



PREFEITURA MUNICIPAL DE VÁRZEA PAULISTA
CHEFIA DE GABINETE

Supervisão Departamental de Informática e Gestão de Redes

prómenino/GT

Pesquisa Fonética

Fundação Telefônica - São Paulo, 7 de Agosto de 2007

Metaphone

- ✎ É um algoritmo fonético para a geração de chaves (índices) de palavras.
- ✎ Palavras com sons parecidos geram chaves iguais.
- ✎ Criado por **Lawrence Philips** em 1990 como forma de suprir as deficiências do algoritmo **Soundex**.
- ✎ O **Metaphone** para o inglês está implementado em PHP.
- ✎ Uma versão em PHP para o espanhol foi desenvolvida por **Israel J. Sustaita** em 2005.
- ✎ A implementação para a língua portuguesa é baseada nesta versão.

Metaphone em português

- ✦ A língua portuguesa possui 19 sons consoantes:
“b”(bacia), “d”(conde), “f”(foca), “g”(fogo), “j”(gente), “k”(doca), “l”(lote),
“λ”¹(alho), “m”(maçã), “n”(cana), “ŋ”²(nhoque), “p”(copo), “r”(tora),
“R”(arroz), “s”(centro), “t”(ponte), “v”(vila), “x”(chão), “z”(asa)
- ✦ Os fonemas “λ”, “R” e “ŋ” são substituídos pelos números “1”, “2” e “3”.
- ✦ Exceto quando está no início das palavras as vogais são descartadas (a língua portuguesa é consonantal).
- ✦ A consoante “h” é descartada, exceto nos dígrafos.
- ✦ As letras “k”, “w” e “y” são convertidas nas suas correspondentes fonéticas.

Metaphone em português

- Casos específicos com as consoantes “c”, “g”, “s”, “x” etc, encontros consonantais e os dígrafos são tratados em seus casos especiais.
- Consoantes duplicadas que não alteram o fonema são ignoradas.
- Faz o tratamento de alguns casos que não existem (mais) na língua portuguesa: “ph”, “th” e “sch”
- Ao contrário da versão em espanhol, faz uso de expressões regulares (REGEX) para a substituição de caracteres.

Expressões regulares definem um padrão a ser utilizado para a procura ou substituição em palavras (ou em grupos delas). é um modo preciso e rápido para a pesquisa em textos.

Exemplo de chaves metafônicas

AS ARMAS E OS BARÕES ASSINALADOS,
QUE DA OCIDENTAL PRAIA LUSITANA,
POR MARES NUNCA DE ANTES NAVEGADOS,
PASSARAM AINDA ALÉM DA TAPROBANA,
EM PERIGOS E GUERRAS ESFORÇADOS,
MAIS DO QUE PROMETIA A FORÇA HUMANA,
E ENTRE GENTE REMOTA EDIFICARAM
NOVO REINO, QUE TANTO SUBLIMARAM;

AS A₂MS E OS BRS A_SNLDS
K D O_SDNTL P₂ L_ZTN
P₂ MRS NNK D ANTS NVGDS
P_SRM AND ALM D TP₂BN
EM PRGS E G₂S ESF₂S₂DS
MS D K P₂MT A F₂S U_MN
E ENT₂ J_NT J₂MT EDFK_RM
NV J₂N K TNT SBLMRM

Fontes de Consulta

- ▣ Verbetes Metaphone na Wikipédia
<http://en.wikipedia.org/wiki/Metaphone>
- ▣ Manual da função *metaphone()* do PHP
http://www.php.net/manual/pt_BR/function.metaphone.php
- ▣ Metaphone em PHP para a língua espanhola
http://www.geocities.com/isloera/spanish_methaphone.txt
- ▣ Metaphone em Ruby
<http://po-ru.com/projects/metaphone/>
- ▣ Gramática e Dicionário On-Line da Priberam Informática:
<http://www.priberam.pt>

Desenvolvimento

✦ Responsáveis:
Rodrigo Lotierzo
Giovanni Nunes

Pesquisa, Codificação e Depuração

Pesquisa (,palpites) e Depuração

✦ Estagiários:
Caio Schleich
Diego de Souza

Tecnologia em Informática - FATEC Jundiaí

Ciências da Computação - UNIP Jundiaí

✦ Ajuda técnica com a língua portuguesa:
Silvia Regina Siqueira
Thaiza de Souza Silva

SMF / estudante de licenciatura em língua portuguesa

SMG / professora de português (cedida)

Pesquisa de nomes via chave metafônica

- ✎ Computadores interpretam as letras como números:
CESAR = 67 69 83 65 82
CEZAR = 67 69 90 65 82
- ✎ Como 83 não é 90 então “CESAR” não é igual a “CEZAR”!
- ✎ A chave metafônica para **Cesar** é “SZ2”, independente de como é escrito (os casos latinos de “æ” viraram “e”).
- ✎ No banco de dados também ficam armazenadas as chaves metafônicas dos nomes.
- ✎ Para facilitar a pesquisa os “**da/das**”, “**de**”, “**do/dos**” nos nomes são removidos antes da criação das chaves.

Conclusão

- ✎ Como se escreve mesmo o nome?

Cesar Luiz Figueiredo de Olivera “SZ2LSFGRDOLVR”

César Luís de Figuerêdo Oliveira “SZ2LSFGRDOLVR”

Ao todo são 64 combinações distintas mas apenas uma chave metafônica

- ✎ Uma pesquisa relativamente complexa é facilmente implementada com um comando simples em SQL:

```
SELECT (...) FROM (...)
```

```
WHERE chave_metafone LIKE "SZ2%";
```

- ✎ Soluções mais sofisticadas podem ser implementadas, um ponto de partida é:

http://en.wikipedia.org/wiki/Category:Algorithms_on_strings

SECRETARIA MUNICIPAL DE
CIDADANIA E ASSISTÊNCIA SOCIAL

Entrando na Rede

prómenino/GT

Pesquisa Fonética



Prefeitura Municipal de Várzea Paulista
Av. Fernão Dias Paes Leme, 284 - Centro
Várzea Paulista - SP - 13220-900
TEL/FAX (11) 4596-9614



prómenino