateDataParser* now also returns detected language in the result dictionary.
- Explicit and lucid timezone conversion for a given datestring using *TIMEZONE*, *TO_TIMEZONE* settings.
- Added Hungarian langauge.
- Added setting, *STRICT_PARSING* to ignore imcomplete dates.

Improvements:

- Fixed quite a few parser bugs reported in issues #219, #222, #207, #224.
- Improved support for chinese language.
- Consistent interface for both Jalali and Hijri parsers.

8.4.5 0.4.0 (2016-06-17)

New features:

- Support for Language based date order preference while parsing ambiguous dates.
- Support for parsing dates with no spaces in between components.
- Support for custom date order preference using *settings*.
- Support for parsing generic relative dates in future.e.g. *tomorrow*, *in two weeks*, etc.
- Added *RELATIVE_BASE* settings to set date context to any datetime in past or future.
- Replaced dateutil.parser.parse with dateparser's own parser.

Improvements:

- Added simplifications for *12 noon* and *12 midnight*.
- Fixed several bugs
- Replaced PyYAML library by its active fork *ruamel.yaml* which also fixed the issues with installation on windows using python35.
- More predictable *date_formats* handling.

8.4.6 0.3.5 (2016-04-27)

New features:

- Danish language support.
- Japanese language support.
- Support for parsing date strings with accents.

Improvements:

- Transformed languages.yaml into base file and separate files for each language.

- Fixed vietnamese language simplifications.
- No more version restrictions for python-dateutil.
- Timezone parsing improvements.
- Fixed test environments.
- Cleaned language codes. Now we strictly follow codes as in ISO 639-1.
- Improved chinese dates parsing.

8.4.7 0.3.4 (2016-03-03)

Improvements:

- Fixed broken version 0.3.3 by excluding latest python-dateutil version.

8.4.8 0.3.3 (2016-02-29)

New features:

- Finnish language support.

Improvements:

- Faster parsing with switching to regex module.
- *RETURN_AS_TIMEZONE_AWARE* setting to return tz aware date object.
- Fixed conflicts with month/weekday names similarity across languages.

8.4.9 0.3.2 (2016-01-25)

New features:

- Added Hijri Calendar support.
- Added settings for better control over parsing dates.
- Support to convert parsed time to the given timezone for both complete and relative dates.

Improvements:

- Fixed problem with caching `datetime.now()` in `FreshnessDateDataParser`.
- Added month names and week day names abbreviations to several languages.
- More simplifications for Russian and Ukranian languages.
- Fixed problem with parsing time component of date strings with several kinds of apostrophes.

8.4.10 0.3.1 (2015-10-28)

New features:

- Support for Jalali Calendar.
- Belarusian language support.
- Indonesian language support.

Improvements:

- Extended support for Russian and Polish.
- Fixed bug with time zone recognition.
- Fixed bug with incorrect translation of "second" for Portuguese.

8.4.11 0.3.0 (2015-07-29)

New features:

- Compatibility with Python 3 and PyPy.

Improvements:

- *languages.yaml* data cleaned up to make it human-readable.
- Improved Spanish date parsing.

8.4.12 0.2.1 (2015-07-13)

- Support for generic parsing of dates with UTC offset.
- Support for Tagalog/Filipino dates.
- Improved support for French and Spanish dates.

8.4.13 0.2.0 (2015-06-17)

- Easy to use *parse* function
- Languages definitions using YAML.
- Using translation based approach for parsing non-english languages. Previously, `dateutil.parserinfo` was used for language definitions.
- Better period extraction.
- Improved tests.
- Added a number of new simplifications for more comprehensive generic parsing.
- Improved validation for dates.
- Support for Polish, Thai and Arabic dates.
- Support for `pytz` timezones.
- Fixed building and packaging issues.

8.4.14 0.1.0 (2014-11-24)

- First release on PyPI.

CHAPTER 9

Indices and tables

- genindex
- modindex
- search

Python Module Index

d

Index