Processing lexical resources

- We have multiple lexical resources (files with words).
  They are:
  EmoSN, NRC, Sentisense, AFINN, ANEW, DAL

- Some of these resources have words associated to a sentiment, some are more general (like positive and negative words) in GI, HL, LIWC, listPosEffTerms and listNegEffTerms.

- In addition some resources have one or more scores associated to the words.
  For instance, for the resource called ANEW we have for each word *Valence (*or *Pleaseness), Arousal and Dominance*;
  for the resource called DAL we have *Activation, Imaginery* and *Pleaseness.*

Storing the lexical resources

- How to store in main memory and in the database the content of lexical resources?

- This depends on the DBMS that we will use.

Purposes of the laboratory

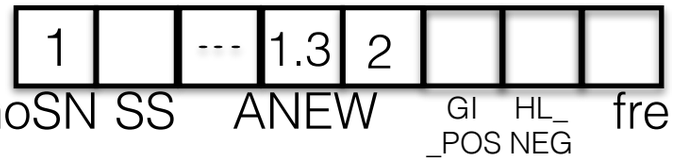
- Our laboratory has two purposes:

- For each word found in the Twitter messages:

1. list the lexical resources containing each word, so that we can rely upon a unique resource source obtained by the fusion of the single resources.

2. count the number of occurrences of each word in the Twitter messages for each emotion so that at the end we can draw a word cloud associated to the most frequent words in each emotion.

Hash table in main memory associated to each sentiment

- nice → 

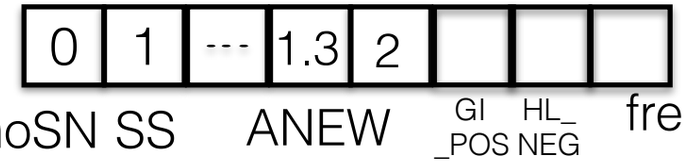
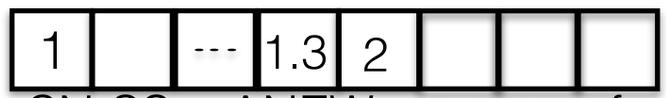
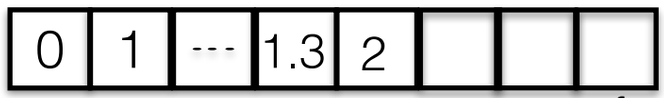
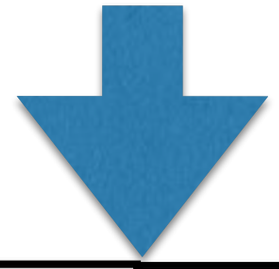
- afraid → 

Table in Oracle associated to each sentiment

- nice → 

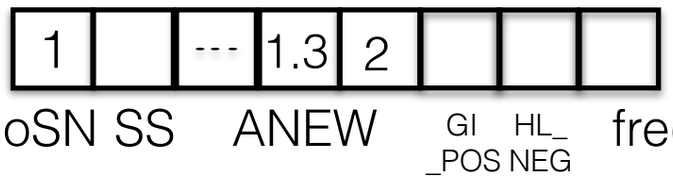
- afraid → 



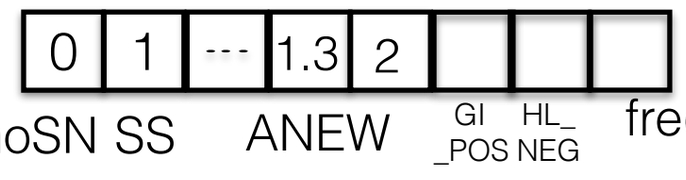
Word	EmoSN	SentiSense	NRC	AFINN	ANEW	…	Frequency
nice	1	0	…	…	1.3	…	123
afraid	0	1	…	…	1.2	…	236
…	…	…	…	…	…	…	…

Collection *words* of JSON documents in MongoDB associated to each sentiment

- nice



- afraid →



```
db.words.insert = {[
  {lemma:“nice”,
  lexical_resources:{
      EmoSN: 1, SentiSense: 1, NRC: 1, GI_POS: 1,
      ANEW:{score:1.3},
      DAL:{arousal: 1.3, dominance: 0.7, pleaseness: 0.5}
   },
   frequency=127},
   {lemma=“afraid”,lexical_resources={…}, frequency=..},
]}
```