Homework 7

Use the pubmed dataset and download 500 entries (you can use easy-pubmed in R or pubmed-look up in Python). Please convert the data into year and month temporal scopes, then identify keywords which are occurring together.

In detail you need to perform followings:

1) Convert the data into different month, temporal scope.
For example, following XML entry:

```
….
  <PubDate>
              <Year>2008</Year>
              <Month>Feb</Month>
  </PubDate>
….
<ArticleTitle>Relation of serum lipids and lipoproteins to obesity.
</ArticleTitle>
```

Conversion to year temporal scope:

```
2008, Relation of serum lipids and lipoproteins to obesity
```

Conversion to month temporal scope:

```
Feb, 2008, Relation of serum lipids and lipoproteins to obesity
```

2) Use bag of words to construct a document term matrix and extract keywords in article title (without stop words).

3) Use two algorithms, one association rule mining and one sequence mining algorithm to extract frequent patterns of words in articles.

The output should be something like this:
"Obesity", "risk"
"cancer", "diagnosis"…

You are free to choose two algorithms from the set of algorithms described in the class. Your implementation should be flexible to work with different minimum support and confidence.

Please do not forget to submit the video presentation of your homework as well. If you have any trouble during your work, please contact the TA or RA before the deadline. We will schedule an appointment to help you progress in your homework.