

Ficha 10

23 de Maio de 2015

Desafio: Detecção de língua

Os dados que se encontram em <http://www.di.uminho.pt/~rcm/aulas/spln1415/hmm> contém os seguintes ficheiros:

treino.pos Ficheiro que deve ser utilizado no treino do HMM.

teste.pos Ficheiro que só deve ser utilizado para testar a capacidade do HMM.

Cada ficheiro contém em cada linha uma palavra e a sua tag separada por *whitespace*. As frases são separadas por uma linha em branco.

1. Crie uma pequena biblioteca de HMMs para tagging baseada na aula sobre HMMs.
2. Use o ficheiro de treino para treinar os HMMs
3. Crie uma forma de tratar as palavras não existentes no ficheiro de treino (mais uma vez inspirado no que aprendeu na aula)
4. Teste as suas estratégias sobre os casos de teste