Relatório Mini Projeto 1 Língua Natural

Grupo 9:

José Guilherme Freitas Gomes Nº 89378

Gonçalo Maria Rodrigues Bustroff Silva Nº 82050

Iniciámos o projeto analisando os exemplos já fornecidos, visto nunca termos trabalhado com o *OpenFST.* Após compreensão do funcionamento do software decidimos, para motivos de facilitação na escrita dos transdutores, começar por desenhar todos os transdutores base (entenda-se, os que não são fruto de nenhuma operação de *software* aplicada a transdutor(es) já existente(s)).

O primeiro transdutor implementado foi o ***mmm2mm.fst****.* Por erro de compreensão da nossa parte, a leitura era efetuada a toda a data e não apenas ao mês, pelo que tanto este como o *dia.fst,* *mes.fst e ano.fst* estavam demasiado complexos sem necessidade. Após leitura do guia *FSTtools* e uma mais cuidada ao enunciado, compreendemos a tarefa. Dito isto, o transdutor em questão foi simplificado e cumpre a tarefa, substituindo o mês (MMM) sempre por 2 dígitos (01 a 12 - MM). Este transdutor foi usado para verificar o mês no ***misto2numerico.fst***. Não foi usada a ferramenta para gerá-lo, a informação foi copiada e adaptada manualmente no .txt correspondente. Mesmo assim foi desafiante visto que decidimos validar parcialmente a data. Não aceitamos dias fora do intervalo 1-31 ou 01-31. Caso o 1º símbolo esteja no intervalo 4-9, o 2º símbolo aceite é apenas “/”. Também, caso o 1º símbolo seja um 0, o 2º encontra-se no intervalo 1-9. Isto são as únicas verificações feitas ao dia, em todos os conversores de datas, sendo que por ex. 31/fevereiro é considerada uma data válida.

O ***en2pt.fst*** inicialmente não aceitava os dias da forma desejada, visto que teve de ser corrigido de igual forma ao transdutor anteriormente referido. A parte relativa à tradução não foi desafiante. O ***pt2en.fst*** foi criado manualmente. Visto ser bastante similar, apenas foi alterado o *Input: Output* dos estados do mês.

O ***dia.fst*** é o nosso transdutor mais complexo visto fazer a verificação do dia e a conversão para texto. A complexidade é principalmente devido aos dias entre 20-31, visto que estes necessitam de um “e” entre a décima e a unidade. Com o *eps* conseguimos modular uma solução. O ***mes.fst*** foi solucionado rapidamente. No ***ano.fst*** tivemos uma ligeira dificuldade em relação aos anos no intervalo 2001-2009. O *dia* e *mês* foram concatenados em *diames.fst* que por sua vez foi concatenado com o *ano,* gerando o ***numerico2texto.fst***, processos feitos no *run.sh*.

O ***misto2texto.fst*** resulta de uma composição do *misto2numerico2* e *numerico2texto2.* O ***data2texto.fst*** resulta de uma união do *misto2texto2* com *numerico2texto.* Ambos foram também criados com as *FSTtools* e desenhados os pdfs com o *Graphviz*.

Testamos todos os transdutores requeridos e obtemos o desejado pelo que consideramos a nossa solução sólida e flexível para todo o intervalo requerido. Apesar disto, sabemos que poderia estar melhor. No início do desenvolvimento a nossa inexperiência com o *FST* fez com que implementássemos manualmente alguns transdutores que poderiam ser resultado de uma simples concatenação de transdutores mais pequenos, mais simples. Com uma melhor gestão de tempo alguns transdutores poderiam ser refeitos, retirando o ligeiro excesso de transdutores.