GOLD STANDARD REGEXES

Listed here are the gold standard regexes for the six entity extraction tasks specified in Table 2. The procedure for generating the gold standard regexes is outlined in the Experiments section. Notice that some of the regexes contain positive and negative lookbehinds ((?<=) and (?<!), respectively).

Table 1: DATE_{MIDEAST}

2	$}/\d{2}/\d{2}$	
\d/\	$d{2}/d{2}$	
\d/\	$d/d{2}$	
$d{2}-d{2}-d{2}$		
2	$.\d{1}\.\d{2}$	

Table 2: DATEWEBKB

\d{2}/\d{2}/\d{2} \d/\d{2}/\d{2} \d{2}/\d/\d{2} \d{2}/\d/\d{2}
\d{2}-\d{2}-\d{2} 1-0ct-96
25-May 4-November
06nov96
26aug96 9-September-96
08-Oct-96 12nov96
12110190

Table 3: DATE_{ENRON}

\d{2}/\d{2}/\d{2} \d/\d{2}/\d{2} \d/\d{2}/\d{2} \d{2}/\d/\d{2} \d/\d/\d{2} \d{2}\.\d{2}\.0\d \d{2}\.03\.\d{2} 01-\d1-\d{2} \d{2}-\d{2}-99 \d{2}-\d{2}-99 \d{2}-\d{2}-0\d 00-\d{2}-\d{2} \d-\d-0\d \d-\d-9\d \d\.\d\2}-\d{2} \d\.\d{2}\.0\d \d\.\d{2}\.0\d \d\2}-\d-0\d \d{2}-\d-0\d \d{2}-\d-9\d \d{2}\.\d\2}\.0\d \d{2}\.1\.9\d	
\d{2}/\d/\d{2} \d/\d/\d{2} \d{2}\.\d{2}\.0\d \d{2}\.03\.\d{2} 01-\d1-\d{2} \d{2}-\d{2}-99 \d{2}-\d{2}-99 \d{2}-\d{2}-0\d 00-\d{2}-\d{2} \d-\d-0\d \d-\d-9\d \d\.\d\.0\d \d-\d{2}-\d{2} \d\.\d{2}\.0\d \d\.\d{2}\.0\d \d{2}-\d-9\d \d{2}-\d-9\d \d{2}\.\d\.0\d	$\d{2}/\d{2}/\d{2}$
\d/\d/\d{2} \d{2}\.\d{2}\.0\d \d{2}\.03\.\d{2} 01-\d1-\d{2} \d{2}-\d{2}-99 \d{2}-\d{2}-0\d 00-\d{2}-\d{2} \d-\d-0\d \d-\d-9\d \d\.\d\.0\d \d-\d{2}-\d{2} \d\.\d{2}\.0\d \d\.\d{2}\.0\d \d{2}-\d-0\d \d{2}-\d-9\d \d{2}\.\d\.0\d	$\d/\d{2}/\d{2}$
\d{2}\.\d{2}\.0\d \d{2}\.03\.\d{2} 01-\d1-\d{2} 01-\d1-\d{2} \d{2}-\d{2}-99 \d{2}-\d{2}-0\d 00-\d{2}-\d{2} \d-\d-0\d \d-\d-9\d \d-\d-9\d \d\.\d\.0\d \d-\d{2}-\d{2} \d\.\d{2}\.0\d \d\.\d{2}\.0\d \d\.2\.9\d \d{2}-\d-9\d \d{2}-\d-9\d \d{2}\.\d-9\d	$\d{2}/\d/\d{2}$
\d{2}\.03\.\d{2} 01-\d1-\d{2} \d{2}-\d{2}-99 \d{2}-\d{2}-0\d 00-\d{2}-\d{2} \d-\d-0\d \d-\d-9\d \d\.\d\.0\d \d-\d{2}-\d{2} \d\.\d{2}\.0\d \d\.\d{2}\.0\d \d\.\d{2}\.0\d \d\.\d{2}\.9\d \d{2}-\d-9\d \d{2}-\d-9\d \d{2}\.\d\.0\d	$\d/\d/\d{2}$
01-\d1-\d{2} \d{2}-\d{2}-99 \d{2}-\d{2}-0\d 00-\d{2}-\d{2} \d-\d-0\d \d-\d-9\d \d\.\d\.0\d \d-\d{2}-\d{2} \d\.\d{2}\.0\d \d\.\d{2}\.9\d \d{2}-\d-0\d \d{2}-\d-9\d \d{2}\.\d\.0\d	$\d{2}\.\d{2}\.\0\d$
\d{2}-\d{2}-99 \d{2}-\d{2}-0\d 00-\d{2}-\d{2} \d-\d-0\d \d-\d-9\d \d\.\d\.0\d \d-\d{2}-\d{2} \d\.\d{2}\.0\d \d\.\d{2}\.9\d \d{2}-\d-9\d \d{2}\.\d\.0\d	$\d{2}\.03\.\d{2}$
\d{2}-\d{2}-0\d 00-\d{2}-\d{2} \d-\d-0\d \d-\d-9\d \d\.\d\.0\d \d-\d{2}-\d{2} \d\.\d{2}\.0\d \d\.\d{2}\.9\d \d{2}-\d-0\d \d{2}-\d-9\d \d{2}\.\d\.0\d	$01 - d1 - d\{2\}$
00-\d{2}-\d{2} \d-\d-0\d \d-\d-9\d \d\.\d\.0\d \d-\d{2}-\d{2} \d\.\d{2}\.0\d \d\.\d{2}\.9\d \d{2}-\d-0\d \d{2}-\d-9\d \d{2}\.\d\.0\d	$d{2}-d{2}-99$
\d-\d-0\d \d-\d-9\d \d\.\d\.0\d \d-\d{2}-\d{2} \d\.\d{2}\.0\d \d\.\d{2}\.9\d \d{2}-\d-0\d \d{2}-\d-9\d \d{2}\.\d\.0\d	$d{2}-d{2}-0d$
\d-\d-9\d \d\.\d\.0\d \d-\d{2}-\d{2} \d\.\d{2}\.0\d \d\.\d{2}\.9\d \d{2}-\d-0\d \d{2}-\d-9\d \d{2}\.\d\.0\d	$00-d\{2\}-d\{2\}$
\d\.\d\.0\d \d-\d{2}-\d{2} \d\.\d{2}\.0\d \d\.\d{2}\.9\d \d{2}-\d-0\d \d{2}-\d-9\d \d{2}\.\d\.0\d	\d-\d-0\d
\d-\d{2}-\d{2} \d\.\d{2}\.0\d \d\.\d{2}\.9\d \d{2}-\d-0\d \d{2}-\d-9\d \d{2}\.\d\.0\d	\d-\d-9\d
\d\.\d{2}\.0\d \d\.\d{2}\.9\d \d{2}-\d-0\d \d{2}-\d-9\d \d{2}\.\d\.0\d	\d\.\d\.0\d
\d\.\d{2}\.9\d \d{2}-\d-0\d \d{2}-\d-9\d \d{2}\.\d\.0\d	$\d-\d\{2\}-\d\{2\}$
\d{2}-\d-0\d \d{2}-\d-9\d \d{2}\.\d\.0\d	$\d\.\d{2}\.\0\d$
\d{2}-\d-9\d \d{2}\.\d\.0\d	$\d\.\d{2}\.\9\d$
\d{2}\.\d\.0\d	$d{2}-d-0$
	$d{2}-d-9$ d
$\d{2}\.1\.9\d$	$\d{2}\.\d\.0\d$
	$\d{2}\.1\.9\d$

Table 4: PHONE_{FORSALE}

Table 5: COURSEWEBKB

\(\d{3}\)_\d{3}-\d{4}	CS\d{3}
$d{3}-d{3}-d{4}$	cs\d{3}
$\d{3}/\d{3}-\d{4}$	EE\d{3}
$\d{3}_{\d}{3}-\d{4}$	ee\d{3}
$\(\d{3}\)\d{3}-\d{4}$	$CSE\d{3}$
$\(\d{3}\) - \d{3} - \d{4}$	$CS\d{3}[A-Z]$
$(d{3})/d{3}-d{4}$	CS\d{4}
\(\d{3}\)_\d{3}_\d{4}	$EE\d{3}[a-zA-Z]$
$\d{3}\.\d{3}\.\d{4}$	$CSE\d{3}[a-zA-Z]$
\d{3}_\d{3}_\d{4}	$PS\d{3}$
$d{3}-d{3}_{d}{4}$	$MIS\d{3}$
$\(\d{3}\)\d{7}$	$MA\d{3}$
$(?$	$ma\d{3}$
	CS_\d{2}
	CS\d{2}
	$CSE\d{3}[A-Z]{2}$

Table 6: PHONE_{ENRON}

```
d{3}-d{3}-d{4}
\d{3}-\d{3}-\d{4}
d{3}/d{3}-d{4}
\d{3}_{,,d}{3} - \d{4}
\d{3}_{uu}\d{3}-\d{4}
\d{3}\.\d{3}\.\d{4}
\d{3}_{\d{3}}\d{4}
\d{3}-\d{3}_{\d{4}}
d{4}-d{3}-d{4}
d{3} \ . \ d{3} - d{4}
\d{3}-\d{3}-\d{4}
\d{3}-\d{3}-[A-Z]{4}
d{3}-[A-Z]{3}-d{4}
d{3}-d{3}-w{4}
\d{3}/\d{3}/\d{4}
\d{3}/\d{3}\. \\ d{4}
d{2}-d{4}-d{4}
1-\d{3}-[A-Z]{6}\d
1-\d{3}-[A-Z]{7}
1-\d{3}-[A-Z]{3}-[A-Z]{4}
1-\d{3}-[A-Z]\d-[A-Z]{5}
1-\d{3}-[A-Z]\d-[A-Z][a-z]{4}
```