

Pràctica d'expressions regulars sobre corpus

Tria d'una mostra segons l'ordre a l'aula

<http://retoc.iula.upf.edu/CGIs/RetocSC.cgi?sc=11®istre=1>

Un cop cada estudiant tindrà la seva mostra i haurà creat un document que es denominarà com a pràctica 3, amb el NIA de l'estudiant, anirem aplicant col·lectivament les REGEX següents, cadascú a la seva mostra i copiarem alguns dels resultats al nostre document, o deixarem constància de no haver-ne obtingut. A més, compararem amb un dels companys per veure si en la seva mostra ha tingut més o menys èxit. Ho comentarem al document i resumirem la funció de la cerca.

Expressions regulars que cercarem:

<code>\W(?:beure cabell caden cell corretj crest deute empènyer estret ovell paret rell sec.\W seques\W peraW peres\W)</code> (l això?)	Troba qualsevol caràcter no alfanumèric seguit de <i>beure, cabell, caden, cell, corretj, corretg, crest, deute, empènyer, estret, ovell, paret, rella, sec</i> (seguit de punt= comodí d'un sol caràcter qualsevol i caràcter no alfanumèric, de fet <i>seca</i> o <i>secs</i>), <i>seques</i> (seguit de caràcter no alfanumèric), i <i>pera</i> o <i>peres</i> (seguit de caràcter no alfanumèric). El grup ?: fa que el parèntesi no sigui retingut en memòria.
<code>[àèèíóú][^aeiouàèèíóú]{1,3}i(?:a es)</code> (l això?)	Troba una vocal accentuada seguida de 1, 2 o 3 consonants, seguida/es de <i>ia</i> o <i>ies</i> .
<code>(?:[s\(\-][ptb][ae]r[aeiouàèèíóú] safare verit esper interess)</code> <code>^[ptb][ae]r[aeiouàèèíóú]</code>	Troba una consonant d'entre p, t i b seguida de les vocals a o e, seguides de r intervocàlica, o bé els radicals de <i>safareig, veritat, esperit</i> o <i>interessant</i> .

<p>(l això?)</p>	<p>La primera variant comença amb <code>[s\(\[-]</code>, o sigui més o menys com <code>[W]</code>.</p> <p>La segona variant comença amb el símbol <code>^</code> que en aquest context vol dir inici absolut de línia.</p>
<p><code>(?:n[tg] [td])mp)s?\$</code> <code>(?:n[tg] [td])mp)s?[W]</code></p> <p>(l això?)</p>	<p>Troba <code>nt</code> o <code>ng</code>, o <code>lt</code> o <code>ld</code>, o <code>mp</code>; tots ells seguits d'una essa opcional (opcionalitat marcada per l'interrogant).</p> <p>El <code>\$</code> de la primera variant significa final absolut de línia.</p> <p>La segona variant va seguida de caràcter no alfanumèric.</p>
<p><code>[aeiouàèòéíóú][bg][eia]</code></p> <p>(l això?)</p>	<p>Troba una vocal accentuada o no accentuada seguida de <code>b</code> o <code>g</code>, seguides de <code>le</code>, <code>li</code> o <code>la</code>.</p>
<p><code>[W](?:ad ab ob)s[ct]</code></p> <p>(l això?)</p>	<p>Troba un caràcter no alfanumèric seguit de (<code>adsc</code> o <code>adst</code>) o de (<code>absc</code> o <code>abst</code>) o de (<code>obsc</code> o <code>obst</code>).</p>
<p><code>[W]ps</code> <code>^ps</code></p> <p>(l això?)</p>	<p>Troba el grup <code>ps</code>, en la primera variant precedit d'un caràcter no alfanumèric; en la segona variant, a l'inici absolut de línia.</p>
<p><code>[W](?:ex[aeiouàèòéíóúh])èx(?:it ode))</code></p> <p>(l això?)</p>	<p>Troba un caràcter no alfanumèric seguit d'<code>ex</code>, seguit de vocal accentuada, o de vocal no accentuada, o de la consonant <code>h</code>. O bé troba caràcter no alfanumèric seguit d'<code>èxit</code> o <code>èxode</code>.</p>
<p><code>[sx][W][aeiouàèòéíóúh]</code></p> <p>(l això?)</p>	<p>Troba <code>s</code> o <code>x</code>, seguit de caràcter no alfanumèric, seguit de vocal accentuada, o de vocal no accentuada, o de la consonant <code>h</code>.</p>
<p><code>[W\s](?:abrev acaric afil agrac agrem ajustic amp anestes anunc apar aprec aprop assoc avar avic balaf bat benefic bescanv cablegraf calumn canv comerc concil conferenc </code></p>	<p>Caràcter no alfanumèric seguit d'espai, seguit de qualsevol dels segments següents, seguit de la vocal <code>i</code>, seguida de les desinències verbals dels presents rizotònics, seguides de</p>

congen|congrac|congr|contag|contrar|cop|custod|denunc|
deprec|descanv|desengab|desgrac|desprestig|diferenc|diligenc|
diluv|dissoc|distanc|divorc|elog|emmurr|encof|encom|ender|
engab|enjudic|enrab|enunc|esbarr|esgarr|espif|espol|estalv|
estud|evidenc|exil|expatr|exprop|extas|extrav|extremunc|fol|
fotocop|fotograf|glor|humil|incend|influenc|inic|injur|intercanv|
interfol|inventar|irrad|llicenc|malf|malic|negoc|notic|obsequ|od|
ofic|pal.l|parod|plag|prelud|prem|presag|presenc|pronunc|
quintaessenc|rad|reconcil|ref|refug|regrac|renunc|repatr|repud|
reverenc|rum|sac|sentenc|som|sogn|substanc|tap|telegraf|
testimon|vanaglor|var|vic|vilipend|i(?:[aoiïr][re][sn])[\W\w]

caràcter no alfanumèric o seguit d'espai.

Recuperarà mots com *estudio*, *diferencia*, *avariïs*, *estudiïn*,
pronuncien.

[\(I això?\)](#)