### Temas

Os temas abaixo foram criados a partir das perguntas enviadas pelos pesquisadores interessados em consultar o DHBB[[1]](#footnote-1), e serão usados para testar a metodologia. A incorporação dos metadados (gênero e cargos ocupados) aos resultados foram particularmente úteis, permitindo o cruzamento com as informações extraídas.

1. **Com que idade o indivíduo iniciou sua carreira pública?**

* Extração dos trechos contendo informação sobre nascimento;
* Metadados contendo as linhas de cargos e respectivos períodos de cada exercício;

1. **Qual a formação acadêmica do político?**

* Extração dos trechos contendo informação sobre formação acadêmica;

1. **O que dizer sobre os vínculos familiares na política?**

* Extração dos trechos contendo informação sobre vínculos familiares com outros personagens também políticos (isto é, outros ‘verbetados’).

Certamente a maioria desses dados são encontrados atualmente na Internet, em sites como os da Câmara[[2]](#footnote-2), do Senado[[3]](#footnote-3) e outras fontes não oficiais. O exercício tem o intuito de medir a recuperação destas informações no DHBB.

## 5.4.3

### Padrões

Esta é a etapa onde: i) observamos em uma amostra de verbetes as construções das frases que trazem a informação desejada, ii) traduzimos estas construções em padrões léxico-sintáticos combinando expressões regulares e etiquetas de anotação, iii) concatenamos todas as expressões (quando houver mais de uma) e iv) aplicamos no corpus.

Para cada tema, apresentamos algumas frases de exemplo e as expressões criadas.

#### Sobre dados de nascimento

De longe, é a informação mais direta de se obter pois sua escrita obedece a um certo padrão: ela é constituída da data e local do evento, localiza-se sempre no primeiro parágrafo, é precedida do nome do biografado e seguida dos nomes dos pais, quando mencionados. Criamos a expressão abaixo para abranger estes casos:

|  |  |
| --- | --- |
| **NASCIMENTO** | |
| Frases de exemplo | **«Moroni Bing Torgan» nasceu em Porto Alegre, no dia 10 de junho de 1956.** |
| «Álvaro Francisco de Sousa» nasceu no dia 28 de fevereiro de 1903. |
| Expressão | [classe="bio.\*" & dicionario="dhbb" & pos="PROP.\*"]+ [: pos!="PROP.\*" :][]{0,1} [lema="nascer" & word!="nascido|nascer"] [pos="PRP.\*"][]{0,21} [pos="NUM.\*|ADJ.\*"] [word="de"]? [pos="N.\*"]? [word="de"]? [pos="NUM.\*"]? [pos="PU"] |
| Ocorrências | 6.464 |

Basicamente, a sintaxe quer dizer: “faça apenas para verbetes biográficos do DHBB, encontre todas as construções onde exista um [nome próprio composto], sucedido do [lema ‘nascer’] (que não pode ser ‘nascido’ nem ‘nascer’) e de uma [preposição]. Siga até encontrar o primeiro [número] ou [modificador] da sentença (dia, ano ou mês) e continue até a próxima vírgula ou ponto final; neste intervalo pode ou não haver outro [número] (eventualmente, a outra parte da data)”.

Abrindo o terminal CQP (Corpus Query Processor), executamos o script que acessa o corpus, aplicamos a expressão de busca e gravamos todas as ocorrências que se encaixam nela em um arquivo que nomeamos como cqpNascimento.txt.

Em paralelo, realizamos algumas operações no R: i) a partir dos arquivos originais dos verbetes, extraímos os metadados e criamos um dataframe com as seguintes colunas: ID, nome do verbete, sexo e cargos (com respectivos períodos de ocupação); ii) abrimos o arquivo gerado no CQP (“cqpNascimento.txt”) e extraímos o ano de nascimento do biografado; iii) incluímos a informação no dataframe criado anteriomente; iv) identificamos a data associada ao primeiro cargo público ocupado pelo indivíduo e calculamos a idade correspondente à época.

Abaixo um extrato das informações extraídas:

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| id | verbete | sexo | nasc |  | id | verbete | sexo | nasc |
| 2 | Nilo de Sousa Coelho | m | 1920 |  | 42 | João Nogueira Adeodato | m | 1902 |
| 5 | John Abbink | m | 1890 |  | 43 | João Adil de Oliveira | m | 1907 |
| 6 | José João Abdalla | m | 1903 |  | 44 | Álvaro Adolfo da Silveira | m | 1882 |
| 8 | Ibrahim Abi-Ackel | m | 1927 |  | 45 | Osório Adriano Filho | m | 1929 |
| 9 | Alarico Abib | m | 1937 |  | 49 | Emiliano Estanislau Afonso | m | 1881 |
| 10 | Armando Abílio Vieira | m | 1944 |  | 50 | Manuel Afonso de Melo Neto | m | 1943 |
| 11 | Junot Abi-Ramia Antônio | m | 1932 |  | 51 | João Agripino Filho | m | 1914 |
| 13 | Elias Abraão | m | 1941 |  | 52 | Francisco Lacerda de Aguiar | m | 1903 |
| 14 | João Abraão Sobrinho | m | 1907 |  | 53 | Anésio Frota Aguiar | m | 1901 |
| 15 | Cláudio Abramo | m | 1923 |  | 54 | Hugo Aguiar | m | 1928 |
| 16 | José Abrão | m | 1945 |  | 55 | Jefferson de Aguiar | m | 1913 |
| 17 | Moisés Abrão | m | 1945 |  | 56 | João Aguiar | m | 1893 |
| 18 | Pedro Abrão Júnior | m | 1958 |  | 57 | Nélson Alves Aguiar | m | 1940 |
| 20 | Dorival Masci de Abreu | m | 1933 |  | 58 | Rafael de Sousa Aguiar | m | 1906 |
| 21 | Fernando de Abreu | m | 1884 |  | 59 | Ubiratan Diniz de Aguiar | m | 1941 |
| 22 | Hugo de Andrade Abreu | m | 1916 |  | 60 | Wilson de Sousa Aguiar | m | 1917 |
| 23 | João Batista de Abreu | m | 1943 |  | 62 | Darcílio Aires Raunheitti | m | 1924 |
| 24 | João d'Abreu | m | 1888 |  | 63 | Ernâni Airosa da Silva | m | 1915 |
| 25 | João Leitão de Abreu | m | 1913 |  | 65 | Jaeder Soares de Albergaria | m | 1904 |
| 26 | José Masci de Abreu | m | 1944 |  | 66 | Albérico de França Ferreira | m | 1950 |
| 28 | Ovídio Xavier de Abreu | m | 1898 |  | 68 | Nion Albernaz | m | 1930 |
| 29 | Paulo Abreu | m | 1912 |  | 69 | Álvaro Alberto da Mota | m | 1889 |
| 30 | Sílvio de Andrade Abreu | m | 1913 |  | 70 | Armanda Álvaro Alberto | f | 1892 |
| 34 | Nicolas C. Accame | m | 1880 |  | 71 | Carlos Alberto de Sousa | m | 1945 |
| 35 | Átila Monteiro Aché | m | 1888 |  | 72 | João Alberto Lins de Barros | m | 1897 |
| 36 | Otávio Monteiro Aché | m | 1890 |  | 73 | Luís Alberto Martins de | m | 1947 |
| 37 | Samir Achoa | m | 1933 |  | 76 | Carlos César Silva de | m | 1940 |
| 40 | Elias Adaime | m | 1929 |  | 77 | Carlos de Albuquerque Filho | m | 1927 |
| 41 | Francisco Mendes Adeodato | m | 1927 |  | 81 | João Pessoa de Albuquerque | m | 1930 |

Com as informações consolidadas, conseguimos agora fazer cruzamentos com os metadados e separar as ocorrências por categorias (gênero, cargos, idade etc.), conforme veremos no próximo capítulo.

1. Conforme explicado na seção 1.1, sobre as principais motivações da pesquisa. A lista completa das perguntas encontra-se no Anexo 1. [↑](#footnote-ref-1)
2. https://www.camara.leg.br/ [↑](#footnote-ref-2)
3. https://www12.senado.leg.br/hpsenado [↑](#footnote-ref-3)