We extract temporal expressions and normalize them simultaneously. The architecture of the rule-based temporal expression identification system is shown in Figure 1.



Figure 1. the architecture of the rule-based temporal expression identification system

class Section:

type: we define five section types as follows.

sec\_type={ 'undef':'UNDEF',

'discharge':'DISCHARGE',

'admission':'ADMISSION',

'history':'HISTORY OF PRESENT ILLNESS',

'hospital course':'HOSPITAL COURSE'

}

start: the row number of the first sentence

context: sentences in current section

class Note:

sec: sections in current note

class TemporalExpression:

sec: which section type does current temporal expression belong to?

text: the context of current temporal expression

row: the row number of the sentence current temporal expression

start: the first character position of current temporal expression in the sentence

end: the last character position of current temporal expression in the sentence

start\_word: the first word position of current temporal expression in the sentence

end\_word: the last word position of current temporal expression in the sentence

value: the normalization of current temporal expression

type: the type of current temporal expression

te\_type = ['DATE', 'TIME', 'DURATION', 'FREQUENCY']

mod: which modifier express current temporal expression

mod\_type = ['NA', 'MORE', 'LESS', 'APPROX', 'START', 'END', 'MIDDLE']

class SectionTemporalExpression:

text: the context of current temporal expression

row: the row number of the sentence current temporal expression

start: the first position of current temporal expression in the sentence

end: the last position of current temporal expression in the sentence

value: the normalization of current temporal expression

type: the type of current temporal expression

sec\_te\_type = ['ADMISSION', 'DISCHARGE']

class Rule:

static:

pattern: pattern used for regular expressions

norm\_pattern: patterns used for normalization, including value and mod

type: which type temporal expressions dose this Rule identify?

'DATE', 'TIME', 'DURATION' or 'FREQUENCY'

expression: regular expression used for searching

attribute: a dictionary of attributes of temporal expressions identified by this rule.

Example:

To identify 'DATE's like MONTH-DAY-2YEAR ( such as 01-29-10), we build the following rule:

expression="%MonthNumber-%DayNumber-%Year2Digit",val="UNDEF-centurygroup(3)-%NormMonth(group(1))-group(2)" (1)

where " MonthNumber ", " DayNumber " and "Year2Digit" are patterns like:

"

Year4Digit:

[12]\d{3}

Year2Digit:

\d{2}

MonthNumber:

1[0-2]

0?[1-9]

DayNumber:

[12][0-9]

3[01]

0?[1-9]

"

" NormMonth " is normal pattern like:

"

NormMonth:

jan 01

1 01

NormDay:

9 09

Mod:

over MORE

"

It means this rule uses regular expression "%MonthNumber-%DayNumber-%Year2Digit" for temporal expression extraction and identifies the value attribute by " UNDEF-centurygroup(3)-%NormMonth(group(1))-group(2)".

Firstly, we parse the expression "%MonthNumber-%DayNumber-%Year2Digit" by looking up the pattern dictionary: "(1[0-2]|0?[1-9])-([12][0-9]|3[01]|0?[1-9])-(\d{2})".

Then, identify the value attribute by matched groups and normal patterns.

For "at 2011-01-29 , we had a meeting . posted time : 1-9-10", a temporal expression "1-9-10" can be extracted by rule (1), and the value attribute can be normalized to "UNDEF-century10-01-09", where " UNDEF-century" is the missing part that should be added later.