NLP Project: Time Extractor

Authors: Hongda Jiang (hj690@nyu.edu), Kailun Wu (kailun@nyu.edu)

Overview

Time Extractor is a time extraction API that can recognize time expressions and convert them to Java Calendar objects. The workflow is:

```
Reading articles → Regex Matching → Normalization → Calendar objects
```

Here's a sample of time expressions and extracted date this API can identify, given the reference time of `05-06-2015` :

```
mon 2:35 == 05-04-2015 02:35 AM
4pm == 05-06-2015 04:00 PM
6 in the morning == 05-06-2015 06:00 AM
friday 1pm == 05-08-2015 01:00 PM
sat 7 in the evening == 05-00-2015 07:00 PM
next month == 06-01-2015 00:00 AM
17:00 == 05-06-2015 05:00 PM
January 5 at 7pm == 01-05-2015 07:00 PM
1979-05-27 05:00 == 05-27-1979 05:00 AM
```

How to test on an article

You can use the `run.sh` script to compile the source code and test on a text body as follows:

```
$ bash run.sh <article_path> <reference_time>
```

The `reference_time` should be of the format `MM-DD-YYYY` . The result can be seen in stdout.

Alternatively, you can test on the training data by running with one argument:

```
$ bash run.sh training
```

Performance

We ran the test on WSJ articles `appointments.txt` provided by Jet and got 1168 time expressions. The running time is:

```
real    0m4.309s
user    0m6.825s
sys     0m0.269s
```

Thanks for your review!