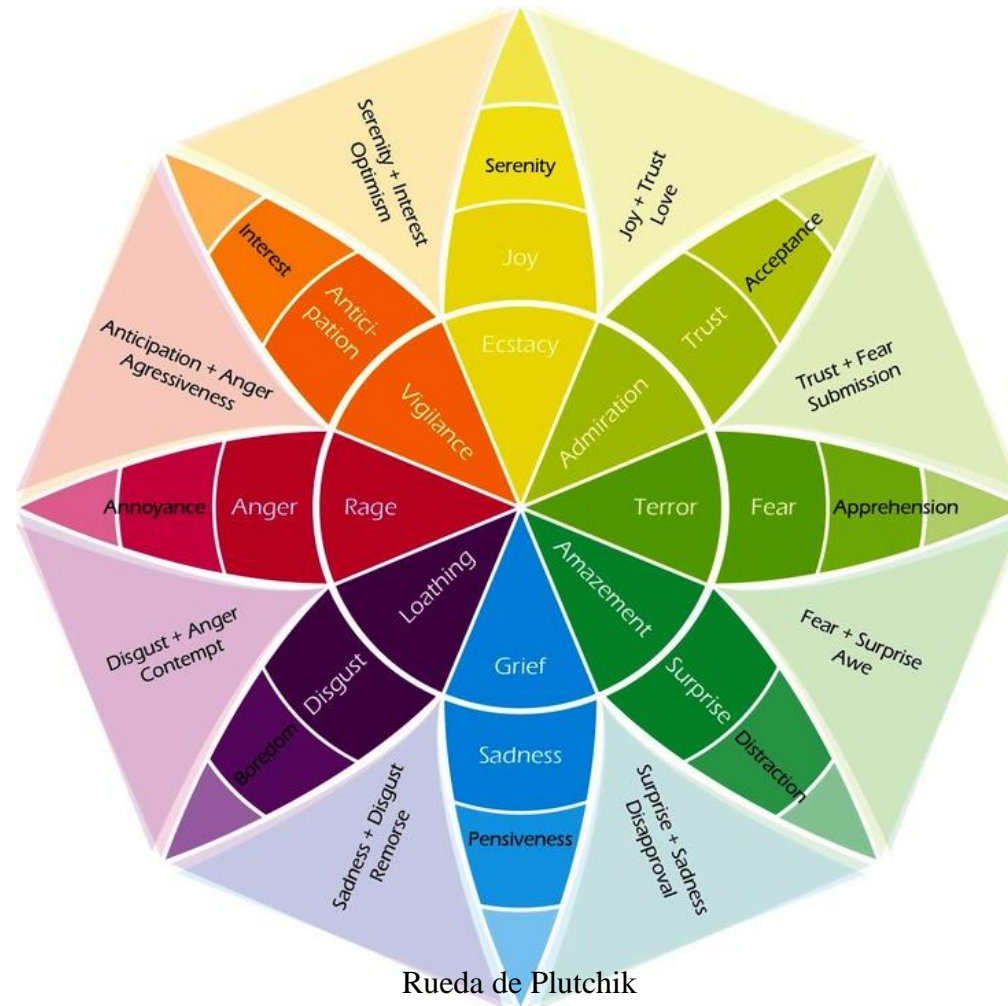
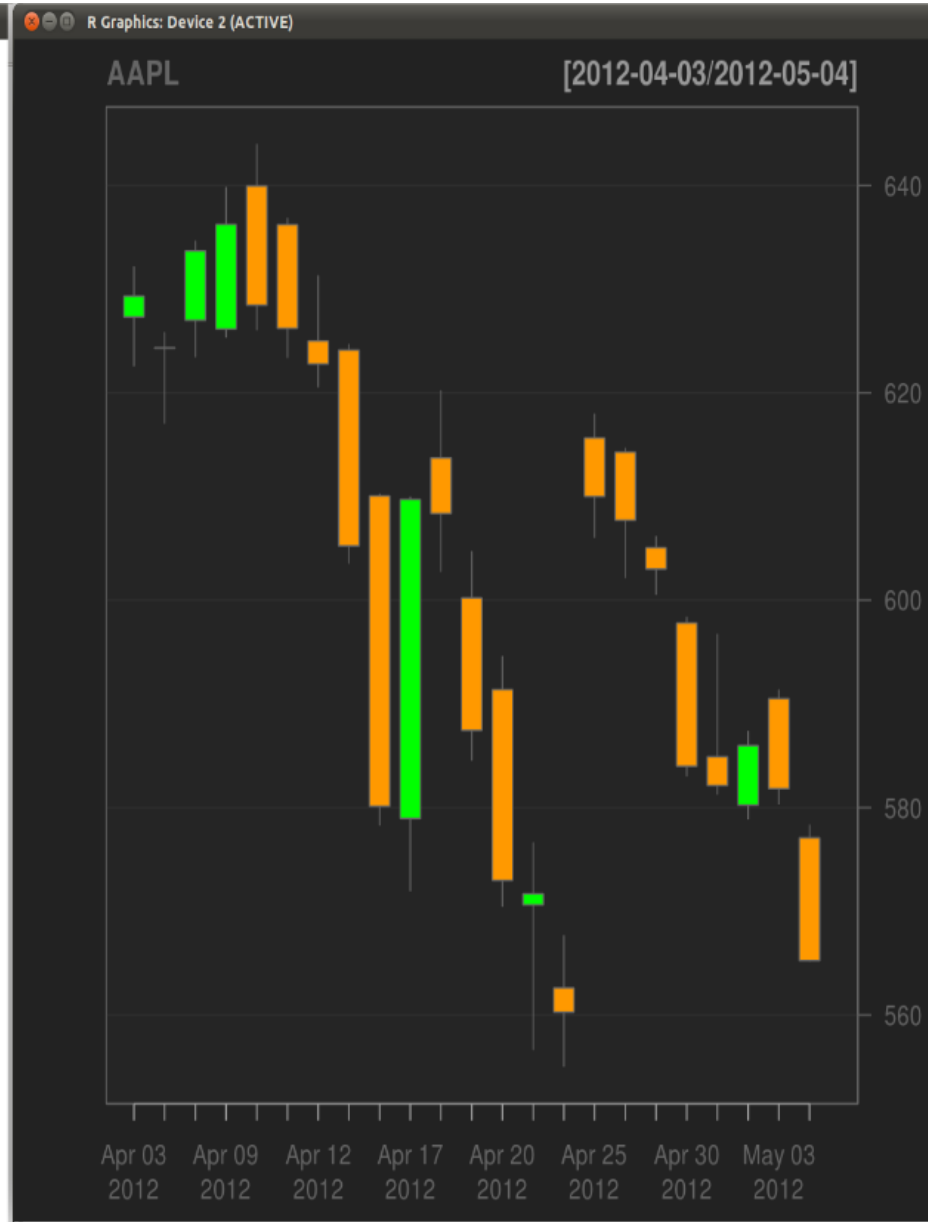
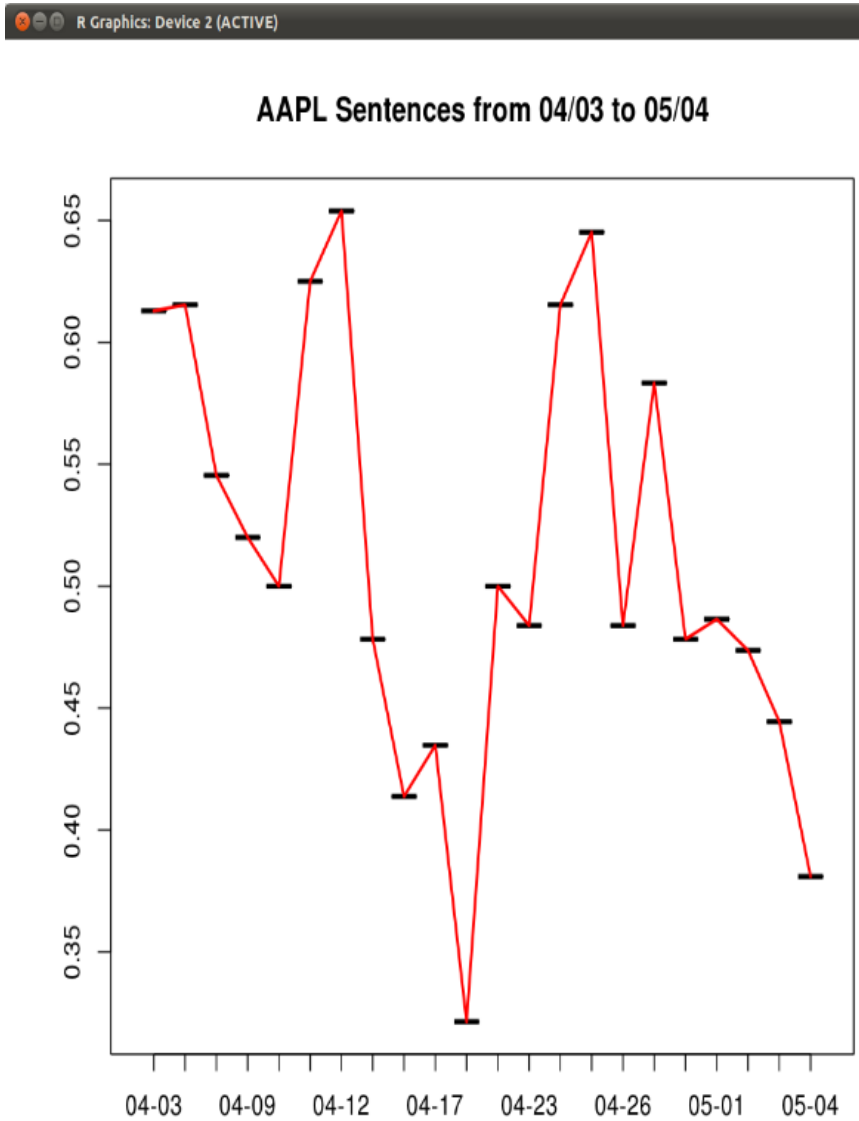


# Análisis de sentimiento aplicado a finanzas



Rueda de Plutchik



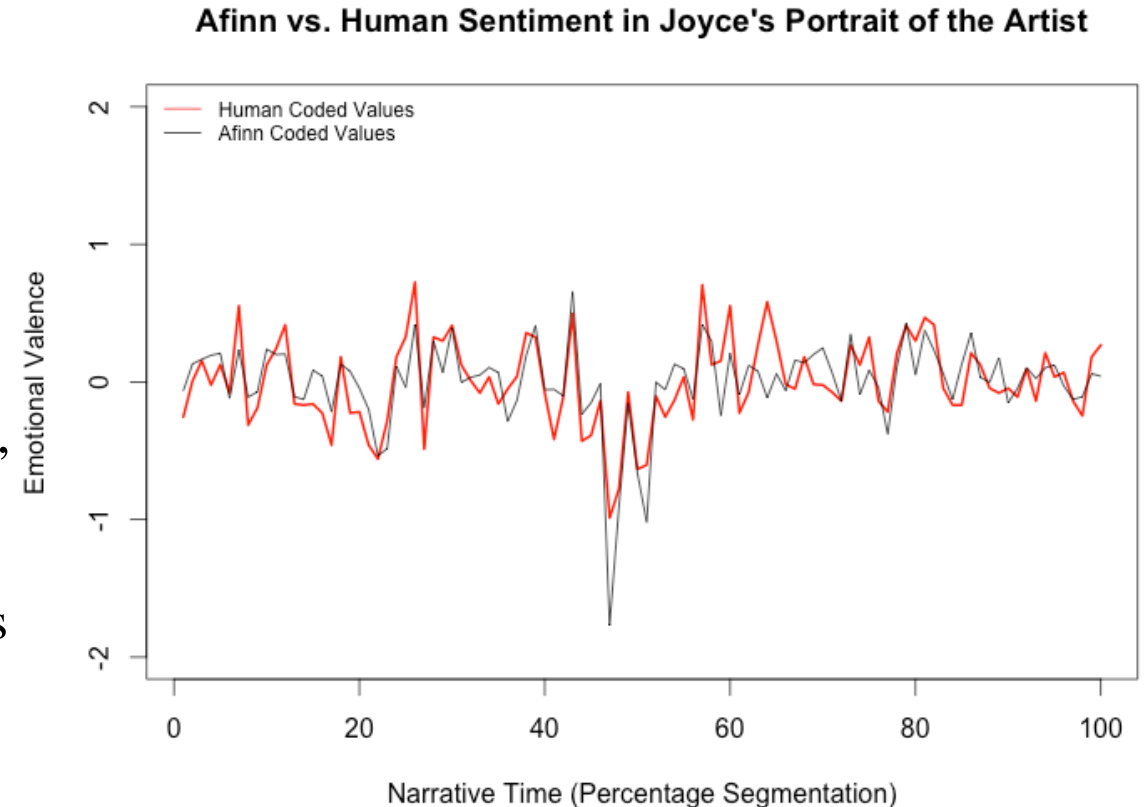
Nagar, A., & Hahsler, M. (2012). Using Text and Data Mining Techniques to extract Stock Market Sentiment from Live News Streams. In *2012 International Conference on Computer Technology and Science* (Vol. 47, pp. 91-95).

# Tipología del problema

- Problema de clasificación:
  - Clasificador bayesiano ingenuo, máxima entropía o SVM (requiere un corpus de entrenamiento)
  - Frecuencia de ocurrencia de palabras (requiere un diccionario)
  - Basado en reglas (requiere un diccionario de sentimiento y de entidades además de un parser)
- Dependencia del dominio léxico
- Problemas propios del lenguaje:
  - Ambigüedades
  - Usos figurados
  - Errores ortográficos
  - Diferencias tipológicas

# Recursos

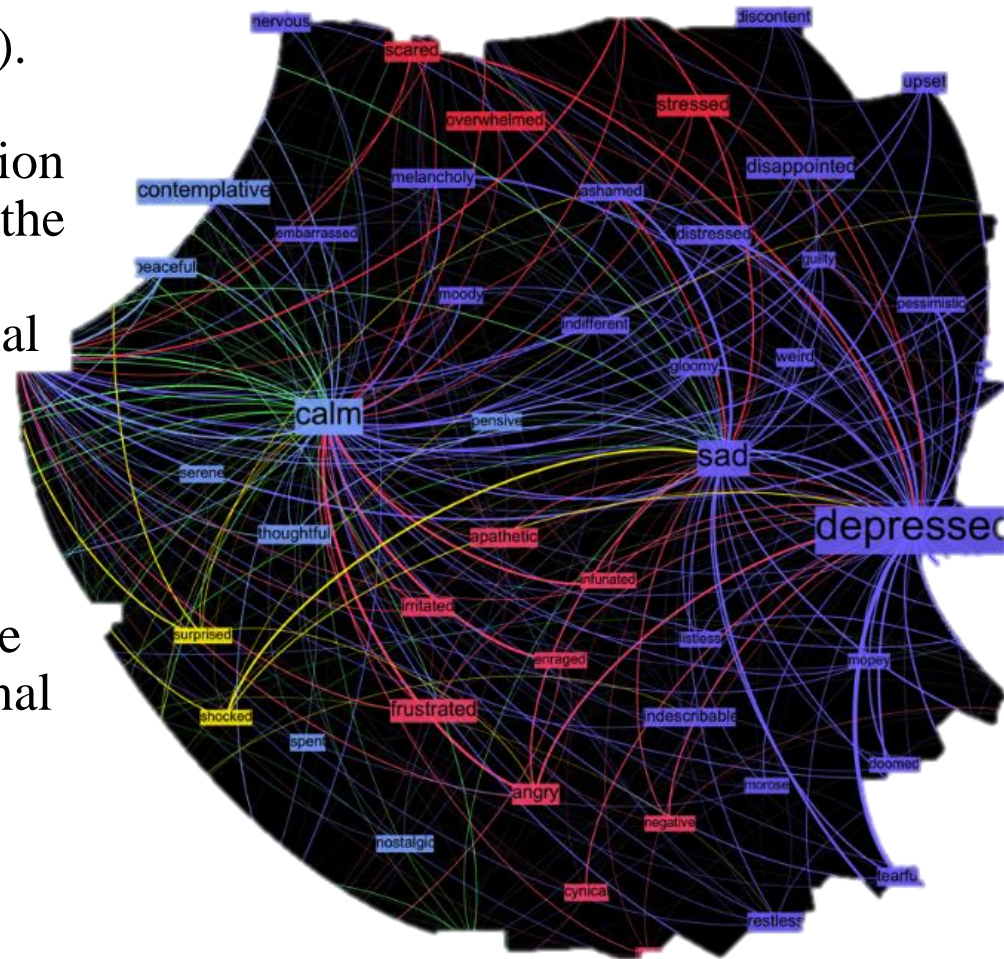
- Diccionarios
  - **Hu and Liu** (ver paquete qdap): 6800 palabras clasificadas como positivas o negativas
  - **AFINN-111**: 2477 palabras clasificadas por una valencia entre -5 y +5
  - **General Inquirer**: 11780 palabras clasificadas según estados mentales (emoción, motivación, orientación cognitiva, referente...)
  - **NRC EmoLex**: 14182 unigramas clasificados en ocho emociones básicas y dos sentimientos.
  - **SentiWordNet**: clasifica los synsets de WordNet en positivo, negativo u objetivo.



<http://www.matthewjockers.net/2015/04/01/my-sentiments-exactly/>

# Recursos

- Coocurrencia y otros patrones sintácticos
  - **Hatzivassiloglou, V., & McKeown, K. R.** (1997, July). Predicting the semantic orientation of adjectives. In Proceedings of the 35th annual meeting of the association for computational linguistics and eighth conference of the european chapter of the association for computational linguistics (pp. 174-181). Association for Computational Linguistics.
  - **Kanayama, H., & Nasukawa, T.** (2006, July). Fully automatic lexicon expansion for domain-oriented sentiment analysis. In Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing (pp. 355-363). Association for Computational Linguistics.



# Transformación

- **Matriz documento-término** (paquete tm):

D1 = "I like databases"

D2 = "I hate databases"

	I	like	hate	databases
D1	1	1	0	1
D2	1	0	1	1

- **Modelo Bag-of-words** (paquetes text2vec y wordVectors):

D1 = "Peter likes to read books. Paul likes too."

D2 = "Peter also likes to read poems"

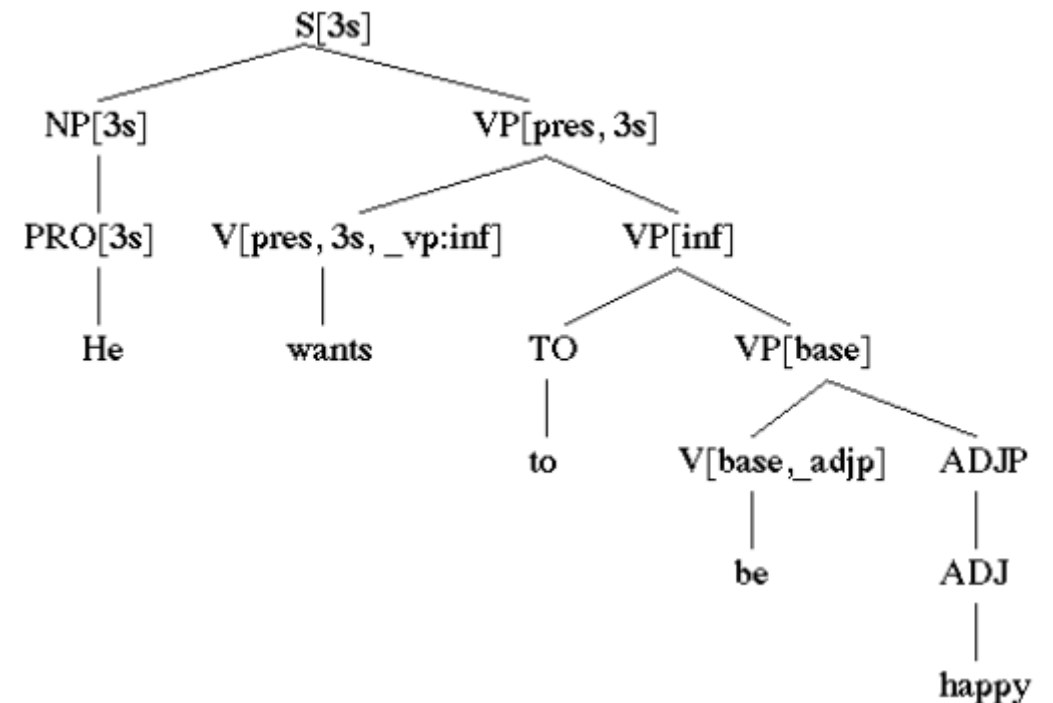
Bag: [ Peter, likes, to, read, books, Paul, too, also, poems ]

Histogram 1: [ 1, 2, 1, 1, 1, 1, 1, 0, 0 ]

Histogram 2: [ 1, 1, 1, 1, 0, 0, 0, 1, 1 ]

- **Árbol sintáctico** (paquete OpenNLP):

- Estadísticos (requieren treebanks)
- Basados en reglas







*That's all Folks!*

- Hu and Liu: <https://www.cs.uic.edu/~liub/FBS/sentiment-analysis.html>
- AFINN: [http://www2.imm.dtu.dk/pubdb/views/publication\\_details.php?id=6010](http://www2.imm.dtu.dk/pubdb/views/publication_details.php?id=6010)
- Generar Inquirer: <http://www.wjh.harvard.edu/~inquirer/homecat.htm>
- NRC EmoLex: <http://saifmohammad.com/WebPages/NRC-Emotion-Lexicon.htm>
- SentiWordNet: <http://sentiwordnet.isti.cnr.it/>
  
- Stanford: <http://nlp.stanford.edu/software/lex-parser.shtml>
- FreeLing: <http://nlp.lsi.upc.edu/freeling/demo/demo.php>