# [EA876 A] Detecção de elementos em uma nota fiscal eletrônica

Guilherme R C (169127) e João T. de A. F. (146649)

## I. INTRODUÇÃO

Notas fiscais são tipicamente disponibilizadas em *xml*, mas não existe nenhuma padronização nacional para este tipo de documento. O objetivo deste trabalho é construir um programa que realiza o *parsing* de uma Nota Fiscal em *xml*, detectando os campos Município gerador, Município prestador, Valor do serviço e ISS retido, gerando um tabela (em formato *csv*) com tais informações.

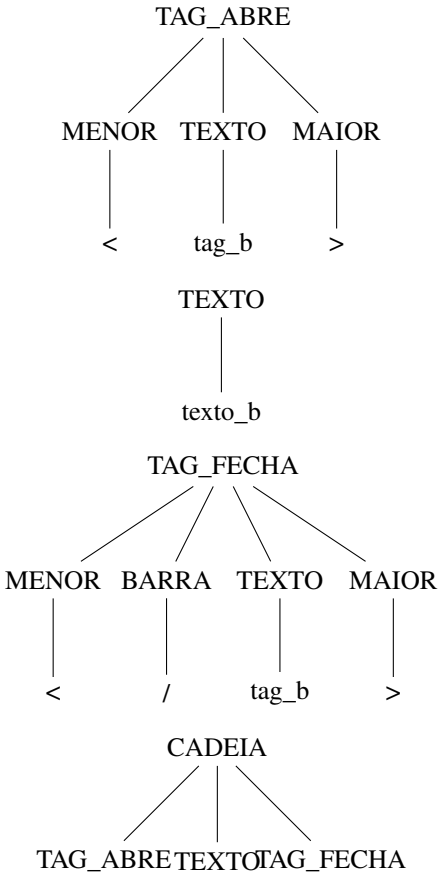
## II. METODOLOGIA

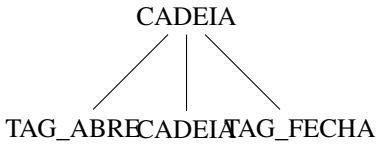
A metodologia adotada para resolução do problema proposto toma como base a estrutura recursiva de um documento *xml*, isto é, a característica de uma cadeia (tag abre, conteúdo, tag fecha) pode ser formada por cadeias.

Suponha que um arquivo *xml* tenha o seguinte conteúdo:

$$< tag\_a >< tag\_b > texto\_b < /tag\_b >< /tag\_a >$$

Primeiramente foi feita a *tokenização* do texto com um analisador léxico, seguida pela construção da árvore sintática do mesmo como demonstrado abaixo com um analisador sintático.




Terminado o *parsing* do *xml*, tratamos agora do reconhecimento das *tags* que indicam as informações da Nota Fiscal requisitadas.

Foram observados dois casos maiores que precisavam de tratamento:

- O primeiro caso é aquele em que a informação pode ser identificada pela simples precedência de uma determinada *tag*.
- O segundo caso é aquele em que a informação só pode ser determinada pelo encadeamento de duas determinadas *tags*.

O tratamento destes casos foi realizado da seguinte forma:
O arquivo *xml* é varrido e assim que uma abertura de *tag* é lida:

1) Procura-se a *tag* recebida em um vetor de *tags* padrão (que resolvem apenas o primeiro caso). Caso ache, o conteúdo da cadeia é o que buscamos e este algoritmo é recomeçado. Caso contrário, 2).
2) Procura-se a *tag* recebida em um vetor de *tags* que contém apenas as *tags* "pais", isto é, as externas de um encadeamento. Caso encontre, as próximas *tags* serão procuradas em um vetor de *tags* que contém apenas as *tags* "filho', isto é, as internas do encadeamento. Quando encontra, o conteúdo da cadeia é o que buscamos e o algoritmo é recomeçado.

Para avaliar o funcionamento do programa, foi implementado um script que itera sobre todos os arquivos *xml* imprimindo as respectivas tabelas *csv*.

## III. RESULTADOS

O programa desenvolvido obteve o resultado esperado para todas as cidades, com exceção de João Monlevade e Rio de Janeiro:

1) Em relação a João Monlevade, a *tag* relativa ao código do município (que é encontrada por um encadeamento de *tags*) aparece vazia em algumas notas. Assim, o programa fica numa busca sem fim pelo conteúdo desta.
2) Não conseguimos identificar o problema para Rio de Janeiro

Caso houvesse mais tempo poderíamos ter contornado tais casos limítrofes.