

**Instituto Politécnico Nacional**

*Unidad Profesional Interdisciplinaria en Ingeniería y Tecnologías Avanzadas*

Teoría de la información

**Practica 1**

Entropía del idioma

**Profesor**

Jorge Rojas Beltrán

**Grupo**

2TM4

**Alumno**

Alvarado Balbuena Jorge Anselmo

2019/02/05

# Información del texto

* **Libro:** Fahrenheit 451.
* **Autor:** Ray Bradbury.
* **Publicación:** 1953.
* **Lengua original:** inglés.
* **Cantidad de caracteres inglés:** 236676.
* **Cantidad de caracteres francés:** 270673.

# Análisis del texto en inglés sin memoria

|  |  |  |
| --- | --- | --- |
| Letra | Probabilidad | Cantidad de información [bits] |
| a | 0.06364397 | 3.97383238 |
| b | 0.01273471 | 6.29509019 |
| c | 0.01639372 | 5.93071296 |
| d | 0.03743937 | 4.73930009 |
| e | 0.09894117 | 3.33728525 |
| f | 0.01682469 | 5.89327639 |
| g | 0.02050483 | 5.60789275 |
| h | 0.05312748 | 4.23439784 |
| i | 0.05508797 | 4.18211893 |
| j | 0.00096757 | 10.0133501 |
| k | 0.00959117 | 6.70407731 |
| l | 0.03637885 | 4.78075637 |
| m | 0.02191604 | 5.51186924 |
| n | 0.05934273 | 4.07478489 |
| o | 0.06340313 | 3.97930205 |
| p | 0.01207135 | 6.37226857 |
| q | 0.000638 | 10.6141492 |
| r | 0.04227721 | 4.56397616 |
| s | 0.04703899 | 4.40999911 |
| t | 0.07551251 | 3.72714042 |
| u | 0.02363569 | 5.40288935 |
| v | 0.00656171 | 7.25171178 |
| w | 0.02165408 | 5.5292176 |
| x | 0.00070983 | 10.4602365 |
| y | 0.01674441 | 5.90017664 |
| z | 0.00052392 | 10.8983576 |
| espacio | 0.18633491 | 2.42403014 |

Tabla 1 Información del texto sin memoria.

Entropía H = **4.09439709** bits/símbolo

# Fuente con memoria – Markov de 1° orden

|  |  |  |  |
| --- | --- | --- | --- |
| Casos más frecuentes | | | |
| Parejas | **Apariciones** | **Probabilidad** | **Cantidad de información [bits]** |
| e-espacio | 9022 | 0.0596642 | 4.06699062 |
| espacio-t | 6872 | 0.15582413 | 2.68200943 |
| he | 6223 | 0.49491013 | 1.01476152 |
| th | 6044 | 0.33818263 | 1.56412553 |
| d-espacio | 5548 | 0.62611443 | 0.67550173 |
| espacio-a | 4934 | 0.11187955 | 3.15998175 |
| t-espacio | 4886 | 0.27338854 | 1.87097532 |
| s-espacio | 4534 | 0.4072577 | 1.29598611 |
| in | 3876 | 0.29728486 | 1.7500821 |
| espacio-s | 3368 | 0.07637015 | 3.71084733 |

Entropía H = **3.315047606** bits/símbolo

# Fuente con memoria – Markov de 2° orden

|  |  |  |  |
| --- | --- | --- | --- |
| Casos más frecuentes | | | |
| Tercias | **Apariciones** | **Probabilidad** | **Cantidad de información [bits]** |
| Espacio-th | 4773 | 0.69455763 | 0.5258337 |
| the | 4085 | 0.6758769 | 0.56516758 |
| he-espacio | 3918 | 0.62959987 | 0.66749285 |
| nd-espacio | 2112 | 0.82532239 | 0.27697031 |
| and | 1950 | 0.59760956 | 0.74272486 |
| espacio-an | 1943 | 0.39379814 | 1.34447181 |
| ing | 1833 | 0.47291022 | 1.08036178 |
| ed-espacio | 1830 | 0.8857696 | 0.17499661 |
| ng-espacio | 1700 | 0.78959591 | 0.34081357 |
| espacio-he | 1248 | 0.37910085 | 1.3993464 |

Entropía H = **2.637543717** bits/símbolo

# Análisis del texto en francés sin memoria

|  |  |  |
| --- | --- | --- |
| Letra | Probabilidad | Cantidad de información [bits] |
| a | 0.06547384 | 3.93293765 |
| b | 0.00841975 | 6.89200625 |
| c | 0.02331596 | 5.42253855 |
| d | 0.02893898 | 5.11084222 |
| e | 0.12212892 | 3.03352328 |
| f | 0.00900718 | 6.79470905 |
| g | 0.00895546 | 6.80301747 |
| h | 0.00689023 | 7.18123154 |
| i | 0.06195668 | 4.01259639 |
| j | 0.00440014 | 7.82823376 |
| k | 0.00014778 | 12.7242634 |
| l | 0.05061827 | 4.30419788 |
| m | 0.02632328 | 5.24751716 |
| n | 0.05736812 | 4.12360706 |
| o | 0.04490658 | 4.47692919 |
| p | 0.02154629 | 5.53641645 |
| q | 0.00957613 | 6.70634146 |
| r | 0.05521053 | 4.17891272 |
| s | 0.06808215 | 3.87657958 |
| t | 0.06009465 | 4.05661956 |
| u | 0.05318595 | 4.23281107 |
| v | 0.01488881 | 6.06962733 |
| w | 0.00014778 | 12.7242634 |
| x | 0.00333243 | 8.22920784 |
| y | 0.00247162 | 8.66032906 |
| z | 0.00205414 | 8.92725039 |
| espacio | 0.19055835 | 2.39169528 |

Entropía H = **3.972733948** bits/símbolo

# Fuente con memoria – Márkov de 1° orden

|  |  |  |  |
| --- | --- | --- | --- |
| Casos más frecuentes | | | |
| Parejas | **Apariciones** | **Probabilidad** | **Cantidad de información [bits]** |
| e-espacio | 13751 | 0.41597846 | 1.26541926 |
| s-espacio | 9334 | 0.50651183 | 0.98133213 |
| t-espacio | 7119 | 0.43766138 | 1.19211301 |
| espacio-l | 5676 | 0.11004479 | 3.18383731 |
| espacio-d | 5655 | 0.10963764 | 3.18918487 |
| le | 4726 | 0.34493833 | 1.53558966 |
| es | 4429 | 0.1339807 | 2.8999029 |
| en | 4182 | 0.12650876 | 2.98269084 |
| espacio-s | 4100 | 0.07948971 | 3.65308799 |
| re | 3951 | 0.26438704 | 1.91927661 |

Entropía H = **3.236088231** bits/símbolo

# Fuente con memoria – Márkov de 2° orden

|  |  |  |  |
| --- | --- | --- | --- |
| Casos más frecuentes | | | |
| Tercias | **Apariciones** | **Probabilidad** | **Cantidad de información [bits]** |
| es-espacio | 3375 | 0.76202303 | 0.3920935 |
| espacio-de | 3017 | 0.53351017 | 0.90641233 |
| espacio-le | 2507 | 0.44168428 | 1.17891259 |
| it-espacio | 2354 | 0.80506156 | 0.31282899 |
| le-espacio | 2335 | 0.49407533 | 1.01719708 |
| de-espacio | 2325 | 0.64727171 | 0.62755663 |
| nt-espacio | 2149 | 0.5723036 | 0.80514742 |
| ent | 1880 | 0.44954567 | 1.1534604 |
| espacio-qu | 1867 | 1 | 0 |
| re-espacio | 1771 | 0.44824095 | 1.15765363 |

Entropía H = **2.649635545** bits/símbolo

# Tabla comparativa entre entropías

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Caracteres | Entropía sin memoria | 1° Orden | 2° Orden |
| Inglés | 236676 | 4.09439709 | 3.315047606 | 2.637543717 |
| Francés | 270673 | 3.972733948 | 3.236088231 | 2.649635545 |

# Conclusiones

En la tabla anterior se puede ver como al aumentar el orden para la cadena de Márkov la entropía disminuye

Es interesante ver como la entropía de ambos lenguajes es bastante cercana entre sí.