

Língua Natural - MP1 (date2text)

Grupo 52

Hugo Ribeiro (95590)
Murillo Teixeira (105038)

_October_third_,_two_thousand_and_twenty_three

1 Divisão do projeto

O trabalho foi dividido igualmente (Hugo: 50% e Murillo: 50%). Os *transducers* foram implementados após discussão conjunta em um [repositório compartilhado](#).

2 Estratégia de implementação

Focamos no desenvolvimento explícito (por ficheiros `.txt`) apenas de transducers simples e usamos transducers auxiliares para a construção de FSTs mais complexos.

O primeiro exemplo disso é o `mmm2mm.fst`, que lista todos os meses em *mixed condensed format* em inglês para os valores numéricos. O transducer `mix2numerical.fst` foi construído com base no anterior, concatenando-o com um transducer auxiliar que repete números e barras (/) por meio do `fstconcat`. De forma similar, o transducer `pt2en.fst` foi construído a partir de um transducer auxiliar (o `pt2en_months.fst`, que apenas traduz os nomes do inglês para o português) e o mesmo transducer referido anteriormente que repete números e barras. Para construir o `en2pt.fst` bastou fazer a inversão do `pt2en.fst` com o `fstinvert`.

Para a criação do transducer `datenum2text.fst` bastou concatenar, na ordem correta, os transducers `month.fst`, `day.fst` e `year.fst`, definidos explicitamente, juntamente com 2 transducers auxiliares. Esses transducers auxiliares lidam com as barras (/) que aparecem no input e substituem-nas por um ϵ ou uma vírgula.

Para criar o `mix2text.fst` fizemos `fstcompose` do `pt2en.fst` com o `mix2numerical`, uma vez que o input também pode vir em português. De seguida usamos o `fstunion` deste novo transducer com o `mix2numerical.fst`, isto permite que independentemente da língua do input o programa converta o mês para números. Por fim, fazemos novamente um `fstcompose` com o `datenum2text.fst`, o que permite obter o output desejado.

O último transducer pedido no enunciado, `date2text.fst`, foi criado através da operação `fstunion` entre o `mix2text.fst` e o `datenum2text.fst`, visto que estes 2 transducers já estão prontos para tratar todos os formatos de input pedidos.

3 Comentários adicionais

O programa foi desenvolvido em