

Há padrões de conversão de caracteres em binários, e um dos mais comuns está registrado na tabela ASCII (*American Standard Code for Information Interchange*). Temos a letra, e a ela há um número associado, a partir deste, é criada uma sequência binária. Por exemplo, a letra C é representada pelo número 67, e resulta na sequência 01000011.

Contudo, temos um problema pois esta tabela engloba todos os caracteres da língua inglesa, mas como sabemos há outros além destes. Por exemplo, aqueles que são acentuados. Há ainda outros alfabetos, com símbolos diferentes.

Para solucionar este problema, foram criadas as codepages, único formato capaz de englobar a quantidade de informação correspondente ao número de línguas e caracteres existentes.

Para tentar unificar os padrões, foi criado o unicode. Trata-se de uma tabela cujo objetivo é apresentar todos os caracteres existentes no mundo.

Ela também conta com um número associado a cada caractere (codepoint).

Contudo, o *unicode* não define a forma como as informações devem ser armazenadas no HD, isto é tarefa dos encodings. Para representar o encoding existem as classes como é o caso dos "UTFs", como o UTF-8 e UTF-16, esta sigla significa "*Unicode Transformation Format*". Ela está vinculada desde o nascimento com a tabela de Unicode, para traduzir os *codepoints* para um formato binário.

Além do UTF há outros exemplos de *Encodings*, como o ASCII e o Windows 1252.

É importante termos a informação de qual é o charset padrão pois é ele que define como traduzir o codepoint em uma sequência de bits e bytes.

