



ENCODE UTILIZADO NOS ARQUIVOS XML - SICAP ANALISADOR

O que é Encode de Arquivo?

Refere-se ao sistema utilizado para representar caracteres (letras, números e símbolos) em arquivos digitais. Considere o encode como uma "linguagem secreta" que os computadores usam para entender e armazenar texto. Existem diferentes tipos de encode, como: **UTF-8**, **ISO-8859-1**, entre outros. Cada tipo de encode define como os caracteres são convertidos em bytes (unidades de dados) e vice-versa.

Por que o Encode é Importante?

Considere a seguinte analogia: você escreve uma carta em português e a envia para alguém que só fala inglês. Sem tradução adequada, o destinatário não conseguirá entender a carta. O encode desempenha um papel semelhante para computadores, se um arquivo é salvo utilizando um encode e lido em outro, pode resultar em caracteres corrompidos ou ilegíveis. Esse aspecto torna-se particularmente relevante ao transferir arquivos entre diferentes sistemas ou softwares.

Conformidade com Requisitos

O Tribunal de Contas do Estado especifica um tipo de encode que deve ser utilizado ao enviar arquivos. Isso garante que todos os documentos recebidos sejam consistentes e possam ser lidos corretamente pelo sistema do tribunal.

Precisão dos Dados

Ao enviar arquivos para o Tribunal de Contas do Estado, é essencial que todas as informações estejam precisas e legíveis. O tribunal utiliza esses arquivos para monitorar e auditar as finanças públicas. Caso os arquivos contenham caracteres ilegíveis ou corrompidos devido a problemas de encode, a análise e visualização dos dados podem ser comprometidas. O encode esperado pelo Tribunal de Contas do Tocantins é o **ISO-8859-1**.

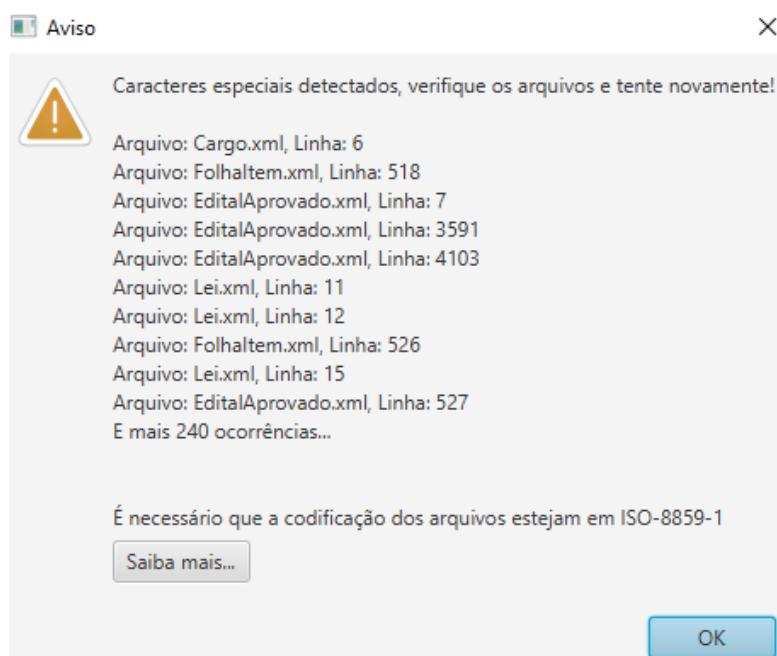


Figura 1. Mensagem de validação do encode

A mensagem acima (Figura 1) é um exemplo de validação que o Analisador irá acusar para casos em que o arquivo não estiver em conformidade com os padrões estabelecidos.

O encode citado anteriormente (**ISO-8859-1**) será validado através do **SICAP Analisador**, que realizará uma varredura no arquivo para verificar se está utilizando o encode correto e se contém caracteres especiais. A Figura 2 abaixo apresenta exemplos de caracteres inválidos.

	A	B	C
1	CIDADE		CIDADE
2	MACAPÃ		MACAPÁ
3	SUMARÃ%		SUMARÉ
4	JUNDIAÃ		JUNDIAÍ
5	CIPÃ"		CIPÓ
6	JAÃS		JAÚ
7	SÃfo PAULO		SÃO PAULO
8	CAÃtAPAVA		CAÇAPAVA
9	ÃÃ%ÃÃ"ÃsÃtÃfÃ,ÃsÃZÃ"Ã		ÁÉÍÓÚÇÃÄÊÎÔ
10			

Figura 2. Exemplos de palavras simples com encode quebrado

Como conferir e ajustar o encode do XML ?

1. Abra o arquivo com algum editor de texto avançado, como por exemplo o [Notepad++](#);
2. Verifique o cabeçalho do arquivo;

```
<?xml version="1.0" encoding="UTF-8" standalone="no"?>
```

3. Em alguns casos o encoding do arquivo está correto, porém os textos estão quebrados, como no exemplo abaixo:

```
<?xml version="1.0" encoding="ISO-8859-1" standalone="no"?>
<Sicap>
<Cargo>
<codigoCargo>5001</codigoCargo>
<nomeCargo>REQUISITADO</nomeCargo>
<atribuicaoCargo>nÃo existente</atribuicaoCargo>
<idCargoNome>11937</idCargoNome>
<regimeJuridico>1</regimeJuridico>
<vinculoCargo>1</vinculoCargo>
<jornadaSemanal>4</jornadaSemanal>
<quantidadeVagas>1</quantidadeVagas>
<natureza>1</natureza>
<escolaridade>5</escolaridade>
<classe>1</classe>
</Cargo>
```

4. Nesses casos em específico apenas o cabeçalho está correto, mas o encoding do arquivo continua em outro formato, o mais comum é que esteja em **UTF-8**.
5. Em ambos os casos é importante verificar como está sendo gerado este XML, para **certificar** que o encoding do arquivo e do texto estejam alinhados e no formato **ISO-8859-1**.