

Perguntas já respondidas sobre o DHBB no AC/DC

Suemi Higuchi

17 de novembro de 2020 – versão 3.0

Documento redigido por Suemi Higuchi com a colaboração de Diana Santos e Cláudia Freitas

Sumário

Visão geral sobre o DHBB:

- [Quem é responsável pela produção e curadoria do DHBB?](#)
- [Quando foi lançado e quantas atualizações já sofreu?](#)
- [Por que muitos verbetes apresentam conteúdo desatualizado?](#)
- [Como era/é a base de dados na sua origem?](#)
- [Como se deu a conversão para arquivos de texto puro e a constituição do corpus?](#)
- [Onde estão estes arquivos de texto do DHBB? Eles estão disponíveis publicamente?](#)

Sobre o processo de integração do DHBB no AC/DC:

- [Como se deu a integração do corpus no AC/DC?](#)
- [Quais são os atributos \(metadados\) específicos do corpus DHBB?](#)
- [Qual a diferença entre entidade e fonte, não se referem ambos ao biografado?](#)
- [O que é o processo de identificação \(grounding\), para que serve e como está sendo feito?](#)
- [Como são tratadas as segmentações erradas?](#)
- [Por que ao realizar busca por parte de um lema \(usando o caractere coringa “*”\), recuperamos linhas de concordância repetidas onde em uma linha a sentença mostra em negrito o termo da busca e em outra\(s\) linha\(s\), outra\(s\) parte\(s\) do mesmo nome?](#)
- [Por que a distribuição dos cargos retorna strings aparentemente sem sentido?](#)
- [Sobre as categorias semânticas usadas no DHBB e a criação de léxicos](#)
- [Onde se encontra a informação semântica?](#)
- [Estão contando colocar mais informação?](#)
- [Como é feito o enriquecimento do corpus com estas informações adicionais?](#)

Sobre como usar o AC/DC (com o DHBB):

- [O que é possível perguntar ao DHBB no AC/DC?](#)
- [Lista de cargos presentes nas biografias](#)

Visão geral sobre o DHBB

1. Quem é responsável pela produção e curadoria do DHBB?

O DHBB é constituído por uma seleção de verbetes cuja produção e curadoria é de responsabilidade do Centro de Pesquisa e Documentação de História Contemporânea do Brasil da Fundação Getúlio Vargas¹. Trata-se de um dos três dicionários histórico-biográficos concebidos com o propósito de fornecer informações sistematizadas sobre personagens e temas considerados relevantes do período e contexto ao qual se referem: 1) Dicionário histórico-biográfico da Primeira República (DHBPR), com dados sobre o período que vai da Proclamação da República, em 1889, até a Revolução de 1930; (2) Dicionário Histórico-Biográfico Brasileiro (DHBB), sobre o período pós-1930 até os dias atuais; e (3) Dicionário da política republicana do Rio de Janeiro (DPRRJ), sobre a história do estado e da cidade do Rio de Janeiro desde a Proclamação da República.

2. Quando foi lançado e quantas atualizações já sofreu?

O DHBB foi lançado em 1984 contendo 4.493 entradas distribuídas em 4 volumes impressos. Sua segunda edição, em 2001 abrangeu 6.620 entradas em 5 volumes e ganhou também uma versão em cd-rom. A terceira atualização foi em 2010, com a disponibilização de todo o conteúdo na Internet, totalizando 7.553 entradas. Desde então a atualização se mantém contínua (na medida do possível) tendo sido acrescentados novos verbetes, chegando em um total de cerca de 7,8 mil entradas até o momento.

3. Por que muitos verbetes apresentam conteúdo desatualizado?

O trabalho de redação e atualização dos verbetes é realizado por especialistas e uma pequena equipe do CPDOC. Depende de recursos captados e aportados pela instituição, e por isso não ocorre de forma contínua. Além disso, como os registros da base de dados de origem do DHBB e os arquivos convertidos em corpus são instâncias separadas (veja abaixo mais informações), é necessário fazer as conversões conforme elas acontecem. Mas na medida do possível, tentamos manter os conteúdos da base oficial e do AC/DC sincronizados.

4. Como era/é a base de dados na sua origem?

O sistema do acervo do CPDOC, datado de 2001, abrange três conjuntos documentais: documentos de arquivos pessoais, entrevistas de história oral e verbetes do Dicionário. A base do DHBB só foi integrada em 2009 e possui poucos metadados, basicamente o id, o título do verbete, o tipo (temático ou biográfico) e o campo para o corpo do texto. Neste último, o conteúdo foi migrado a partir de arquivos do Word, capturando junto a formatação produzida pelo software, o que demandou um processo de limpeza mais cuidadoso no momento de constituição do corpus.

A consulta a essa base pode ser feita aqui: <http://www.fgv.br/cpdoc/acervo>

5. Como se deu a conversão para arquivos de texto puro e a constituição do corpus?

Com vistas ao processamento automático do DHBB, em 2013 foi iniciado um estudo para a conversão dos verbetes em um formato mais adequado, e decidiu-se pelo markdown e YAML, que resultam em textos puros com marcações simples. Cada verbete se tornou um arquivo identificado pelo id da base (que é um número sequencial) e ganhou uma área de metadados em seu cabeçalho. As fontes bibliográficas foram separadas do texto, em arquivos próprios. A revisão manual de cada entrada permitiu a normalização dos metadados

¹ <http://cpdoc.fgv.br/acervo/dhbb>

da seguinte forma: primeiro, a indicação do título trazido da base, em seguida o sexo do biografado seguido dos cargos dispostos um por linha (mantendo-se as abreviaturas adotadas na origem), e por fim a autoria do texto, quando houvesse. Exemplo de como ficou o cabeçalho de metadados:

```
---
título: TOFFOLI, José Antônio
natureza: biográfico
sexo: m
cargos:
- adv. geral União 2007-2009
- min. TSE 2009-
- min. STF 2009-
autor:
- Leonardo Sato
---
```

6. Onde estão estes arquivos texto do DHBB? Eles estão disponíveis publicamente?

Sim. Todos os verbetes estão no Github (<https://github.com/cpdoc/dhbb>). Dois repositórios foram criados: o repo dhbb e o repo dhbb-nlp, sendo este último uma cópia do primeiro para fins de processamento. No dhbb-nlp os cabeçalhos de metadados estão separados dos textos em si e as referências bibliográficas foram suprimidas (esta é a versão enviada para o AC/DC).

Toda atualização de conteúdo é feita no repo dhbb e depois replicada automaticamente no dhbb-nlp.

Perguntas sobre o processo de integração do DHBB no AC/DC

7. Como se deu a integração do corpus no AC/DC?

Uma versão anotada desta coleção de textos foi criada a partir da análise pelo PALAVRAS e também de outros processamentos comuns do projeto AC/DC². A estruturação de cada verbete – agora em um formato pseudo-xml – faz a separação de parágrafos e frases, e atribui marcas sintáticas e semânticas aos tokens. O processamento inclui tarefas para lidar com a tokenização de certos fenômenos - como contrações, clíticos e MWEs – e também para incorporar metadados/atributos aos verbetes³.

O resultado é um corpus contendo a identificação das palavras (word), lemas (lema), categorias gramaticais (pos), tempos verbais (temcagr), pessoa e número (pessnum), função sintática (func) e algumas informações semânticas para os nomes (sema). Uma descrição destas informações se encontram no próximo tópico.

O endereço do DHBB no AC/DC é: <https://www.linguateca.pt/aceso/corpus.php?corpus=DHBB>.

E este é o corpus anotado: <https://www.linguateca.pt/aceso/corpos/dhbb/dhbb.anotado.txt>

Temos no momento cerca de 7.800 verbetes no sistema, que podem ser conhecidos aqui: <https://www.linguateca.pt/aceso/entidadesDHBB.html>.

² O AC/DC é um sistema que usa o Open Corpus Workbench (CWB), voltado para interrogação de corpora linguisticamente anotados e enriquecidos com informação semântica. As consultas retornam concordâncias, distribuições e frequências simples e complexas e são feitas através da linguagem CQP (Corpus Query Processor) através de uma combinação de expressões regulares e propriedades linguísticas atribuídas ao corpus.

³ Para saber mais sobre o processo de entrada de um corpus no AC/DC, leia o seguinte artigo: <https://www.linguateca.pt/Diana/download/SantosBickLREC2000.pdf>.

8. Quais são os atributos (metadados) específicos do corpus DHBB?

Os atributos específicos do DHBB no AC/DC, associados a cada palavra ou unidade, são os seguintes:

- **dicionário** -> como mencionado anteriormente, a coleção abriga três dicionários distintos. Caso não seja especificado o dicionário, a busca varrerá todos eles, mas se o usuário desejar, pode selecionar apenas um: [dhbb] para o Dicionário histórico-biográfico brasileiro; [dhbpr] para o Dicionário histórico-biográfico da Primeira República ; [dprj] para o Dicionário da política republicana do Rio de Janeiro.
- **classe** -> identifica se o verbete é [temático] ou [biográfico].
- **fonte** -> indica o verbete em si que é exatamente o nome completo da entrada, como por exemplo: [José Ribamar Ferreira de Araújo Costa] ou [REVOLUÇÃO DE 1930]. Por agora os verbetes temáticos têm o nome em maiúsculas.
- **entidade** -> identifica uma personalidade única, atribuída (por enquanto) apenas a nomes próprios humanos (todas as outras palavras têm id 0). Se essa personalidade estiver descrita no DHBB, o valor da entidade é o número que corresponde ao id do verbete na base de origem [dhbb_10, dhbb_5458]. Cada verbete tem uma identificação única, e a lista encontra-se em:
<https://www.linguateca.pt/aceso/entidadesDHBB.html>.
Se essa personalidade não estiver descrita no DHBB ou ainda não tiver sido identificada, o valor de id será "NS" (não sabemos) ou "NV" (não verbetado).
 - O processo de identificação dos nomes humanos que aparecem no DHBB é um processo em andamento, e a sua identificação é por vezes chamada "grounding", que explicaremos mais à frente.
- **sexo** -> [m] para biografias de homens, [f] para biografias de mulheres.
- **cargos** -> identifica todos os principais cargos ou papéis associados ao biografado. Este atributo concatena em uma única string todos os valores encontrados no campo "cargos" oriundos do cabeçalho de metadados do verbete em questão, respeitando a sequência original. Assim, teremos para o verbete de Pedro Aleixo a seguinte string:
[const1934_depFedMG1935-1937_depFedMG1959-1966_minEduc1966_depFedMG1966-1967_vice-presRep1967-1969]
- **autores** -> caso o verbete possua autoria, a informação é fornecida por este atributo.

Os atributos descritos acima foram adicionados a todos os tokens de um verbete biográfico. Isso significa que cada palavra/unidade terá agregado a si uma 'coluna' preenchida com o valor para dicionário, outra para fonte, outra para entidade, outra para cargos e assim por diante – o que nos permite, por exemplo, procurar quem "nasceu em Cataguazes" (utilizando uma busca léxico-sintática) e foi "presidente da República" (buscando pelo valor correspondente na coluna "cargos").

9. Qual a diferença entre entidade e fonte, não se referem ambos ao biografado?

É muito diferente: **fonte** e verbete são a mesma coisa, ou seja, trata-se do titular do verbete, seja ele biográfico ou temático. Sempre vai indicar de onde veio um texto ou pedaço de texto, pois como vimos, cada token tem associado a si uma fonte. Já as **entidades** só podem se referir a nomes de pessoas: quando estas coincidem com os próprios verbetados, então o valor da entidade será o próprio ID do verbete, caso contrário, será NS ou NV. É assim que conseguimos identificar as pessoas verbetadas quando elas aparecem nos textos. Ainda que os verbetes temáticos tenham um ID próprio, eles NÃO são entidades (ao menos por enquanto; no futuro pode ser que outras entidades além dos nomes de pessoas venham a ser identificadas também).

Para ilustrar, imagine que se deseje saber quais as fontes/verbetes que mencionam a palavra “ditadura”: basta buscar por [lema="ditadura"] e pedir a distribuição de fonte. A mesma expressão, se se pedir distribuição de entidade, devolverá Zero, porque "ditadura" não é uma entidade do DHBB.

Já se buscar por [entidade="dhbb_5458"], id que corresponde a Getúlio Dornelles Vargas no DHBB, e pedir distribuição de fonte, se verá em quais verbetes Getúlio Dornelles Vargas é mencionado (e pode sê-lo de diversas formas, como resultado do processo de *grounding* aplicado a esta personalidade).

10. O que é o processo de **identificação (grounding)**, para que serve e como está sendo feito?

Vamos chamar “identificação” ao que em inglês em linguística se chama *grounding* https://en.wikipedia.org/wiki/Symbol_grounding_problem e em linguística computacional se chama *entity linking*, *entity normalization* ou *entity disambiguation*, https://en.wikipedia.org/wiki/Entity_linking. No caso do DHBB, interessa-nos a criação de correspondências entre os nomes mencionados em certos contextos e os objetos do mundo real aos quais esses nomes se referem. No trabalho de processamento automático de corpus isso se torna necessário porque, se o texto traz entidades referentes ao mesmo objeto expressas de várias maneiras, a máquina precisa identificar todas estas expressões como sendo exatamente aquele a quem o humano está se referindo. No DHBB encontramos muitos casos de pessoas sendo mencionadas pelos seus nomes abreviados, por apelidos e até mesmo com erros de grafia. Há ainda casos em que homônimos só são resolvidos quando associados aos seus nomes completos de família. Lugares também são passíveis de terem várias formas de escrita.

Por regra, o nome completo da pessoa biografada virá sempre na primeira linha do texto do verbe, delimitado pelos caracteres «».

«Álvaro Augusto Ribeiro da Costa» nasceu em Fortaleza em 17 de abril de 1947.

Este será o valor do atributo “fonte” e também o referente a que corresponderão todas as possíveis derivações do nome da pessoa em questão.

Para realizar a tarefa de *grounding*, as seguintes etapas foram realizadas:

1. Recuperamos todos os nomes próprios a que o PALAVRAS atribuiu como sendo “humano”, e que não foram identificados como um titular de verbe; para isso usamos a expressão de busca [dicionario="dhbb" & sema="hum.*" & entidade!="*[1-9][0-9]*"].⁴
2. Da lista resultante, fazemos uma checagem manual indo aos trechos onde estes lemas são mencionados a fim de confirmarmos a identificação e evitarmos correspondências erradas. Alguns nomes são óbvios, como por ex. “Lula” (que sempre irá se referir a “Luís Inácio da Silva”) e “Getúlio=Vargas” (sempre será “Getúlio Dornelles Vargas”). Mas alguns, como “Vargas” e “Castelo=Branco”, podem se referir a mais de uma pessoa.
3. Após a checagem, criamos uma tabela indicando as correspondências entre os lemas e os IDs dos verbetes. Em geral, quando a ocorrência se verifica em um verbe próprio (por ex., “Vargas” sendo mencionado no verbe de “José Israel ‘Vargas’”), se está a referir ao seu titular (“José Israel Vargas”). Mas não podemos tomar esta regra como certa pois existem casos onde os lemas podem se referir

⁴ Anterior à criação do sema=hum, utilizávamos a sintaxe [pos="PROP.*hum" & entidade!="*[1-9][0-9]*"]

a outras pessoas, em geral parentes de mesmo sobrenome. No exemplo abaixo, os dois primeiros seguem a regra (o lema se refere ao próprio titular do verbete), mas os seguintes não:

AC/DC lema	Verbetes onde ocorre	Verbetes (id) a que corresponde
Vargas	Getúlio Dornelles Vargas	Getúlio Dornelles Vargas
Vargas	Jose Israel Vargas	Jose Israel Vargas
Vargas	Alzira Vargas do Amaral Peixoto	Getúlio Dornelles Vargas
Vargas	Benjamim Dornelles Vargas	Getúlio Dornelles Vargas
Vargas	Lutero Sarmanho Vargas	Getúlio Dornelles Vargas

Este é um processo iterativo: obtém-se a lista de lemas, fazem-se as correspondências; obtém-se nova lista, novas correspondências e assim sucessivamente. Sempre que se cria o corpus de novo, o processo de construção passa por um filtro que adiciona os identificadores, com base na lista de correspondências.

11. Como são tratadas as segmentações erradas?

Invariavelmente detectamos que alguns nomes próprios foram erroneamente segmentados pelo *parser*. Por exemplo, “Eugênia=Lopes=de=Oliveira=Prestes=de=Macedo=Soares” deveria ser um único nome, mas foi considerado como dois nomes separados: “Eugênia=Lopes=de=Oliveira=Prestes” e “Macedo=Soares”. Outro exemplo: “República=Jânio=Quadros” deveria ser apenas “Jânio=Quadros” (ex-presidente da República), mas teve “República” agregado ao seu nome.

Estes erros são corrigidos através de uma ferramenta baseada em regras para revisão de anotação em corpus, o “Corte-e-costura”⁵. O arquivo de regras para corrigir as segmentações erradas é este: https://www.linguateca.pt/acesso/corpos/dhbb/regras_corr_PALAVRAS_roupa_excl.txt.

12. Por que ao realizar busca por parte de um lema (usando o caractere coringa “*”), recuperamos linhas de concordância repetidas onde em uma linha a sentença mostra em negrito o termo da busca e em outra(s) linha(s), outra(s) parte(s) do mesmo nome?

Por exemplo, na busca por [lema="*.Bolsonaro"], temos como resultado:

[722](#) : «**Jair** Messias Bolsonaro» nasceu em Campinas (SP) no dia 21 de março de 1955, filho de Perci Geraldo Bolsonaro e de Olinda Bonturi Bolsonaro .

[722](#) : «Jair **Messias** Bolsonaro» nasceu em Campinas (SP) no dia 21 de março de 1955, filho de Perci Geraldo Bolsonaro e de Olinda Bonturi Bolsonaro .

[722](#) : «Jair Messias **Bolsonaro**» nasceu em Campinas (SP) no dia 21 de março de 1955, filho de Perci Geraldo Bolsonaro e de Olinda Bonturi Bolsonaro .

O que aparece em negrito é a forma que satisfaz a procura, ou seja, é onde há o encontro com o termo buscado. No exemplo acima, Bolsonaro faz parte de um lema constituído por três termos, e por isso a

⁵ Para saber mais, ler: Santos, D., Mota, C.: Experiments in human-computer cooperation for the semantic annotation of Portuguese corpora. In: Calzolari et al. (eds), Proceedings of LREC 2010. European Language Resources Association (2010).

sentença será recuperada três vezes. Para evitar essa repetição, pode-se usar um truque: pedir o lema do nome próprio usando a sintaxe *look ahead*: `[lema="*.Bolsonaro"]+ [: pos != "PROP.*" :]`

[722](#) : «**Jair Messias Bolsonaro**» nasceu em Campinas (SP) no dia 21 de março de 1955, filho de Perci Geraldo Bolsonaro e de Olinda Bonturi Bolsonaro .

13. Por que a distribuição dos cargos retorna strings aparentemente sem sentido?

`const1934_depfedMG1935-1937_depfedMG1959-1966_minEduc1966_depfedMG1966-1967_vice-presRep1967-1969`

Conforme mencionado no item relacionado aos atributos dos verbetes, as informações sobre cargos foram extraídas dos cabeçalhos de metadados, concatenando-se todos os valores na sequência em que foram encontrados. Assim, temos o cargo (abreviado na origem) seguido do Estado de atuação (quando houver) e o período de mandato (quando houver).

No exemplo acima, a pessoa de Pedro Aleixo foi deputado constituinte de 1934, deputado federal em três períodos diferentes, foi ministro da Educação durante alguns meses do ano de 1966 e exerceu o cargo de vice-presidente entre 1967 e 1969.

Veja mais adiante uma tabela com as abreviaturas dos principais cargos, para usar nas buscas.

14. Sobre as categorias semânticas usadas no DHBB e a criação de léxicos

O PALAVRAS possui marcações semânticas para algumas classes de nomes. A lista completa destas marcas pode ser consultada nesta página: <https://visl.sdu.dk/visl/pt/info/portsymbol.html>. Por ora, o que vai nos interessar de fato é apenas uma seleção pequena:

- `<hum>` = pessoa
- `<org>` e `<inst>` = organização/instituição
- `<party>` = partido político
- `<occ>` e `<event>` = evento/acometimento
- `<civ>` = lugar
- `<tit>` = documento/obra
- `<hprof>` = cargo/papel

Além destas, criamos também uma nova:

- `<titfpol>` = formulações políticas (tais como planos, programas, acordos, tratados, leis, decretos, códigos etc.)

Como era de se esperar, o analisador não consegue atribuir de forma 100% correta a classe dos nomes próprios presentes em um corpus. Para melhorar a identificação, usamos listas com entidades revisadas manualmente. Como se deu este processo? Da seguinte forma: por meio da interface do AC/DC realizamos consultas linguísticas no DHBB capazes de indicar a presença de nomes do tipo organização `<org>`, e assim obtemos uma lista inicial de quase 28 mil instâncias. Manualmente revisamos cada entrada para validar a atribuição semântica dada pelo PALAVRAS e no final restaram cerca de 22.500 organizações, que agora podemos considerar como sendo uma lista “golden”. Os nomes identificados erroneamente popularam as outras listas: pessoa, formulação política, evento, lugar, documento, partido e papel. Dessa forma, forçamos uma identificação mais correta desse conjunto de nomes próprios presentes no DHBB.

15. Onde se encontra a informação semântica?

Consideramos oportuno usar o atributo <sema> para identificar as classes semânticas que nos interessam. Dessa forma, as regras de anotação seguem a mesma lógica daquelas usadas para os outros léxicos criados circunstancialmente, como o das relações de parentesco⁶ (veja mais adiante), das emoções⁷, do corpo humano⁸, etc., podendo ser incorporadas nas buscas realizadas no corpus. Assim, as seguintes classes se tornaram um atributo <sema> no AC/DC:

- organizações -> [sema="org"]
- formulações políticas -> [sema="titfpol"]
- documentos -> [sema="tit"]
- pessoas -> [sema="hum"]
- lugares -> [sema="civ "]
- eventos -> [sema="evento "]
- partidos -> [sema="party"]
- papéis -> [sema="Hprof"]

16. Estão contando colocar mais informação?

A construção do DHBB segue uma filosofia de enriquecimento cíclico, em que à medida que vamos obtendo mais informação a adicionamos ao corpus, de forma que as próximas versões tenham sempre mais informação ou informação mais correta. Por exemplo, além das relações de parentesco, estamos a estudar vínculos institucionais, formação académica, datas de nascimento e morte. É bem provável que essas relações fiquem mais explícitas em futuras versões, e que as datas sejam juntadas aos metadados.

17. Como é feito o enriquecimento do corpus com estas informações adicionais?

No caso das relações de parentesco, em primeiro lugar definimos um léxico com palavras pertencentes a este domínio a fim de que sejam usadas na anotação dos verbetes (este léxico pode ser conhecido aqui: <https://www.linguateca.pt/Gramateca/Familia.html>). Em alguns casos é preciso especificar e/ou desambiguar os contextos que indicam o uso do termo naquele domínio ou não (como no caso de “*mulher de*” ou “*separado de*”, termos bastante usados fora do contexto de família). Essas informações encontram-se em um arquivo de regras em formato VISLCG3⁹. Para casos mais específicos e mais difíceis de generalizar, colocamos outras regras no formato corte-e-costura.

Exemplos de como podemos fazer uso deste enriquecimento:

1. Quais relações de parentesco são mais frequentes no DHBB?

```
[sema=".*familia:lacos.*" & dicionario="dhbb"]
```

Pedir distribuição por lema:

filho	10876	companheiro	482	sobrinho	221
pai	1779	tio	369	parente	207
família	1703	primo	335	casal	156
irmão	1624	esposa	298	irmã	155
filha	1334	mãe	274	marido	149
...					

⁶ <https://www.linguateca.pt/Gramateca/Familia.html>

⁷ <https://www.linguateca.pt/Gramateca/Emocionario.html>

⁸ <https://www.linguateca.pt/acesso/Esqueleto/>

⁹ https://www.linguateca.pt/acesso/corpos/visl3g3/regas_emodizer.utf8.txt

2. Como se dão as relações especificadas?

`<mwe> [sema=".*familia:lacos.*"] []+ </mwe>`

Pedir concordância:

[dhbb1086](#) : Seu **tio paterno**, Gileno De Carli, foi deputado federal por Pernambuco de 1959 a 1963, e seu primo João Carlos de Petribu De Carli também foi deputado federal por esse estado de 1979 a 1987 .

[dhbb10901](#) : Da **família materna**, herdou o título de marquês de Salamanca .

[dhbb110](#) : Além de um **irmão gêmeo**, teve ainda outros 12 irmãos .

[dhbb11043](#) : Entre os que foram aprovados, figuram a lei que concede licença maternidade para a **mãe adotiva** e a do primeiro emprego .

[dhbb11076](#) : Em 1951, a fim de cursar o ginasial, Gil se mudou, juntamente com sua irmã, para Salvador, residindo na casa de uma **tia paterna** .

....

3. Quais as relações de parentesco atribuídas a Getúlio Vargas (entidade 5458) usando o padrão [relação de parentesco] de [Getúlio Vargas]?

`[dicionario="dhbb"] @[sema=".*familia:lacos.*"] "de" [entidade=".*5458"]+[: entidade!=".*5458" :]`

Pedir distribuição por lema e concordância:

[dhbb1503](#) : Aproveitando o descontentamento causado pela nomeação de Benjamim Vargas , **irmão de Getúlio Vargas**, para a Chefatura de Polícia do Distrito Federal, Dutra e Góis Monteiro articularam o golpe de Estado que em 29 de outubro de 1945 derrubou o Estado Novo .

[dhbb1503](#) : Em junho, depois que os trabalhistas já se haviam praticamente definido a favor do ex-presidente, Canrobert pediu ao deputado federal fluminense Ernâni Amaral Peixoto , **genro de Vargas**, que o aconselhasse a retirar sua candidatura para não desagradar ao Exército .

[dhbb1754](#) : Seu pai era irmão de Cândida Dornelles Vargas , **mãe de Getúlio Vargas** .

[dhbb1755](#) : Era **primo de Getúlio Dornelles Vargas** .

[dhbb1756](#) : Seu avô paterno, Ernesto Francisco Dornelles, militar e natural de São Borja (RS) , era irmão de Cândida Dornelles Vargas , **mãe de Getúlio Vargas** .

...

Sobre como usar o AC/DC (com o DHBB)

18. O que é possível perguntar ao DHBB no AC/DC?

A redação dos verbetes segue algumas orientações mínimas sugeridas pela equipe do DHBB no que diz respeito à construção das sentenças, conteúdo textual e disposição da narrativa, e assim, utilizando padrões léxico sintáticos simples podemos tentar tirar algum proveito. Obviamente, não é possível garantir 100% da recuperação das informações, pois o DHBB possui muitos autores que desenvolvem sua própria forma de escrita. Ou ainda, não encontrar a informação que estamos a pesquisar simplesmente porque ela não existe em alguns verbetes. A seguir, alguns exemplos simples.

- Quais os nomes próprios do DHBB que contêm “Vargas”?

`[dicionario="dhbb" & word="Vargas" & sema=".*hum.*"]`

Pedir distribuição por lema

Vargas 4892

Getúlio=Vargas 2370

Ivete=Vargas 129

Benjamim=Vargas 69

Lutero=Vargas 38

André=Vargas 33

Alzira=Vargas=do=Amaral=Peixoto 18

...

- Em quais verbetes biográficos o ex-presidente Eurico Gaspar Dutra (entidade 1790) é mencionado?

[classe="biográfico" & entidade="dhbb_1790" & sema="hum.*"]+ [: sema!="hum.*" :]

Pedir concordância para visualizar as ocorrências

Pedir distribuição por lema para visualizar as variantes do nome, pós-identificação

Pedir distribuição por fonte para saber em quais verbetes ocorre

[dhbb100](#) : Realizado o pleito na data marcada, foi eleito para a presidência da República o candidato do PSD, general **Eurico Gaspar Dutra**, que contava com o apoio do PTB .

[dhbb1005](#) : Entretanto, a articulação golpista prosperava, capitaneada pelos dois homens fortes do Exército, os generais Góis Monteiro (chefe do Estado-Maior do Exército) e **Eurico Gaspar Dutra** (ministro da Guerra) .

[dhbb1005](#) : No dia 8, o texto da Constituição a ser outorgada foi examinado na casa de Francisco Campos por Góis Monteiro e pelos ministros da Guerra, general **Eurico Dutra**, e da Marinha, almirante Aristides Guilhem .

[dhbb1005](#) : O assalto fracassou diante da resistência, mesmo precária, dos ocupantes do palácio e da ação improvisada do general **Dutra**, ministro da Guerra .

[dhbb1005](#) : **Dutra** aprovou plenamente suas recomendações .

[dhbb1018](#) : Vargas foi derrubado em 29 de outubro de 1945 por um golpe militar chefiado pelos generais **Eurico Dutra** e Pedro Aurélio de Góis Monteiro .

...

- Quais as menções de parentesco encontradas no verbete de Fernando Collor de Melo?

[dicionario="dhbb" & fonte="Fernando Afonso Collor de Melo"] @[sema="*.familia:lacos.*"]

Pedir distribuição por concordância e lema

[dhbb1418](#) : Em agosto de 1995, Collor transferiu-se para Miami, nos Estados Unidos, onde passou a viver, segundo afirmou, do pró-labore de cem mil dólares mensais que receberia das empresas de **sua família** .

[dhbb1418](#) : Os cartazes coincidiram com uma campanha que as emissoras de rádio **da família** fizeram para a sua volta à vida política .

[dhbb1418](#) : De seu primeiro casamento, teve **dois filhos** .

[dhbb1418](#) : Teve **outro filho**, Fernando James, eleito vereador em Rio Largo (AL) .

[dhbb1418](#) : De seu terceiro casamento, com Caroline Medeiros, teve **duas filhas** .

...

- Quantas e quais relações de parentesco entre personagens políticas podem ser identificadas em todo o corpus, com a construção [biografado], [relação de parentesco] de [biografado]

[dicionario="dhbb"] [entidade="d.*"]+"", "[sema="*.familia:lacos.*"] "de" [entidade="d.*"]+ [: entidade!="d.*" :]

Pedir concordância

Pedir distribuição por lema

[dhbb1570](#) : Foi nomeado secretário de Planejamento e Coordenação Geral de Minas Gerais por indicação de **Francisco Dornelles, sobrinho de Tancredo Neves**, então governador do estado (1983-1984) .

[dhbb25](#) : Tal movimento teria sido sustado graças a uma enérgica intervenção do ministro do Exército, o **general Orlando Geisel, irmão de Ernesto Geisel** .

[dhbb2538](#) : Esse estado era governado por **Leonel Brizola, cunhado de Goulart** .

[dhbb2645](#) : Na eleição para governador do estado do Rio de Janeiro , **Ernâni Amaral Peixoto, genro de Vargas** e candidato de ampla aliança entre o PSD, o PTB, o PR, o Partido Social Trabalhista (PST) , o Partido Trabalhista Nacional (PTN) e o Partido Rural Trabalhista (PRT) , recebeu cerca de 280 mil votos, contra pouco mais de 120 mil dados a Prado Kelly .

[dhbb2995](#) : O movimento era mais forte no Rio Grande do Sul, sob a liderança do governador do estado , **Leonel Brizola, cunhado de Goulart** .

...

- Quantos e quais são os verbetes de ministros do STF?

[dicionario="dhbb" & cargos=".*minSTF.*"]

Pedir distribuição por fonte

- Quantos e quais são os verbetes de presidentes eleitos que também foram militares?

[dicionario="dhbb" & cargos=".*_presRep.*" & cargos=".*militar.*"]

Pedir distribuição por fonte

- Quais são os verbetes dos que concorreram à eleição presidencial de 1994 e 1998?

[dicionario="dhbb" & cargos=".*cand.?resRep199.*"]

Pedir distribuição por fonte

- Quantos e quais são os verbetes de deputadas federais pelo Rio de Janeiro?

[dicionario="dhbb" & cargos=".*depfedRJ.*" & sexo="f"]

Pedir distribuição por fonte

- Quando e onde nasceram os ministros do Supremo Tribunal Federal?

[cargos=".*minSTF.*" & dicionario="dhbb"] [pos="PROP.*"]+ [: pos!="PROP.*" :] []{0,1} [lema="nascer"]
[pos="PRP.*"] []{0,21} [pos="NUM.*|ADJ.*"] [word="de"]? [pos="N.*"]? [word="de"]? [pos="NUM.*"]?
[pos="PU"]

Pedir concordância

- Quais locais de nascimento dos biografados que exerceram cargo de deputado federal pelo Rio de Janeiro?

[dicionario="dhbb" & cargos=".*depfedRJ.*"] [pos="PROP.*"]+ [: pos!="PROP.*" :] []{0,1} [lema="nascer"]
[lema="em(\.)*"] ([lema="estado|cidade|município"] [lema="de(\.)*"])* @[pos="PROP.*"]
[pos="PROP.*"]* [: pos!="PROP.*":]

Pedir fonte para saber quantos deputados são

Pedir distribuição por lema para saber os lugares distintos

- Quais verbetes foram redigidos por determinado autor?

[dicionario="dhbb" & autor="Paulo Brandi"]

Pedir distribuição por fonte

19. ABREVIATURAS DOS CARGOS PRESENTES NAS BIOGRAFIAS

Constituintes, deputados e senadores

const.
dep. fed. AC
dep. fed. AL
dep. fed. AM
dep. fed. AP
dep. fed. BA
dep. fed. CE
dep. fed. DF
dep. fed. ES
dep. fed. GB
dep. fed. GO
dep. fed. GR
dep. fed. MA
dep. fed. MG
dep. fed. MS
dep. fed. MT
dep. fed. PA
dep. fed. PB
dep. fed. PE
dep. fed. PI
dep. fed. PR
dep. fed. prof.
dep. fed. RB
dep. fed. RJ
dep. fed. RN
dep. fed. RO
dep. fed. RR
dep. fed. RS
dep. fed. SC
dep. fed. SE
dep. fed. SP
dep. fed. TO
sen. AC
sen. AL
sen. AM
sen. AP
sen. BA
sen. CE
sen. DF
sen. ES
sen. GB
sen. GO
sen. MA
sen. MG
sen. MS
sen. MT
sen. PA
sen. PB
sen. PE

sen. PI
sen. PR
sen. RJ
sen. RN
sen. RO
sen. RR
sen. RS
sen. SC
sen. SE
sen. SP
sen. TO

Presidentes, vice-presidentes e ministros

cand. pres. Rep.
cand. vice-pres. Rep.
pres. Rep.
pres. Rep. eleito
vice-pres. Rep.
prim.-min.
min. Ação Social
min. Adm.
min. Adm. Ref. Est.
min. Aer.
min. Agric.
min. Agric. Abast. e Ref. Agr.
min. Amazônia Legal e Meio Ambiente
min. Assist. e Prom. Social
min. Assist. Social
min. Assunt. Estrat.
min. Bem-Estar Soc.
min. Casa Civ. Pres. Rep.
min. Casa Mil. Pres. Rep.
min. CGU
min. ch. EMFA
min. ch. SAF
min. ch. Secretaria de Governo
min. ch. Seplan
min. ch. SNI
min. Cidades
min. Ciên. Tecn.
min. Ciên. Tecn. e Inov.
min. Comunic.
min. Comunic. Pres. Rep.
min. Controle e Transparência
min. Cultura
min. Defesa
min. Desenv. Agr.
min. Desenv. Ind. Com.
min. Desenv. Regional

min. Desenv. Soc.
 min. Desenv. Urb. e Meio Ambiente
 min. Direitos Humanos
 min. Econ.
 min. Educ.
 min. Educ. e Cultura
 min. Educ. e Desp.
 min. Esporte
 min. Esporte e Turis.
 min. Exérc.
 min. Extr. Ass. Fundiários
 min. Extr. Ass. Irrigação
 min. Extr. Ass. Parl.
 min. Extr. Coord. Órgãos Reg.
 min. Extr. Criança
 min. Extr. Desburoc.
 min. Extr. Integr. América Latina
 min. Extr. Pol. Fund. e Desenv. Agr.
 min. Extr. Ref. Admin.
 min. Extr. Ref. Inst.
 min. Faz.
 min. Gab. Seg. Inst.
 min. Guerra
 min. Hab. Urb. e Meio Amb.
 min. Ind. Com. e Tur.
 min. Ind. e Com.
 min. Infraestrutura
 min. Integr. Nac.
 min. Integr. Reg.
 min. Interior
 min. Interior e Just.
 min. Just.
 min. Mar.
 min. Meio Amb. e Amazônia Legal
 min. Meio Ambiente
 min. Meio Ambiente, Recursos Hídricos e da
 Amazônia Legal
 min. Minas e En.
 min. Orç.
 min. Pesca
 min. Planej.
 min. Planejamento, Orçamento e Gestão
 min. Pol. Fund. e Desenv. Agr.
 min. Prev. e Assist. Social
 min. Prev. Social
 min. Ref. e Desenv. Agr.
 min. Rel. Ext.
 min. Rel. Inst.
 min. SAE
 min. Saúde
 min. Sec. Ass. Estrat.
 min. Sec. Comunic.
 min. Sec. Des. Urbano

min. Sec. Esp. Pol. Mulheres
 min. Sec. Esp. Pol. Prom. Iguald. Racial
 min. Sec. Esp. Portos
 min. Sec. Especial de Aquicultura e Pesca
 min. Sec. Ger. Pres. Rep.
 min. Sec. Micro e Pequena Empresa
 min. Sec. Pol. Region.
 min. STF
 min. STJ
 min. STM
 min. TCU
 min. Trab.
 min. Trab. Admin.
 min. Trab. e Emprego
 min. Trab. e Prev. Soc.
 min. Transp.
 min. Transp. e Com.
 min. TSE
 min. Tur.
 min. Tur. e Esp.
 min. Viação
 min. Viação e Agric.

Membros de juntas

pres. junta Gov. Prov.
 junta gov. AM
 junta gov. GO
 junta gov. MA
 junta gov. PA
 junta gov. prov.
 junta gov. RN
 junta gov. RS
 junta gov. SC
 junta mil.

Militares, comandantes e chefes

militar
 insp. 1º Gr. RM
 insp. 2º Gr. RM
 comdo. naval Brasília
 comte. 12ª RM
 comte. 1ª RM
 comte. 2ª RM
 comte. 3ª RM
 comte. 4ª RM
 comte. 5ª RM
 comte. 7ª RM
 comte. Art. FEB
 comte. CFN
 comte. Comdo. Mil. Amazônia
 comte. Comdo. Mil. Brasília
 comte. Comdo. Mil. Planalto
 comte. Depto. Eng. Comunic. Ex.

comte. EMA
comte. ESG
comte. FEB
comte. Força Naval NE
comte. ger. PMDF
comte. ger. PMRJ
comte. I DN
comte. I Ex.
comte. I ZA
comte. II Comar
comte. II DN
comte. II Ex.
comte. II Exército
comte. II ZA
comte. III Comar
comte. III Dist. Naval
comte. III DN
comte. III Ex.
comte. III ZA
comte. Inf. FEB
comte. IV Comar
comte. IV DN
comte. IV Ex.
comte. IV ZA
comte. V DN
comte. V ZA
comte. VI Comar
comte. VI DN
comte. VI ZA
comte. Zona Mil. Centro
comte. Zona Mil. Leste
comte. Zona Mil. Norte
comte. Zona Mil. Sul
comte.-em-ch. Esquadra

Chefes

ch. VI Comar
ch. EM FEB
ch. EM Geral
ch. EM Gov. Prov.
ch. EM Pres. Rep.
ch. Depto. Eng. Comunic. Ex.
ch. Depto. Ens. Pesq. Ex.
ch. Depto. Ger. Admin. Ex.
ch. Depto. Ger. Pess. Ex.
ch. Depto. Ger. Serv. Ex.
ch. Depto. Mat. Bél. Ex.
ch. Depto. Prod. Obras Ex.
ch. Depto. Prov. Ger. Ex.
ch. Depto. Téc. Prod. Ex.
ch. AERP
ch. CGI
ch. CODI

ch. EMA
ch. Emaer
ch. EME
ch. EMFA
ch. DFSP
ch. SNI
chefe DPF
ch. pol. DF

ch. Gab. Mil. Pres. Rep.
ch. Gab. Civ. Pres. Rep.

Assessores e auxiliares

ass. econ. Pres. Rep.
ass. Pres. Rep.
aux. gab. Pres. Rep.
consult. ger. Rep.

Delegados

deleg. mil. SP
deleg. pol. SP

Advogados, juízes, magistrados, procuradores

adv. geral União
juiz TSN
jurista
magistrada
magistrado
proc. ger. Rep.
promotor TSN

Reitores

reitor UB
reitor UDF
reitor UFRJ
reitor UnB
reitor Univ. RJ
reitor USP

Interventores e governadores

ch. gov. prov. SP
interv. AC
interv. AL
interv. AM
interv. BA
interv. CE
interv. DF
interv. ES
interv. GO
interv. MA
interv. MG
interv. MT

interv. PA
 interv. PB
 interv. PE
 interv. PI
 interv. PR
 interv. RJ
 interv. RN
 interv. RS
 interv. SC
 interv. SE
 interv. SP
 gov. AC
 gov. AL
 gov. AM
 gov. AP
 gov. BA
 gov. CE
 gov. DF
 gov. ES
 gov. GB
 gov. GO
 gov. MA
 gov. MG
 gov. mil. AM
 gov. mil. SP
 gov. MS
 gov. MT
 gov. PA
 gov. PB
 gov. PE
 gov. PI
 gov. PR
 gov. prov. CE
 gov. prov. PE
 gov. prov. SE
 gov. prov. SP
 gov. RB
 gov. RJ
 gov. RN
 gov. RO
 gov. RR
 gov. RS
 gov. SC
 gov. SE
 gov. SP
 gov. TO
 vice-gov. AL
 vice-gov. RN
 vice-gov. SP
 vice-gov. TO

Diplomatas, embaixadores, encarregados, representantes

diplomata
 diplomata alemão
 diplomata argentino
 diplomata francês
 diplomata inglês
 diplomata italiano
 diplomata japonês
 diplomata norte-americano
 diplomata suíço
 emb. Alemanha no Brasil
 emb. Alemanha Ocid. no Brasil
 emb. Argentina no Brasil
 emb. Bras. Alemanha
 emb. Bras. Alemanha Ocid.
 emb. Bras. Argentina
 emb. Bras. Bolívia
 emb. Bras. Chile
 emb. Bras. Estados Unidos
 emb. Bras. EUA
 emb. Bras. França
 emb. Bras. Inglaterra
 emb. Bras. Itália
 emb. Bras. Liga das Nações
 emb. Bras. OEA
 emb. Bras. OMC
 emb. Bras. ONU
 emb. Bras. Paraguai
 emb. Bras. Peru
 emb. Bras. Portugal
 emb. Bras. Rússia
 emb. Bras. URSS
 emb. Bras. Uruguai
 emb. Bras. Vaticano
 emb. EUA no Brasil
 emb. França no Brasil
 emb. Inglaterra no Brasil
 emb. Itália no Brasil
 emb. Suíça no Brasil
 encar. neg. Bras. EUA
 encar. neg. EUA
 repr. França no Brasil
 adido mil. EUA no Brasil
 ch. missão econ. Inglaterra no Brasil
 ch. missão econ. norte-americana no Brasil

Secretários

sec. Adm. Fed.
 sec. Ass. Estratég.
 sec. Ciên. Tecn.
 sec. Cons. Téc. Econ. Fin.
 sec. Coord. Pol.
 sec. Cultura
 sec. esp. Agricultura e Pesca

sec. esp. Cons. Desenvol. Econ. Soc.
sec. esp. Pol. Prom. Igualdade Racial
sec. ger. ANL
sec. ger. CNBB
sec. Impr. e Divulg. Pres. Rep.
sec. Nac. Comun.
sec. Nac. Desportos
sec. Nac. Dir. Hum.
sec. Nac. Meio Ambiente
sec. Nac. Pol. Mulheres
sec. Pres. Rep.

**Presidentes e diretores de agências, instituições
públicas e privadas**

superint. Sudene
superint. Sumoc
pres. Acad. Bras. Letras
pres. ACRJ
pres. ANL
pres. BA
pres. Bco. Bras.
pres. Bco. Central
pres. BNDE
pres. BNDES
pres. BNH
pres. CAT
pres. CGT
pres. CGTB
pres. CIB
pres. CIESP
pres. CIRJ
pres. CIRJ/FIDF
pres. CIRJ/Fiega
pres. CIRJ/FIRJ
pres. CNA
pres. CNBB
pres. CNC
pres. CNI
pres. CNP
pres. CNPq
pres. CNTC
pres. CNTI
pres. CNTTMFA
pres. CNTTT
pres. Cons. Nac. Café
pres. Contag
pres. Contec
pres. CUT
pres. CVRD
pres. DASP
pres. DNC
pres. Eletrobrás
pres. FGV

pres. FIEGA/CIRJ
pres. FIESP
pres. FIESP/CIESP
pres. FIRJ
pres. Firjan
pres. Firjan-CIRJ
pres. Força Sindical
pres. IAPC
pres. IBC
pres. INPS
pres. IPASE
pres. Light
pres. MA
pres. MG
pres. MT
pres. NCST
pres. PB
pres. Petrobras
pres. PTB
pres. RJ
pres. RS
pres. SC
pres. SDS
pres. SE
pres. SNA
pres. SP
pres. STF
pres. TCU
pres. Telebrás
pres. TSE
pres. UGT
pres. UNE
dir. Ag. Nac.
dir. Av. Mil.
dir. CGT
dir. DASP
dir. Depto. Nac. Prop.
dir. Depto. Prop. Dif. Cult.
dir. DIP
dir. ger. ABIN
dir. ger. DFSP
dir. ger. DPF
ch. oper. Petrobras

Prefeitos

pref. Aracaju
pref. Belém
pref. Belo Horizonte
pref. Boa Vista
pref. Cuiabá
pref. Curitiba
pref. DF
pref. Florianópolis

pref. Fortaleza
pref. Goiânia
pref. Imperatriz
pref. João Pessoa
pref. Macapá
pref. Maceió
pref. Manaus
pref. Natal
pref. Palmas
pref. Porto Alegre
pref. Porto Velho
pref. Recife
pref. Rio Branco
pref. Rio de Janeiro
pref. Salvador
pref. Santa Inês
pref. São Luís
pref. São Paulo
pref. Sapeaçu
pref. Teresina
pref. Vitória

**Líderes soc. civil e participantes de atentados,
revoluções, movimentos**

atent. Toneleros
col. Prestes
Contestado
líder estudantil
líder guerrilheiro
líder seringueiro
líder soc. civil
mov. comunista
mov. integralista
mov. tenentista
rev.
rev. 1892
rev. 1922
rev. 1923
rev. 1924
rev. 1925
rev. 1926
rev. 1930
rev. 1932
rev. 1935
rev. 1938
rev. 1930
rev. 1964
rev. Aragarças
rev. Jacareacanga
rev. marinheiros
rev. Princesa
rev. sargentos
sind.

membro ANL

Arcebispos e religiosos

arceb. Belo Horizonte
arceb. Brasília
arceb. Olinda e Recife
arceb. Porto Alegre
arceb. primaz de Salvador e do Brasil
arceb. Rio de Janeiro
arceb. Salvador
arceb. Salvador e primaz Brasil
arceb. São Paulo
religioso

Profissionais liberais

arquiteto
ecologista
empresário
escritor
intelectual
jornalista
pensador político
urbanista