Troubleshooting XML - UML Class Mapping

The XML parser throws a fatal error.

Problem

The XML parser throws error XMLLM/3 with a message similar to:

```
Native XML parser fatal error in "XML_Buffer" line 1, column 39, location :
"invalid byte 'u' at position 2 of a 3-byte sequence"
```

You cannot find a reason for this error in the parsed data.

Solution

This kind of message is nearly always caused by an **invalid encoding** of the input. The XML parser can deal with different encodings but they have to be correctly declared in the XML processing instruction. In most cases when this error occurs, the encoding is declared as UTF-8 but the input (or just a part of it) is in another encoding - in most scenarios windows-125{0,1,2}.

1. To troubleshoot, **always** start with examining the input blob.

    - If you decode it in an editor (e.g. Notepad++), you should immediately see the wrong bytes.
    - On linux, save the blob in question to a file and use the following command:

        ```
        base64 -d xml_blob.b64 | hexdump -C
        ```

2. Then, trace the wrongly-encoded strings to their origin, figure out why they are in another encoding and what you can do properly convert them to UTF-8.

Example

You have the following input blob:

```
PD94bWwgdmVyc2lvbj0iMS4wIiBlbmNvZGluZz0idXRmLTgiPz48cmVjaG51bmdMaXN0ZT48ZGF0
0
ZWlfZGVmX3JlY29yZCBlbXBmYWVuZ2VyQ29kZVBmbGVnZT0iS+R1ZmVyIiBlbXBmYWVuZ2VyT3J0
0
PSJN/G5jaGVuIi8+PC9yZWNobnVuZ0xpc3RlPgo=
```

Decoding the blob shows:

```
<?xml version="1.0" encoding="utf-8"?><rechnungListe><datei_def_record
empfaengerCodePflege="K×E4ufer" empfaengerOrt="M×FCnchen"/></rechnungListe>
```

Here one can see why the parser complains - **e4** (character "ä" in windows-1251) is a beginning of 3-byte sequence in UTF-8 but 75 ("u") is not a valid continuation.

Related Pages:

- XML - UML Class Mapping
- xmlToClass() Operation for Blobs