

1, 2, 3, кодирај! – Научни осврт – Представљање информација

Идеја представљања неког објекта, конкретног или апстрактног, другим објектом је врло стара, јер се и сви језици заснивају на репрезентацији објеката помоћу гласа. [Пројект « 1, 2, 3... кодирај! » предлаже више активности, у циклусу 3, секвенца III-3.]

Објекта је сувише много да би сваки био представљен различитим знаком, зато су они представљени низом знакова одабраних у неком малом скупу. На пример, класични француски језик има 36 слова, или фонема, помоћу којих се формира реч. Порука у неком језику је коначан низ одабраних јединки у оквиру неког коначног скупа. Ова идеја, међутим, није потпуно реализована у говорном језику где јачина, интонација, начин говора...дају знатно више значења поруци него сам низ јединки које је чине.

Колико информација садржи неки текст?

Писањем, а нарочито његовом употребом код језичког изражавања, елиминишу се ови помоћни елементи па оно, бар у прво време представља неку врсту регресије јер се првим писмима језика продружује различит симбол сваком објекту. Само се код азбуке појављује идеја придруживања више јединки неком објекту, ипак се тај низ јединки своди на мали скуп. На пример, француска азбука садржи 110 карактера: 42 мала слова (26 основних слова, 13 акцентованих самогласника, са квачицом и 2 лигатуре), исто толико великих слова, 10 цифара, и двадесетак знакова интерпункције. (На часу српског језика направите сличну анализу за ћирилично и латинично писмо! п.п.) Број низова од n карактера које можемо формирати бирајући их у некој азбуци која има k елемената је k^n , тј., за први карактер има k могућности, k могућности за други, ... и k могућности за n -ти, па је укупан број могућности $k * k * \dots * k$ (n пута) тј., k^n .

Што је тај број већи, то и сваки од ових низова садржи више информација. На пример, неки текст од 4 000 карактера одабран међу 110 карактера француске азбуке садржи више информација од неког текст од два карактера одабрана од три могућа.

Јединица информације се назива **бит**.

Често, из историјских разлога, користимо другу јединицу информација попут октета који садржи 8 бита, као и њихове мултипликације: килооктет (ко) који има хиљаду октета, или осам хиљада бита, мегаоктет (Мо) који има милион октета, гигаоктет (Го) који има милијарду октета, тераоктет (То) који има хиљаду милијарди октета, петаоктет (Ро) који има милион милијарди октета.

Квантитет информација садржан у неком карактеру азбуке је реда октета. На пример, страница текста од око 2000 карактер је реда 2 килооктета, нека књига од шесто страна је реда 1 мегаоктета, нека мала библиотека од хиљаду књига је реда 1 гигаоктета. Национална библиотека Француске која има 14 милиона књига (не узимају се у обзир слике и филмови...) је реда 1 тераоктета (прецизније 14 тераоктета). Један тераоктет је капацитет диска који се може купити за десетак евра. Један петаоктет одговара величини од сто Националних библиотека Француске. У Европском центру за нуклеарна истраживања- CERN (centre européen pour la recherche nucléaire) се сваке године произведе података реда петнаест петаоктета.

<< [Les réseaux](#)

[Eclairages](#)

[Le codage binaire](#) >>

Extrait de "[1, 2, 3... codez !](#)", Editions Le Pommier, 2016-2017. Publié sous licence [CC by-nc-nd 3.0](#).