## A intenção de se aplicar um algoritmo de fonética é poder assimilar palavras, no exemplo deste trabalho, falaremos sobre nomes e sobrenome. A aplicação do algoritmo abaixo é de um caso real dentro de uma base de dados de uma seguradora.

## Mostraremos porque as técnicas conhecidas não resolveriam determinados problemas. Essa técnica não foi aplicada de maneira isolada no caso real, que possui diversas técnicas e a construção de uma engine própria que tornou possível todo objetivo que é o agrupamento de pessoas similares afim de, evitar fraudes bem como a sua detecção.

**Esse algoritmo foi testado de supervisionados até alcançar 97% de assertividade na base da seguradora. O mesmo algoritmo já possui algumas oportunidades de melhoria para possibilitar que o nível de assertividade aumente.**

# “Lógica Funcional” - Fonética

A “Lógica Funcional” da Fonética tem como objetivo gerar uma chave para os nomes (Pessoa Física) que estão relacionados ao sinistro (evento/acidente). Após a geração dessa chave é possível buscar os nomes por similaridade. *Exemplo: Felipe – Felipi.*

A “Lógica Funcional” da Fonética será aplicada na base (ENGINE) como em todos os dados de entrada (input).

A “Lógica Funcional” da Fonética deve estar sendo sempre observada, pois existe a possibilidade de que uma nova situação não esteja sendo coberta pelo conjunto de regras. Porém essa observação deve ser realizada mediante a um estudo e necessita de um aval para que possa ser incluída nas regras atuais.

## Fonética – Criação da Chave Fonética.

Esse item tem o objetivo de descrever de maneira estruturada e inteligível, quais são os passos necessários para a criação da chave fonética.

### Extrair preposição e abreviação.

Excluir todas as preposições e abreviaçõesmantendo cada palavra do nome separada individualmente com um espaço (somente um carácter de espaço). Ao final desse passo temos a palavra tratada sem preposições e abreviações (nome sem preposições).

*Ex.: JOSE DA SILVA DOS SANTOS*

*Resultado esperado: JOSE SILVA SANTOS*

### Contar a quantidade de Palavras.

Contar a quantidade de palavras após a execução do passo 2.1.1. Nessa contagem será possível obter a quantidade de repetições de código dentro do algoritmo até que ele chegue ao final (repetições do laço).

*Ex.: JOSE SILVA SANTOS*

*Quantidade de palavras = 3*

### Tratar palavra por palavra.

Devido ao fato de que cada palavra pode cair em uma condição diferente de regras, trataremos as palavras de forma separada. O passo 2.1.2 indica quantas vezes o algoritmo será executado, ou seja, uma execução para cada palavra até que seja executado pela última vez.

*Ex.: JOSE SILVA SANTOS*

*1ª palavra: JOSE*

*2ª palavra: SILVA*

*3ª palavra: SANTOS.*

### Excluir as letras repetidas da palavra tratada.

Excluir as letras repetidas (“RR”; “SS”; “AA” e etc.). A aplicação desse tratamento possibilita um nível de aproximação entre as palavras.

*Ex.: FELLIPPE*

*Resultado esperado: FELIPE*

### Identificar vogais e consoantes da palavra tratada.

Identificar as vogais e consoantes de cada palavra, pois cada letra da palavra receberá um tratamento específico. Nos passos adiante veremos a aplicação desse tratamento.

*Ex.: JOSE*

*Posição 0: letra J – Consoante*

*Posição 1: Letra O – Vogal*

*Posição 2: Letra S – Consoante*

*Posição 3: Letra E - Vogal*

### Aplicar o tratamento sobre as consoantes da palavra tratada.

Aplicação do tratamento de todas as consoantes da palavra tratada tendo como base o seguinte quadro de regras:

|  |  |
| --- | --- |
| B | Sempre irá ser “B”; |
| C | Sempre será “C”, se a letra posterior for “A, O, U, L ou R”. Quando a letra “C” anteceder as letras “E, I e Y” ganhará a chave fonética de “S”. Se o “C” tiver a cedilha, a chave fonética será “S”. |
| D | Sempre será “D”. |
| F | Sempre será “F”. |
| G | Sempre será “G”, quando a letra posterior for “A, O, U, L ou R”, e será considerado “J” quando a letra posterior for “E, I ou Y”. |
| H | Se o “H” vier posterior a letra “P” o que formaria um “PH” ganharia a chave fonética de “F”. Se a letra “H” vier posterior a um “C” ou a um “S” formando um dígrafo “CH ou SH”, ganharia a chave fonética de “X”. Se o “H” vier depois de um “N” ou de um “L”, formando o dígrafo “NH ou LH”, ganhará a chave fonética de “I”. Obs.: Nos dígrafos “NH e LH” as consoantes “N” e “L” não são desprezadas nas chaves fonéticas, nos demais são dígrafos.  Desprezar a letra “H” quando ela estiver entre consoantes. |
| J | Sempre será “J”. |
| K | Sempre será “C”. |
| L | Sempre será “L”. |
| M | Sempre será “M”. |
| N | Sempre será “N”. |
| P | Sempre será “P”. |
| Q | Sempre será considerada letra “C”. |
| R | Sempre será “R”. |
| S | Sempre será “S”. |
| T | Sempre será “T”. |
| V | Sempre será “U”. |
| X | Sempre será “X”. |
| W | Sempre será “U”. |
| Y | Sempre será “I”. |
| Z | Sempre será “S”. |

### Aplicar o tratamento a letra “e” na palavra tratada.

Substituir todas as letras “e” por “i” exceto quando o “e” vier precedido de “i” ou sucedido de “u”. A aplicação desse tratamento possibilita um nível de aproximação entre as palavras.

*Ex.: CLOVES*

*Resultado esperado: CLOVIS*

### Aplicar o tratamento a letra “H” na palavra tratada.

Quando a letra “H” for a 1ª letra da palavra tratada, registrar em uma variável (y) o valor 1. Caso a letra “H” não seja a 1ª letra da palavra tratada registrar em uma variável (y) o valor 0. Nos passos adiante veremos a aplicação dessa variável.

*Ex.: ELDER*

*Resultado esperado: y = 0*

*Ex.: HELDER*

*Resultado esperado: y = 1*

### Excluir as letras repetidas da palavra tratada. Repetição do passo 2.1.4.

Excluir as letras repetidas (“RR”; “SS”; “AA” e etc.). A aplicação desse tratamento possibilita um nível de aproximação entre as palavras. A repetição desse passo se deve, pois os passos anteriores podem trazer novas repetições.

*Ex.: FELLIPPE*

*Resultado esperado: FELIPE*

### Calcular a quantidade de letras da palavra tratada.

Contar a quantidade de letras que uma palavra possui e registrar em uma variável (x). Nos passos adiante veremos a aplicação dessa variável.

*Ex.: JOSE*

*Resultado esperado: x = 4*

### Aplicar o tratamento do encontro de vogais na palavra tratada.

Trazer todos os encontros de vogais contidos na palavra para a chave fonética. Quando houver um encontro de vogais com mais de 2 vogais sequenciadas trazer as duas últimas letras do encontro para chave fonética.

*Ex.: JOAO*

*Resultado esperado: “AO”*

### Aplicar o tratamento das variáveis de apoio na palavra tratada.

Subtrair o valor da variável de apoio (x) do passo 2.1.4 pelo valor da variável de apoio (y) do passo 2.1.9. A aplicação desse tratamento irá registrar uma nova quantidade de letras para a palavra tratada. Passo fundamental para o tratamento aplicado no passo seguinte.

*Ex.: 2.1.6: JOSE - x=4 ; 2.1.9: JOSE - y=0*

*Resultado esperado: x-y = 4;*

*Ex.: 2.1.6: HELDER - x=6 ; 2.1.9: HELDER - y=1*

*Resultado esperado: x-y = 5;*

### Aplicar o tratamento da quantidade de letras na palavra tratada.

A aplicação desse tratamento objetiva através, da “nova” quantidade de letras da palavra (tratamento do passo 2.1.12), extrair as vogais para a chave fonética da palavra tratada tendo como base o seguinte quadro de regras:

**Três letras ou menos** 🡪 Caso durante o tratamento do passo 2.1.11 **não tiver** nenhum encontro de vogais extraídos para a chave então, **trazer a 1ª vogal** da palavra tratada para a chave fonética.

*Ex.: ANE*

*Vogal extraída = A*

*Ex.: RUI*

*Vogal extraída = não tem extração de vogal nesse caso*

**Quatro letras** 🡪 Caso durante o tratamento do passo 2.1.11 **não tiver** nenhum encontro de vogais extraídos para a chave então, **trazer a 1ª vogal** da palavra tratada para a chave fonética.

*Ex.: JOSE*

*Vogal extraída = O*

*Ex.: CAIO*

*Vogal extraída = não tem extração de vogal nesse caso.*

**Cinco letras** 🡪 Caso durante o tratamento do passo 2.1.11 **tiver** encontro de vogais extraídos para a chave então, não será necessária mais nenhuma extração de vogal para a chave fonética.

Caso durante o tratamento do passo 2.1.11 **não tiver** nenhum encontro de vogais extraídos para a chave então, dividir a “nova” quantidade de letras da palavra (tratamento do passo 2.1.12) por 2 com arredondamento para baixo.

* Caso na primeira metade da palavra (duas primeiras letras) **tiver** um encontro de consoantes então, **trazer a última vogal** da palavra tratada para a chave fonética.
* Caso na primeira metade da palavra (duas primeiras letras) **não** **tiver** um encontro de consoantes então, **trazer a 1ª vogal** da palavra tratada para a chave fonética.

*Ex.: FABIO*

*Vogal extraída = não tem extração de vogal nesse caso.*

*Ex.: BRUNO*

*Verifique as duas primeiras posições:* ***BR***

*Teve encontro de consoantes:* ***SIM***

*Vogal extraída = “O”*

*Ex.: TARSO*

*Verifique as duas primeiras posições:* ***TA***

*Teve encontro de consoantes:* ***NÃO***

*Vogal extraída = “A”*

**Seis letras** 🡪 Caso durante o tratamento do passo 2.1.11 **tiver** encontro de vogais extraídos para a chave então, não será necessária mais nenhuma extração de vogal para a chave fonética.

Caso durante o tratamento do passo 2.1.11 **não tiver** nenhum encontro de vogais extraídos para a chave então, será efetuado o seguinte tratamento:

* Caso na palavra tratada, a **última letra** **seja** uma **consoante** então, **trazer a 1ª vogal** da palavra tratada para a chave fonética.
* Caso na palavra tratada, a **última letra** **não seja** uma **consoante** então, **trazer a última vogal** da palavra tratada para a chave fonética.

*Ex.: SERGIO*

*Vogal extraída = não tem extração de vogal nesse caso.*

*Ex.: NILSON*

*A última letra é uma consoante:* ***SIM***

*Vogal extraída = “I”*

*Ex.: CARINA*

*A última letra é uma consoante:* ***NÃO***

*Vogal extraída = “A”*

**Sete letras ou mais** 🡪 **Independente** do resultado dos **passos anteriores**, **trazer a última vogal** da palavra tratada para a chave fonética.

Ex.: Palavra MARIANE

*Vogal extraída = “E”*

### Tratar a próxima palavra.

Nesse passo somamos mais 1 ao laço de repetição, caso esse número seja menor que a quantidade de palavras (passo 2.1.2) então, reiniciamos o processo com a nova palavra a ser tratada. Senão passamos para o próximo passo (2.1.15).

### Retornar a chave fonética.

Após o tratamento de todas as palavras, retornar a chave fonética para o conjunto de palavras solicitado.

Segue abaixo o exemplo prático de como é feita a aplicação da “Lógica Funcional” de Fonética:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **CRIAÇÃO DA CHAVE FONÉTICA** | | | | | | |
|  |  |  |  |  |  |  |
| **Dado de entrada (Input) = JOSSEPH ARCADIANO DA CRUZ** | | | | | |  |
|  |  |  |  |  |  |  |
|  | **REGRAS** |  | **RESULTADO DAS REGRAS** | | |  |
|  | **Regra** |  | **Chave Fonética** | | |  |
|  | 2.1.1. Extrair preposição e abreviação |  | JOSEPH ARCADIANO CRUZ | | |  |
|  | 2.1.2. Contar a quantidade de Palavras. |  | 3 Palavras | | |  |
|  | 2.1.3. Tratar palavra por palavra. |  | JOSSEPH | ARCADIANO | CRUZ |  |
|  | 2.1.4. Excluir as letras repetidas da palavra tratada. |  | JOSEPH | não aplicado nesta palavra | não aplicado nesta palavra |  |
|  | 2.1.5. Identificar vogais e consoantes da palavra tratada. |  | JSSPH consoantes ; OE vogais | RCDN consoantes ; AAIAO vogais | CRZ consoantes ; U vogais |  |
|  | 2.1.6. Aplicar o tratamento sobre as consoantes da palavra tratada. |  | JOSEF | ARCADIANO | CRUS |  |
|  | 2.1.7. Aplicar o tratamento a letra “e” na palavra tratada. |  | JOSIF | ARCADIANO | CRUS |  |
|  | 2.1.8. Aplicar o tratamento a letra “H” na palavra tratada. |  | y = 0 | y = 0 | y = 0 |  |
|  | 2.1.9. Excluir as letras repetidas da palavra tratada. Repetição do passo 2.1.4. |  | não aplicado nesta palavra | não aplicado nesta palavra | não aplicado nesta palavra |  |
|  | 2.1.10. Calcular a quantidade de letras da palavra tratada. |  | x = 6 | x = 9 | x = 4 |  |
|  | 2.1.11. Aplicar o tratamento do encontro de vogais na palavra tratada. |  | não aplicado nesta palavra | IA | não aplicado nesta palavra |  |
|  | 2.1.12. Aplicar o tratamento das variáveis de apoio na palavra tratada. |  | x - y = 6 | x - y = 9 | x - y = 4 |  |
|  | 2.1.13. Aplicar o tratamento da quantidade de letras na palavra tratada. |  | JOSF | RCDIANO | CRUS |  |
|  | 2.1.14. Tratar a próxima palavra. |  | passar para a palavra seguinte | passar para a palavra seguinte | não aplicado nesta palavra |  |
|  | 2.1.15. Retornar a chave fonética |  | JOSF RCDIANO CRUS | | |  |
|  |  |  |  |  |  |  |