

Documentação

Introdução

A aplicação usa um compilador baseado numa gramática [PEG.js](#) para processar o input do utilizador e gerar o dataset pretendido. A gramática mencionada define uma linguagem específica de domínio (DSL), com sintaxe semelhante a JSON, disponibilizando muitas ferramentas que permitem a geração de datasets complexos e diversificados. Estas ferramentas incluem capacidades relacionais e lógicas, fornecendo meios para os datasets satisfazerem vários tipos de limitações - o que facilita a utilização de frameworks declarativas com esta especificação -, bem como capacidades funcionais, permitindo uma gestão e processamento facilitados de certas propriedades dos datasets.

A primeira e mais fundamental das ferramentas implementadas é a sintaxe semelhante a JSON - o utilizador pode especificar propriedades chave-valor, onde o valor pode tomar qualquer tipo básico ou estrutura de dados JSON, desde inteiros a objetos e arrays. O utilizador pode também aninhar estes valores para criar uma estrutura com qualquer profundidade que pretenda.

```
1 nome: {
2   first: ["Hugo", "Cardoso"],
3   last: "Miguel"
4 },
5 age: 21
```

Definição de Pares Chave-Valor

Um par chave-valor é composto por dois elementos separadas por dois pontos (:).

A chave não pode conter espaços brancos (exceto entre o último caractere da *String* e o separador) nem qualquer outro caractere que não pertença ao alfabeto ou que não seja um *underscore*.

Por exemplo, `lorem_ipsum` é uma chave válida enquanto `lorem ipsum`, `lorem:ipsum` ou `lorem-ipsum` não são.

A única exceção é a diretiva `repeat`, que está entre plicas e que recebe como argumento um inteiro. Esta é responsável por gerar um *array* de objetos cujo comprimento é o dado por argumento.

Os valores podem ser um dos seguintes:

- Número
- *String*
- *Array*
- Boleano
- null
- Objeto DSL
- Função "Moustache"

Interpolações (Funções "Moustache")

Para definir o valor de uma propriedade, o utilizador pode também usar interpolação. Para aceder a uma função de interpolação, esta necessita de estar envolta em chavetas duplas. Há dois tipos de funções de interpolação:

- **Funções de geração espontânea:** geram valores espontâneos em tempo de execução, de acordo com as instruções do utilizador - por exemplo, existe uma função de geração de números inteiros aleatórios, onde o utilizador precisa de indicar, no mínimo, a gama de valores que pretende para o resultado:

```
1 id: '{{objectId()}}',
2 int: '{{integer(50,100)}}',
3 random: '{{random(23, "olá", [1,2,3], true)}}'
```

- **Funções com suporte de datasets:** retornam valores aleatórios de um grupo de datasets incorporados na aplicação por detrás de uma API, onde cada dataset possui informação de uma dada categoria, por exemplo nomes e partidos políticos:

```
1 nome: '{{fullName()}}',
2 partido: '{{political_party()}}'
```

Estas funções de interpolação podem também ser interligadas entre si e com *Strings* elementares para gerar *Strings* mais estruturadas, nomeadamente moradas. Algumas destas funções recebem argumentos e, nesse caso, o utilizador tanto pode introduzir manualmente os valores, como pode referenciar outras propriedades definidas acima no modelo, através da variável local `this`, permitindo assim estabelecer relações entre vários campos.

```
1 freguesia: '{{pt_parish()}}',
```

```
2 distrito: '{{pt_district("parish", this.parish)}}',
3 morada: 'Rua {{fullName()}}, {{pt_city("district", this.district)}}'
```

O utilizador pode ainda chamar a função `.string()` à frente de uma interpolação para converter o seu valor para *String*. Isto é redundante se o valor de retorno já for uma *string*, mas pode ser usado para converter resultados de funções que retornem outros tipos.

```
1 int: '{{integer(5,15)}}'.string(),
2 bool: '{{boolean()}}'.string()
```

De seguida, estão explícitas todas as funções de interpolação que se encontram atualmente disponíveis (os argumentos que têm um asterisco podem ser referenciados localmente através da chave `this`).

Funções de geração espontânea:

MOUSTACHE	TIPO	ARGUMENTO(S)	DESCRIÇÃO
boolean	Boolean	Nenhum	Gera um boleano aleatório. Exemplo: boolean() = true
date	String	*Init:: String	Gera uma data aleatória entre a data atual e a data Init. A <i>String</i> do Init tem de ter o formato "DD[./-]MM[./-]YYYY". O resultado é dado em versão JS raw. Exemplo: date("12-12-2100") = "2056-04-11T01:22:38.174Z"
date	String	*Init:: String, Form:: String	Gera uma data aleatória entre a data atual e a data Init, no formato dado. A <i>String</i> do argumento Init tem de ter o formato "DD[./-]MM[./-]YYYY". Os formatos possíveis para o argumento Form são os seguintes: DD.MM.YYYY, DD.MM.AAAA, MM.DD.YYYY, MM.DD.AAAA, YYYY.MM.DD e AAAA.MM.DD, onde o . pode ser / - ou . Exemplo: date("12-12-2100","DD.MM.AAAA") = "09.12.2083"
date	String	*Init:: String, *Fim:: String	Gera uma data aleatória entre as datas argumentos. A <i>String</i> dos argumentos tem de ter o formato "DD[./-]MM[./-]YYYY". O resultado é dado em versão JS raw. Exemplo: date("12-12-2100","20-12-2100") = "2100-12-16T12:11:08.049Z"
date	String	*Init:: String, *Fim:: String, Form:: String	Gera uma data aleatória entre as datas argumentos, no formato dado. A <i>String</i> dos argumentos Init e Fim tem de ter o formato "DD[./-]MM[./-]YYYY". Os formatos possíveis para o argumento Form são os seguintes: DD.MM.YYYY, DD.MM.AAAA, MM.DD.YYYY, MM.DD.AAAA, YYYY.MM.DD e AAAA.MM.DD, onde o . pode ser / - ou . Exemplo: date("12-12-2100","20-12-2100","DD.MM.AAAA") = "18.12.2100"
float	Float	*Min:: Float, *Max:: Float	Gera um número decimal aleatório entre Min e Max. O número de casas decimais do resultado vai ser igual ao do argumento com mais casas decimais. Exemplo: float(-180.4, 180.29) = -19.10
float	Float	*Min:: Float, *Max:: Float, *Dec:: Integer	Gera um número decimal aleatório entre Min e Max com um total de Dec casas decimais. Caso o número gerado acabe com 0s à direita na parte decimal, estes são omitidos. Exemplo: float(-180, 180, 2) = -19.11
formattedInteger	String	*Min:: Integer, *Max:: Integer, *Pad:: Integer, Unid:: String	Gera um inteiro aleatório entre Min e Max, garante que tem pelo menos tantos algarismos quantos especificados no Pad e acrescenta uma <i>String</i> Unid no final. Caso não queira padding, pode colocar um 0 no Pad e caso não queira colocar a unidade, pode colocar um "" no Unid. Exemplo1: formattedInteger(2, 400, 3, "\$") = "100\$" Exemplo2: formattedInteger(1, 3, 3, "") = "002"
formattedFloat	String	*Min:: Integer, *Max:: Integer,	Gera um número decimal aleatório entre Min e Max com um total de Dec casas decimais e garante que tem pelo menos tantos algarismos na parte inteira quantos especificados no

		<div><div>*Dec:: Integer,</div><div>*Pad:: Integer,</div><div>Form:: String</div></div>	<div>Pad.</div> <div>Por fim, recebe um formato no argumento Form na forma "0#0#00?", onde o primeiro # é um caractere para separar cada 3 algarismos de inteiros, o segundo # separa a parte inteira da decimal e ? é uma string a concatenar no fim (unidades).</div> <div>Caso não queira padding, pode colocar um 0 no Pad e caso não queira colocar uma unidade, pode acabar a <i>String</i> do formato logo a seguir aos últimos 00.</div> <div>Exemplo1: formattedFloat(2, 400, 3, 4, "0.0,00\$") = "0.181,306\$"</div> <div>Exemplo2: formattedFloat(2, 5, 2, 0, "0.0,00") = "2,05"</div>
guid	String	Nenhum	<div>Gera um UUID aleatório.</div> <div>Exemplo: guid() = "3d16d5d0-4b11-4de8-9e26-6668b52d9219"</div>
index	Integer	Nenhum	<div>Retorna o índice atual no array em que se encontra (normal ou gerado pelo 'repeat').</div> <div>Exemplo: index() = 1</div>
integer	Integer	<div>*Min:: Integer,</div> <div>*Max:: Integer</div>	<div>Gera um inteiro aleatório entre Min e Max.</div> <div>Exemplo: integer(2,4) = 3</div>
integerOfSize	Integer	*Size:: Integer	<div>Gera um número inteiro com o número de dígitos igual a Size.</div> <div>Exemplo: integerOfSize(4) = 1234</div>
letter	Character	Nenhum	<div>Gera um caractere aleatório.</div> <div>Exemplo: letter() = 'a'</div>
letter	Character	Type:: String	<div>Gera um caractere aleatório em maiúscula ou minúscula, dado em Type. O Type pode ser dado em português ou inglês.</div> <div>Exemplo: letter("uppercase") = 'A'</div>
lorem	String	<div>Unid:: String,</div> <div>*Num:: Integer</div>	<div>Gera Num palavras, frases ou parágrafos de <i>lorem ipsum</i>. A variável Unid tem de corresponder a "palavras", "frases" ou "parágrafos", podendo ser dada em português ou inglês.</div> <div>Exemplo: lorem("words", 3) = "mollit fugiat officia"</div>
lorem	String	<div>Unid:: String,</div> <div>*Min:: Integer,</div> <div>*Max:: Integer</div>	<div>Gera entre Min e Max palavras, frases ou parágrafos de <i>lorem ipsum</i>. A variável Unid tem de corresponder a "palavras", "frases" ou "parágrafos", podendo ser dada em português ou inglês.</div> <div>Exemplo: lorem("words", 1, 5) = "aute voluptate dolore veniam"</div>
objectID	String	Nenhum	<div>Gera um ID aleatório com 24 bytes.</div> <div>Exemplo: objectID() = "6048e87b9281fc9a1afe8e61"</div>
position	String	Nenhum	<div>Gera um conjunto de coordenadas geográficas aleatórias.</div> <div>Exemplo: position() = "(67.95632, -55.44137)"</div>
position	String	<div>*[*MinLat,*MaxLat]:: [Float],</div> <div>*[*MinLon,*MaxLon]:: [Float]</div>	<div>Gera um conjunto de coordenadas cartesianas aleatórias, dentro dos limites dados.</div> <div>Exemplo: position([0.03,3],[-5,-2.4]) = "(0.26275, -4.03904)"</div>
pt_phone_number	String	Nenhum	<div>Gera um número de telemóvel português.</div> <div>Exemplo: pt_phone_number() = "911 154 239"</div>
pt_phone_number	String	Bool:: Boolean	<div>Gera um número de telemóvel português e com Bool==true é colocada a extensão.</div> <div>Exemplo: pt_phone_number(true) = "+351 911 154 239"</div>
random	Object	<div>*Arg1:: Object,</div> <div>...</div> <div>*ArgN:: Object</div>	<div>Retorna aleatoriamente um dos argumentos passados à função.</div> <div>Exemplo: random("blue", true, false, 23, 17.56) = 23</div>
time	String	<div>Form:: String,</div> <div>Inter:: Integer,</div> <div>Unid:: Boolean</div>	<div>Gera uma hora do dia aleatória.</div> <div>Form corresponde ao formato em termos de horas, minutos e segundos e pode ser um dos seguintes: "hh:mm:ss", "hh:mm", "hh:ss", "hh:mm", "hh", "mm", "ss".</div> <div>Inter pode ser 12 ou 24 e representa o intervalo de horas usado. Se for selecionado o formato de relógio de 12 horas, as horas do tempo gerado estarão sempre entre a 01 e as 12,</div>

e o tempo virá com "AM" ou "PM" à frente para indicar se é de manhã ou de tarde.

Unid é um booleano que indica se se pretende unidades no tempo gerado ou não. Em caso afirmativo, em vez do formato "hh:mm:ss", o tempo virá com as unidades "h min s".

Exemplo1: time("hh:mm:ss", 12, false) = "09:21:01 AM"

Exemplo2: time("hh:mm:ss", 24, true) = "22h 46min 30s"

Exemplo3: time("hh:mm", 12, true) = "11h 43min PM"

time	<code>String</code>	Form:: <code>String</code> , Inter:: <code>Integer</code> , Unid:: <code>Boolean</code> , *Init:: <code>String</code> , *Fim:: <code>String</code>	Gera uma hora do dia aleatória entre os tempos Init e Fim. Form corresponde ao formato em termos de horas, minutos e segundos e pode ser um dos seguintes: "hh:mm:ss", "hh:mm", "hh:ss", "hh:mm", "hh", "mm", "ss". Inter pode ser 12 ou 24 e representa o intervalo de horas usado. Se for selecionado o formato de relógio de 12 horas, as horas do tempo gerado estarão sempre entre a 01 e as 12, e o tempo virá com "AM" ou "PM" à frente para indicar se é de manhã ou de tarde. Unid é um booleano que indica se se pretende unidades no tempo gerado ou não. Em caso afirmativo, em vez do formato "hh:mm:ss", o tempo virá com as unidades "h min s". A <i>String</i> dos argumentos Init e Fim tem de ter o formato "hh:mm:ss" e corresponder a uma hora do dia válida. Exemplo1: time("hh:mm:ss", 12, false, "13:30:00", "15:15:15") = "02:18:08 PM" Exemplo2: time("hh:mm:ss", 24, true, "13:30:00", "15:15:15") = "14h 43min 03s" Exemplo3: time("hh:mm", 24, true, "13:30:00", "15:15:15") = "13h 39min"
-------------	---------------------	--	---

Funções com suporte de datasets:

MOUSTACHE	TIPO	ARGUMENTO(S)	DESCRIÇÃO
actor	<code>String</code>	Nenhum	Gera um nome de um ator aleatório. Exemplo: actor() = "Ian McKellen"
animal	<code>String</code>	Nenhum	Gera um animal aleatório. Exemplo: animal() = "Baleia"
brand	<code>String</code>	Nenhum	Gera uma marca aleatória. Exemplo: brand() = "Ford"
buzzword	<code>String</code>	Nenhum	Gera uma buzzword aleatória. Exemplo: buzzword() = "Ansiedade"
capital	<code>String</code>	Nenhum	Gera uma capital aleatória. Exemplo: capital() = "Cairo"
car_brand	<code>String</code>	Nenhum	Gera uma marca de carro aleatória. Exemplo: car_brand() = "Lamborghini"
continent	<code>String</code>	Nenhum	Gera um continente aleatório. Exemplo: continent() = "África"
country	<code>String</code>	Nenhum	Gera um país aleatório. Exemplo: country() = "Alemanha"
cultural_center	<code>String</code>	Nenhum	Gera um centro cultural aleatório. Exemplo: cultural_center() = "Stonehenge"
firstName	<code>String</code>	Nenhum	Gera um nome próprio aleatório. Exemplo: firstName() = "Hugo"
fullName	<code>String</code>	Nenhum	Gera um nome completo aleatório. Exemplo: surname() = "Isa Filipa Oliveirinha"
gov_entity	<code>String</code>	Nenhum	Gera uma entidade governamental aleatória. Exemplo: gov_entity() = "ASAE"
hacker	<code>String</code>	Nenhum	Gera um nome de um hacker aleatório. Exemplo: hacker() = "ioerror"
job	<code>String</code>	Nenhum	Gera uma profissão aleatório. Exemplo: iob() = "Sapateiro"

			Exemplo: <code>gerar_geracao() = "Portugal"</code>
month	<code>String</code>	Nenhum	Gera um mês aleatório. Exemplo: <code>month()</code> = "Março"
musician	<code>String</code>	Nenhum	Gera um músico aleatório. Exemplo: <code>musician()</code> = "John Lennon"
nationality	<code>String</code>	Nenhum	Gera uma nacionalidade aleatória. Exemplo: <code>nationality()</code> = "Australiano"
political_party	<code>Object</code>	Nenhum	Gera um partido político aleatório e devolve um objeto com a abreviação e o nome correspondentes. Exemplo: <code>political_party()</code> = { "party_abbr": "Fr", "party_name": "Aliança de Paz" }
political_party	<code>Object</code>	<code>*Cty:: String</code>	Gera um partido político aleatório do país dado em <code>Cty</code> e devolve um objeto com a abreviação e o nome correspondentes. O nome do país pode ser dado em português ou inglês. Exemplo: <code>political_party("Portugal")</code> = { "party_abbr": "BE", "party_name": "Bloco de Esquerda" }
political_party_abbr	<code>String</code>	Nenhum	Gera a abreviatura de um partido político aleatório. Exemplo: <code>political_party_abbr()</code> = "Fr"
political_party_abbr	<code>Object</code>	<code>*Cty:: String</code>	Gera a abreviatura de um partido político aleatório do país dado em <code>Cty</code> . O nome do país pode ser dado em português ou inglês. Exemplo: <code>political_party_abbr("Portugal")</code> = "BE"
political_party_name	<code>String</code>	Nenhum	Gera o nome de um partido político aleatório. Exemplo: <code>political_party_name()</code> = "Aliança de Paz"
political_party_name	<code>Object</code>	<code>*Cty:: String</code>	Gera o nome de um partido político aleatório do país dado em <code>Cty</code> . O nome do país pode ser dado em português ou inglês. Exemplo: <code>political_party_name("Portugal")</code> = "Bloco de Esquerda"
pt_businessman	<code>String</code>	Nenhum	Gera um empresário português aleatório. Exemplo: <code>pt_businessman()</code> = "Manuel Cordo Boullosa"
pt_city	<code>String</code>	Nenhum	Gera uma cidade portuguesa aleatória. Exemplo: <code>pt_city()</code> = "Braga"
pt_city	<code>String</code>	<code>Def:: String,</code> <code>*Dist:: String</code>	Gera uma cidade portuguesa aleatória do distrito dado em <code>Dist</code> . O primeiro argumento tem de corresponder à's <i>Strings</i> "district" ou "distrito". Exemplo: <code>pt_city("district","Braga")</code> = "Barcelos"
pt_county	<code>String</code>	Nenhum	Gera um concelho português aleatório. Exemplo: <code>pt_county()</code> = "Beja"
pt_county	<code>String</code>	<code>Def:: String,</code> <code>*Dist:: String</code>	Gera um concelho português aleatório do distrito dado em <code>Dist</code> . O primeiro argumento tem de corresponder às <i>Strings</i> "district" ou "distrito". Exemplo: <code>pt_county("district","Braga")</code> = "Barcelos"
pt_county	<code>String</code>	<code>Def:: String,</code> <code>*Coun:: String</code>	Gera o concelho português da freguesia dada em <code>Coun</code> . O primeiro argumento tem de corresponder às <i>Strings</i> "parish" ou "freguesia". Exemplo: <code>pt_county("parish","Tadim")</code> = "Braga"
pt_district	<code>String</code>	Nenhum	Gera um distrito português aleatório. Exemplo: <code>pt_district()</code> = "Braga"
pt_district	<code>String</code>	<code>Def:: String,</code> <code>*Loc:: String</code>	Gera o distrito português do concelho, freguesia ou cidade dada em <code>Loc</code> . O primeiro argumento tem de corresponder às <i>Strings</i> "county", "parish" ou "city", que também podem ser dadas em português. Exemplo1: <code>pt_district("county","Braga")</code> = "Braga" Exemplo2: <code>pt_district("parish","Fermentelos")</code> = "Aveiro" Exemplo2: <code>pt_district("city","Torres Vedras")</code> = "Lisboa"

pt_entity	Object	Nenhum	Gera uma entidade portuguesa aleatória e devolve um objeto com a sigla e a designação correspondentes. Exemplo: pt_entity() = { "abbr": "CMBRG", "name": "Câmara Municipal de Braga" }
pt_entity_abbr	String	Nenhum	Gera a sigla de uma entidade portuguesa aleatória. Exemplo: pt_entity_abbr() = "CMBRG"
pt_entity_name	String	Nenhum	Gera a designação de uma entidade portuguesa aleatória. Exemplo: pt_entity_name() = "Câmara Municipal de Braga"
pt_parish	String	Nenhum	Gera uma freguesia portuguesa aleatória. Exemplo: pt_parish() = "Tadim"
pt_parish	String	Def:: String, *Loc:: String	Gera uma freguesia portuguesa aleatória do distrito ou concelho dado em Loc. O primeiro argumento tem de corresponder às <i>Strings</i> "district" ou "county", que também podem ser dadas em português. Exemplo1: pt_parish("district","Braga") = "Macieira de Rates" Exemplo2: pt_parish("county","Braga") = "Adaúfe"
pt_politician	String	Nenhum	Gera um político português aleatório. Exemplo: pt_politician() = "Francisco Manuel Costa Fernandes"
pt_public_figure	String	Nenhum	Gera uma figura pública portuguesa aleatória. Exemplo: pt_public_figure() = "João Manzarra"
pt_top100_celebrity	String	Nenhum	Gera uma celebridade portuguesa aleatória. Exemplo: pt_top100_celebrity() = "Ricardo Araújo Pereira"
religion	String	Nenhum	Gera uma religião aleatória. Exemplo: religion() = "Protestantismo"
soccer_club	String	Nenhum	Gera um clube de futebol aleatório. Exemplo: soccer_club() = "Liverpool FC"
soccer_club	Object	*Cty:: String	Gera um clube de futebol aleatório do país dado em Cty. O nome do país pode ser dado em português ou em inglês e, para já, só são reconhecidos Portugal, Espanha, Itália, Inglaterra e Alemanha. Exemplo: soccer_club("Portugal") = "SC Braga"
soccer_player	String	Nenhum	Gera um jogador de futebol aleatório. Exemplo: soccer_player() = "Eden Hazard"
sport	String	Nenhum	Gera um desporto aleatório. Exemplo: sport() = "Voleibol"
surname	String	Nenhum	Gera um apelido aleatório. Exemplo: surname() = "Rodrigues"
top100_celebrity	String	Nenhum	Gera uma celebridade aleatória. Exemplo: top100_celebrity() = "Lady Gaga"
weekday	String	Nenhum	Gera um dia da semana aleatório. Exemplo: weekday() = "Quinta-feira"
writer	String	Nenhum	Gera um escritor aleatório. Exemplo: writer() = "J. R. R. Tolkien"

Primitiva 'repeat'

Para especificar o tamanho do dataset, ou de um array aninhado, existe a diretiva **repeat**, onde o utilizador indica a estrutura que pretende replicar (que pode ser qualquer uma, desde um tipo JSON primitivo até um objeto complexo), bem como o número de cópias, ou intervalo de números.

```
1 nomes: [ 'repeat(150,200)': {
2   first: '{{firstName()}}',
3   last: '{{surname()}}'
4 }]
```

O utilizador pode também definir tantas coleções quantas pretenda num único modelo (são consideradas coleções as propriedades chave-valor no nível superior do modelo) e a aplicação retornará o dataset resultante em sintaxe *json-server* - um objeto com uma propriedade por coleção. Durante o processamento do input, a aplicação constrói recursivamente tanto o dataset final como o modelo Strapi para a estrutura especificada, concorrentemente, de maneira a permitir a integração posterior numa API RESTful.

```
1 {
2   nomes: [ 'repeat(10)': '{{fullName()}}' ],
3   animais: [ 'repeat(20)': '{{animal()}}' ]
4 }
```

Primitiva 'range'

Esta implementação da função 'range' é equivalente à implementação do *Python* e retorna uma sequência de números inteiros. A função tem a seguinte sintaxe: **range(start, stop, step)**, onde apenas o argumento *start* é obrigatório.

Se for dado apenas o argumento *start*, a sequência de números resultante começa em 0 e para antes de *start*, incrementando 1 unidade de cada vez.

Se forem dados apenas os argumentos *start* e *stop*, a sequência resultante começa em *start* e para antes de *stop*, incrementando 1 unidade de cada vez.

Se forem dados os 3 argumentos, a sequência resultante começa em *start* e para antes de *stop*, incrementando *step* unidades de cada vez.

Caso o utilizador tente fazer um ciclo infinito, o programa dará erro e avisá-lo-á disso.

```
1 {
2   range1: range(5),
3   range2: range(1,-8),
4   range3: range(-5,20,3)
5 }
```

Primitiva 'unique'

A gramática também disponibiliza uma ferramenta chamada **unique()**, à qual o utilizador pode dar uma função de interpolação, ou uma string interpolada com uma função dessas, como argumento. O **unique** garante que as funções de interpolação às quais é aplicado retornam sempre valores únicos, no contexto do dataset final. Isto é especialmente relevante no que toca a funções de interpolação que vão buscar informação aleatória aos datasets de suporte do DataGen, dentro de uma diretiva **repeat**, visto que não há nenhuma garantia base de que retornarão valores sempre diferentes e o utilizador pode querer que isso não aconteça.

Como tal, a função **unique** apenas tem algum efeito quando aplicada em funções de interpolação dos datasets de suporte ou com a função **random**. Desde que seja um destes dois casos (possivelmente interpolados com strings normais) e haja entradas distintas suficientes nos datasets para todos os elementos do **repeat**, o **unique** garante que todos os objetos do dataset resultante terão um valor diferente na propriedade em questão. Se o utilizador usar como argumento uma string com mais do que uma função de interpolação, também não haverá nenhum efeito - poderá haver combinações de valores repetidas no fim.

De seguida, são apresentados dois exemplos: o primeiro demonstra a utilização correta da função **unique**; o segundo mostra casos de uma abordagem errada (não haver valores distintos suficientes para o **repeat**; não ser uma função de interpolação dos datasets de suporte; mais do que uma função de interpolação na string) que ou não funcionarão ou não terão nenhuma garantia de exclusividade mútua para os valores resultantes:

```
1 // Primeiro caso
2 [ 'repeat(6)': {
3   continente: unique('{{continent()}}'),
4   país: unique('Country: {{country()}}'),
5   aleatório: unique('{{random(1,2,3,4,5,6)}}')
6 } ]
7 // Segundo caso
8 [ 'repeat(10)': {
9   continente: unique('{{continent()}}'),
10  inteiro: unique('{{integer(5,20)}}'),
11  aleatório: unique('{{firstName()}} {{surname()}}')
12 } ]
13
```

Primitivas map/filter/reduce

A gramática também disponibiliza uma implementação das ferramentas de programação funcional fundamentais - `map`, `filter` e `reduce`. O utilizador pode utilizar uma ou várias destas funções com um valor array (de qualquer uma das várias formas de declaração de arrays disponibilizadas na gramática). Sintaxe *shorthand* não é permitida, pelo que o utilizador deve sempre abrir chavetas para o bloco de código dentro da função. Tirando isso, esta implementação funciona exatamente como a implementação nativa do Javascript: o utilizador pode declarar a função dentro `map/filter/reduce` ou usar sintaxe anónima para as variáveis; nas variáveis, pode declarar apenas o valor atual ou, adicionalmente, qualquer uma das outras variáveis complementares menos comuns. De seguida, é possível observar vários exemplos distintos do uso destas ferramentas:

```
1 map: range(5).map(value => { return value+1 }),
2 filter: [0,1,2].filter(function(value, index) {return [0,1,2][index]>0}),
3 reduce: range(5).reduce((accum, value, index, array) => {
4     return accum + array[index]
5     }),
6 combinados: range(5).map((value) => { return value+3 })
7     .filter(x => { return x >= 5})
8     .map(x => { return x*2 }).reduce((a,c) => {return a+c})
```

Execução de Código / Funções

De volta às propriedades do modelo, o utilizador pode também usar funções de Javascript para definir o seu valor. Existem dois tipos de funções: funções assinadas, onde o nome do método corresponde à chave da propriedade e o resultado do corpo da função ao valor; funções *arrow* anónimas, que são usadas para indicar apenas o valor da propriedade (a chave precisa de ser especificada antes da função).

```
1 nome: "André",
2 email(gen) {
3     var i = gen.integer(1,30);
4     return `${this.nome}.${gen.surname()}${i}@gmail.com`.toLowerCase();
5 },
6 probabilidade: gen => { return Math.random() * 100; }
```

Dentro destas funções, o utilizador é livre de escrever todo o código de Javascript que pretenda, que será posteriormente executado para determinar o valor da propriedade. Desta forma, torna-se possível incorporar algoritmos mais complexos na lógica de construção do dataset, permitindo uma ferramenta de geração mais especializada e versátil. Dado que o utilizador tem acesso a toda a sintaxe de Javascript, pode também fazer uso de operadores relacionais e lógicos para elaborar condições sobre os dados pretendidos, bem como de métodos funcionais (por exemplo `map` e `filter`, que são implementados pelo Javascript).

Dentro destes blocos de código, o utilizador tem acesso total a qualquer propriedade declarada acima no modelo da DSL, através da variável local `this`, bem como a qualquer função de interpolação disponível na gramática, através da variável local `gen` - sempre que usar uma função para definir o valor de uma propriedade do modelo, o utilizador tem a necessidade de declarar este argumento na sua assinatura, ao qual pode aceder posteriormente no corpo da função para aceder às ditas funções de interpolação. Tudo isto pode ser observado no exemplo acima.

É ainda de realçar que, caso o utilizador tente fazer uma invocação inválida de uma função de interpolação através do `gen`, por exemplo um `integer("string", 10)` ou um `political_party("Oceano Pacífico")`, será retornado um valor `null` ou `false` dentro da função Javascript, por isso o utilizador precisa de verificar com bastante cuidado se o nome da função e os argumentos estão corretos.

Geração Difusa

A gramática também permite geração difusa de propriedades, isto é, a imposição de restrições sobre a existência de certas propriedades, com base em condições lógicas ou probabilidades. A gramática possui quatro ferramentas diferentes para este fim:

- Diretivas `missing/having` - como argumento, recebem a probabilidade de as propriedades contidas nelas (não) existirem no dataset final; esta probabilidade é calculada para cada elemento, originando assim um dataset onde alguns elementos podem ter as propriedades em questão e outros não:

```
1 missing(50) { prop1: 1, prop2: 2 },
2 having(80) { prop3: 3 }
```

- Condições `if... else if... else` - estas funcionam como em qualquer linguagem de programação: o utilizador pode usar operadores relacionais e outros condicionais para criar condições e juntá-las com a ajuda de operadores lógicos. O objeto final terá as propriedades especificadas no bloco da primeira condição que se verificar (ou eventualmente nenhuma delas, se todas as condições forem falsas). Nestas condições, semelhante ao modo de funcionamento das funções, o utilizador tem acesso ilimitado a todas as propriedades declaradas acima no modelo da DSL, bem como a todas as funções de interpolação, o que cria a possibilidade de relacionar causalmente diferentes propriedades:

```
1 tipo: '{{random("A","B","C")}}',
2 if (this.tipo == "A") { A: "tipo A" }
3 else if (this.tipo == "B") { B: "tipo B" }
4 else {C: "tipo C" }
```


- A diretiva **or** - a gramática disponibiliza este operador lógico para permitir a prototipagem ágil de propriedades mutuamente exclusivas, onde apenas uma delas será selecionada aleatoriamente para cada objeto (note-se que não faz sentido criar uma diretiva **and**, pois isso seria equivalente a simplesmente listar as propriedades pretendidas normalmente no modelo da DSL):

```
1 or() {
2     prop1: 1,
3     prop2: 2,
4     prop3: 3
5 }
```

- A diretiva **at_least** - dentro deste bloco, o utilizador escreve um conjunto de propriedades e dá como argumento o número mínimo dessas propriedades que deve estar presente no objeto final. O compilador seleciona um conjunto dessas propriedades aleatoriamente, entre o mínimo dado e o total:

```
1 at_least(2) {
2     prop1: 1,
3     prop2: 2,
4     prop3: 3
5 }
```

Datasets

Em relação à API de datasets, a equipa realizou uma pesquisa extensiva por datasets com informação útil, utilizou os datasets bem-estruturados que encontrou e aproveitou a informação que pode dos restantes, processando-a para remover erros e normalizar o seu conteúdo, posteriormente agrupando todos os dados da mesma categoria de forma a criar datasets maiores e dotados de maior complexidade para colocar ao serviço do utilizador.

A equipa também criou alguns datasets originais manualmente, para tópicos considerados relevantes, e introduziu suporte bilingue - português e inglês - em todos os datasets disponibilizados na aplicação, de forma a dar ao utilizador a liberdade de escolher a linguagem que melhor se adequa ao seu objetivo. Para indicar a sua linguagem de eleição, o modelo do utilizador deve começar com a seguinte sintaxe:

```
1 <!LANGUAGE pt> // ou alternativamente
2 <!LANGUAGE en>
```

Atualmente, o DataGen tem datasets de suporte para todas as categorias listadas de seguida:

- atores;
- animais;
- capitais;
- centros culturais;
- cidades;
- clubes de futebol;
- continentes;
- desportos;
- dias da semana
- distritos, cidades, concelhos e freguesias portuguesas;
- empresários portugueses;
- entidades governamentais;
- escritores;
- figuras públicas portuguesas;
- futebolistas;
- hackers;
- marcas;
- marcas de carros;
- meses;
- músicos;
- nacionalidades;
- nomes;
- países;
- *buzzwords*
- partidos políticos;
- profissões
- políticos portugueses;
- religiões;
- top 100 celebridades;
- top 100 celebridades portuguesas.

Rotas Aplicacionais

As seguintes rotas REST são disponibilizadas como alternativa à interface gráfica:

- **POST /api/datagen/**
O corpo tem de conter o modelo da gramática (DSL) em formato *raw*, isto é, não envolvido por aspas.

Esta rota gera um objeto JSON com o dataset (também em JSON) e campos adicionais necessários para a sua inserção no Strapi.
- **POST /api/datagen/json**
O corpo tem de conter o modelo da gramática (DSL) em formato *raw*, isto é, não envolvido por aspas.

Esta rota retorna apenas o dataset resultante em formato JSON.
- **POST /api/datagen/xml**
O corpo tem de conter o modelo da gramática (DSL) em formato *raw*, isto é, não envolvido por aspas.

Esta rota retorna apenas o dataset resultante em formato XML.
- **POST /api/datagen/csv**
O corpo tem de conter o modelo da gramática (DSL) em formato *raw*, isto é, não envolvido por aspas.

Esta rota retorna apenas o dataset resultante em formato CSV.

Para além destas rotas, são também disponibilizadas rotas para obter os datasets utilizados na aplicação:

- **GET /api/dataset/:nome**
Este **nome** terá de ser um dos seguintes:
 - actors
 - animals
 - brands
 - buzzwords
 - capitals
 - car_brands
 - continents
 - countries
 - cultural_centers
 - gov_entities
 - hackers
 - jobs
 - months
 - musicians
 - names
 - nationalities
 - political_parties
 - pt_businessmen
 - pt_districts
 - pt_entities
 - pt_politicians
 - pt_public_figures
 - pt_top100_celebrities
 - religions
 - soccer_clubs
 - soccer_players
 - sports
 - top100_celebrities
 - weekdays
 - writers

Informação Adicional

O principal propósito desta ferramenta é a geração de datasets (arrays de objetos) a partir do modelo do utilizador, pelo que foi projetada na expectativa de que o valor de cada coleção do modelo seja definido por um 'repeat'.

Caso se definam as coleções de outra maneira (tipos primitivos, estruturas de dados explícitas ou funções), o dataset será gerado corretamente na mesma, mas a API não.

Modelos Exemplo

O primeiro exemplo refere-se a um conjunto de utilizadores que utilizam uma aplicação de leitura de livros. As seguintes informações são necessárias:

- **nome**
- **idade**
- **número do BI/CC**
- **descrição** que é opcional (há uma chance de 70%)
- **número de livros** lidos no total
- **listagem dos livros** lidos:
 - **título**
 - **suporte** (físico ou digital)
 - **rating** do utilizador
- **livro favorito** da lista anterior

```
1 <!--LANGUAGE pt-->
2 {
3   perfil: [
4     'repeat(3)': {
5       nome: '{{fullName()}}',
6       idade: '{{integer(15,60)}}',
7       or() {
8         BI: '{{integerOfSize(8)}}-{{integer(0,9)}}',
9         CC: '{{integerOfSize(8)}}-{{integer(0,9)}}-{{letter("uppercase")}}{{letter("uppercase")}}{{integer(0,9)}}'
10      },
11      having(70) {
12        descrição: '{{lorem("sentences", 1)}}'
13      },
14      nr_livros: '{{integer(1,10)}}',
15      livros: [ 'repeat(this.nr_livros)': {
16        titulo: '{{lorem("words",1)}}',
17        suporte: '{{random("Físico","Digital")}}',
18        rating: '{{integer(1,5)}}'
19      } ],
20      livro_favorito(gen) {
21        var titulos = this.livros.map(x => x.titulo)
22        return gen.random(...titulos)
23      }
24    ]
25  ]
26 }
```

O segundo exemplo refere-se a um caso académico da representação, utilizando grafos, de 100 cidades portuguesas únicas, que atuam como os seus nodos. Estas têm:

- **id** que é dado por `c{N}` onde N é um número inteiro
- **nome**
- **população** que é um número inteiro entre 1500 e 550000
- **descrição**
- **distrito** da cidade

As arestas são as ligações entre as cidades (2000 no total):

- **id** que é dado por `l{id_ligação}-{id_origem}-{id_destino}`
- **origem** que é o **id** da cidade onde a ligação começa
- **destino** que é o **id** da cidade onde a ligação termina
- **distância** que é um número entre 5 e 600

```
1 <!--LANGUAGE pt-->
2 {
3   cidades: [ 'repeat(100)': {
4     id_cidade: 'c{{index(1)}}',
5     nome: unique('{{pt_city()}}'),
6     população: '{{integer(1500, 550000)}}',
7     descrição: '{{lorem("paragraphs", 1)}}',
8     distrito: '{{pt_district("city", this.nome)}}'
9   }],
10   ligações(gen) {
11     var id = 1
12     var cidades = this.cidades.map(x => x.id_cidade)
13     var possiveis = cidades.flatMap((v, i) => cidades.slice(i+1).map( w => v + '|' + w ))
14     var ligs = []
15
16     for (var i = 0; i < 2000; i++) {
17       let l = gen.random(...possiveis)
18       possiveis.splice(possiveis.indexOf(l), 1)
```

```

19
20     var split = l.split('|')
21     ligs.push({
22         id_ligação: `l${id++}-${split[0]}-${split[1]}`,
23         origem: split[0],
24         destino: split[1],
25         distância: gen.float(5, 600)
26     })
27 }
28
29 return ligs
30 }
31 }

```

O terceiro exemplo refere-se a um pequeno excerto de Autos de Eliminação que são, resumidamente, uma estrutura que deve ser criada e cuidadosamente preenchida de maneira a eliminar documentação que atinja o prazo da sua conservação administrativa de forma segura.

Neste exemplo serão abordadas apenas as seguintes secções desse documento:

- **tipo** da fonte de legitimação, que poderá ser "PGD/LC", "TS/LC", "PGD", "RADA" ou "RADA/CLAV"
- **fundos** que dependem do **tipo**:
 - se o tipo for "PGD/LC", "TS/LC" ou "PGD", só tem um fundo (entidade)
 - se não, tem entre 1 a 5 fundos (entidades)
- **classes** que são entre 2 e 5 e que dependem do **tipo**:
 - se o tipo for "PGD/LC" ou "TS/LC", tem um código (que pode ter 3 ou 4 níveis)
 - se não, pode ter um código, uma referência ou ambos

```

1 <!LANGUAGE pt>
2 {
3   autoEliminação: {
4     fonteLegitimação: {
5       tipo: '{{random("PGD/LC", "TS/LC", "PGD", "RADA", "RADA/CLAV")}}'
6     },
7     fundos(gen) {
8       if ([ "PGD/LC", "TS/LC", "PGD" ].includes(this.fonteLegitimação.tipo))
9         return [gen.pt_entity()]
10      else {
11        var arr = []
12        for (var i = 0; i < gen.integer(1,5); i++) arr.push(gen.pt_entity())
13        return arr
14      }
15    },
16    classes: [ 'repeat(2,5)': {
17      if ([ "PGD/LC", "TS/LC" ].includes(this.fonteLegitimação.tipo)) {
18        código: gen => {
19          var nivel1 = gen.random(...gen.range(100,950,50))
20          var nivel2 = gen.random(10,20,30,40,50)
21          var nivel3 = gen.integer(1,999,3)
22          var nivel4 = gen.random("01", "02")
23
24          var classe = nivel1 + '.' + nivel2 + '.' + nivel3
25          if (Math.random() > 0.5) classe += '.' + nivel4
26          return classe
27        }
28      }
29      else {
30        at_least(1) {
31          código(gen) {
32            var nivel1 = gen.random(...gen.range(100,950,50))
33            var nivel2 = gen.random(10,20,30,40,50)
34            var nivel3 = gen.integer(001,999)
35            var nivel4 = gen.random("01", "02")
36
37            var classe = nivel1 + '.' + nivel2 + '.' + nivel3
38            if (Math.random() > 0.5) classe += '.' + nivel4
39            return classe
40          },
41          referência: '{{random(1,2,3,55,56)}}'
42        }
43      }
44    }
45  }
46 }

```

