Cálculo de Programas
Trabalho Prático
LEI — 2022/23

Departamento de Informática
Universidade do Minho

Janeiro de 2023

Grupo nr.	29
93282	Pedro Ferreira
95076	Pedro Oliveira
96547	Rodrigo Freitas
96794	Ricardo Oliveira

Preâmbulo

Cálculo de Programas tem como objectivo principal ensinar a programação de computadores como uma disciplina científica. Para isso parte-se de um repertório de *combinadores* que formam uma álgebra da programação (conjunto de leis universais e seus corolários) e usam-se esses combinadores para construir programas *composicionalmente*, isto é, agregando programas já existentes.

Na sequência pedagógica dos planos de estudo dos cursos que têm esta disciplina, opta-se pela aplicação deste método à programação em Haskell (sem prejuízo da sua aplicação a outras linguagens funcionais). Assim, o presente trabalho prático coloca os alunos perante problemas concretos que deverão ser implementados em Haskell. Há ainda um outro objectivo: o de ensinar a documentar programas, a validá-los e a produzir textos técnico-científicos de qualidade.

Antes de abodarem os problemas propostos no trabalho, os grupos devem ler com atenção o anexo A onde encontrarão as instruções relativas ao sofware a instalar, etc.

Problema 1

Suponha-se uma sequência numérica semelhante à sequência de Fibonacci tal que cada termo subsequente aos três primeiros corresponde à soma dos três anteriores, sujeitos aos coeficientes a, b e c:

$$
\begin{aligned}
&f\ a\ b\ c\ 0 = 0\\
&f\ a\ b\ c\ 1 = 1\\
&f\ a\ b\ c\ 2 = 1\\
&f\ a\ b\ c\ (n+3) = a * f\ a\ b\ c\ (n+2) + b * f\ a\ b\ c\ (n+1) + c * f\ a\ b\ c\ n
\end{aligned}
$$

Assim, por exemplo, $f\ 1\ 1\ 1$ irá dar como resultado a sequência:

$$1, 1, 2, 4, 7, 13, 24, 44, 81, 149, \ldots$$

$f\ 1\ 2\ 3$ irá gerar a sequência:

$$1, 1, 3, 8, 17, 42, 100, 235, 561, 1331, \ldots$$

etc.

A definição de f dada é muito ineficiente, tendo uma degradação do tempo de execução exponencial. Pretende-se otimizar a função dada convertendo-a para um ciclo *for*. Recorrendo à lei de recursividade mútua, calcule $loop$ e $initial$ em

$$fbl\ a\ b\ c = wrap \cdot \mathsf{for}\ (loop\ a\ b\ c)\ initial$$

por forma a f e fbl serem (matematicamente) a mesma função. Para tal, poderá usar a regra prática explicada no anexo B.

Valorização: apresente testes de *performance* que mostrem quão mais rápida é fbl quando comparada com f.

Problema 2

Pretende-se vir a classificar os conteúdos programáticos de todas as UCs lecionadas no *Departamento de Informática* de acordo com o ACM Computing Classification System. A listagem da taxonomia desse sistema está disponível no ficheiro `Cp2223data`, começando com

```
acm_ccs = ["CCS",
"    General and reference",
"        Document types",
"            Surveys and overviews",
"            Reference works",
"            General conference proceedings",
"            Biographies",
"            General literature",
"            Computing standards, RFCs and guidelines",
"        Cross-computing tools and techniques",
```

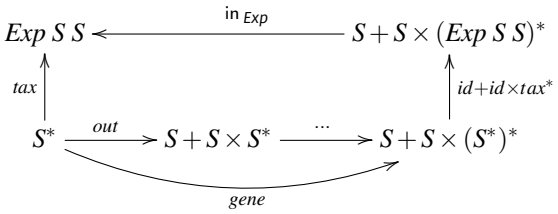
(10 primeiros ítens) etc., etc.[1]

Pretende-se representar a mesma informação sob a forma de uma árvore de expressão, usando para isso a biblioteca Exp que consta do material padagógico da disciplina e que vai incluída no zip do projecto, por ser mais conveniente para os alunos.

1. Comece por definir a função de conversão do texto dado em acm_ccs (uma lista de *strings*) para uma tal árvore como um anamorfismo de Exp:

   $$tax :: [String] \to Exp\ String\ String$$
   $$tax = [\![gene]\!]_{Exp}$$

   Ou seja, defina o $gene$ do anamorfismo, tendo em conta o seguinte diagrama[2]:

   Para isso, tome em atenção que cada nível da hierarquia é, em acm_ccs, marcado pela indentação de 4 espaços adicionais — como se mostra no fragmento acima.

   Na figura 1 mostra-se a representação gráfica da árvore de tipo Exp que representa o fragmento de acm_ccs mostrado acima.

2. De seguida vamos querer todos os caminhos da árvore que é gerada por tax, pois a classificação de uma UC pode ser feita a qualquer nível (isto é, caminho descendente da raiz `"CCS"` até um subnível ou folha). [3]

[1] Informação obtida a partir do site ACM CCS selecionando *Flat View*.
[2] S abrevia $String$.
[3] Para um exemplo de classificação de UC concreto, pf. ver a secção **Classificação ACM** na página pública de Cálculo de Programas.

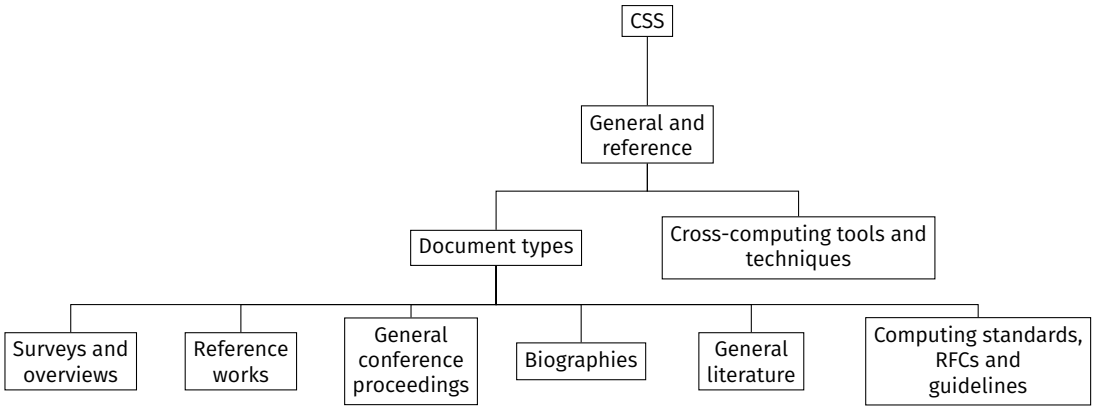


Figura 1: Fragmento de acm_ccs representado sob a forma de uma árvore do tipo Exp.

Precisamos pois da composição de tax com uma função de pós-processamento $post$,

$$tudo :: [String] \to [[String]]$$
$$tudo = post \cdot tax$$

para obter o efeito que se mostra na tabela 1.

CCS			
CCS	General and reference		
CCS	General and reference	Document types	
CCS	General and reference	Document types	Surveys and overviews
CCS	General and reference	Document types	Reference works
CCS	General and reference	Document types	General conference proceedings
CCS	General and reference	Document types	Biographies
CCS	General and reference	Document types	General literature
CCS	General and reference	Cross-computing tools and techniques	

Tabela 1: Taxonomia ACM fechada por prefixos (10 primeiros ítens).

Defina a função $post :: Exp\ String\ String \to [[String]]$ da forma mais económica que encontrar.

Sugestão: Inspecione as bibliotecas fornecidas à procura de funções auxiliares que possa re-utilizar para a sua solução ficar mais simples. Não se esqueça que, para o mesmo resultado, nesta disciplina *"ganha" quem escrever menos código*!

Sugestão: Para efeitos de testes intermédios não use a totalidade de acm_ccs, que tem 2114 linhas! Use, por exemplo, $take\ 10\ acm\_ccs$, como se mostrou acima.

Problema 3

O tapete de Sierpinski é uma figura geométrica fractal em que um quadrado é subdividido recursivamente em sub-quadrados. A construção clássica do tapete de Sierpinski é a seguinte: assumindo um quadrado de lado l, este é subdivido em 9 quadrados iguais de lado $l/3$, removendo-se o quadrado central. Este passo é depois repetido sucessivamente para cada um dos 8 sub-quadrados restantes (Fig. 2).

NB: No exemplo da fig. 2, assumindo a construção clássica já referida, os quadrados estão a branco e o fundo a verde.

A complexidade deste algoritmo, em função do número de quadrados a desenhar, para uma profundidade n, é de 8^n (exponencial). No entanto, se assumirmos que os quadrados a desenhar são os que estão a verde, a complexidade é reduzida para $\sum_{i=0}^{n-1} 8^i$, obtendo um ganho de $\sum_{i=1}^{n} \frac{100}{8^i}\%$. Por exemplo, para $n = 5$, o ganho é de 14.28%. O objetivo deste problema é a implementação do algoritmo mediante a referida otimização.

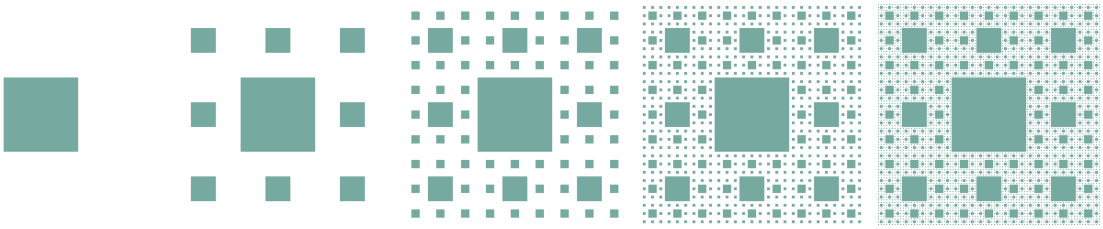


Figura 2: Construção do tapete de Sierpinski com profundidade 5.

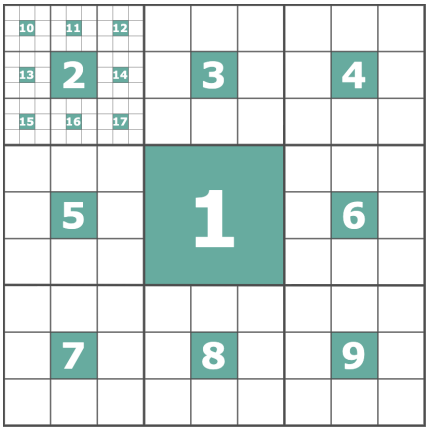


Figura 3: Tapete de Sierpinski com profundidade 2 e com os quadrados enumerados.

Assim, seja cada quadrado descrito geometricamente pelas coordenadas do seu vértice inferior esquerdo e o comprimento do seu lado:

type $Square = (Point, Side)$
type $Side = Double$
type $Point = (Double, Double)$

A estrutura recursiva de suporte à construção de tapetes de Sierpinski será uma Rose Tree, na qual cada nível da árvore irá guardar os quadrados de tamanho igual. Por exemplo, a construção da fig. 3 poderá[4] corresponder à árvore da figura 4.

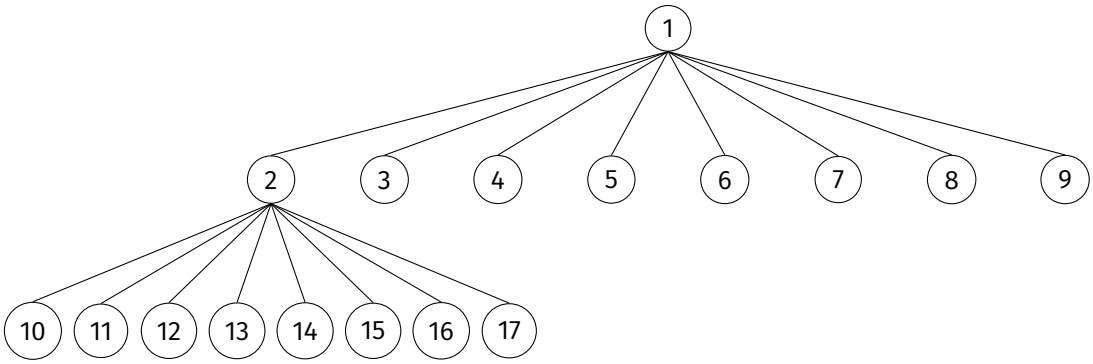


Figura 4: Possível árvore de suporte para a construção da fig. 3.

Uma vez que o tapete é também um quadrado, o objetivo será, a partir das informações do tapete (coordenadas do vértice inferior esquerdo e comprimento do lado), desenhar o quadrado central, subdividir o tapete nos 8 sub-tapetes restantes, e voltar a desenhar, recursivamente, o quadrado nesses 8 sub-tapetes. Desta forma, cada tapete determina o seu quadrado e os seus 8 sub-tapetes. No exemplo em cima, o tapete que contém o quadrado 1 determina esse próprio quadrado e determina os sub-tapetes que contêm os quadrados 2 a 9.

[4] A ordem dos filhos não é relevante.

Portanto, numa primeira fase, dadas as informações do tapete, é construida a árvore de suporte com todos os quadrados a desenhar, para uma determinada profundidade.

$$squares :: (Square, Int) \to Rose\ Square$$

NB: No programa, a profundidade começa em 0 e não em 1.

Uma vez gerada a árvore com todos os quadrados a desenhar, é necessário extrair os quadrados para uma lista, a qual é processada pela função *drawSq*, disponibilizada no anexo D.

$$rose2List :: Rose\ a \to [a]$$

Assim, a construção de tapetes de Sierpinski é dada por um hilomorfismo de *Rose Trees*:

$$sierpinski :: (Square, Int) \to [Square]$$
$$sierpinski = [\![gr2l, gsq]\!]_R$$

Trabalho a fazer:

1. Definir os genes do hilomorfismo *sierpinski*.
2. Correr

$$\begin{array}{l} sierp4 = drawSq\ (sierpinski\ (((0,0),32),3)) \\ constructSierp5 = \mathbf{do}\ drawSq\ (sierpinski\ (((0,0),32),0)) \\ \quad await \\ \quad drawSq\ (sierpinski\ (((0,0),32),1)) \\ \quad await \\ \quad drawSq\ (sierpinski\ (((0,0),32),2)) \\ \quad await \\ \quad drawSq\ (sierpinski\ (((0,0),32),3)) \\ \quad await \\ \quad drawSq\ (sierpinski\ (((0,0),32),4)) \\ \quad await \end{array}$$

3. Definir a função que apresenta a construção do tapete de Sierpinski como é apresentada em *construcaoSierp5*, mas para uma profundidade $n \in \mathbb{N}$ recebida como parâmetro.

$$constructSierp :: Int \to \mathsf{IO}\ [()]$$
$$constructSierp = present \cdot carpets$$

Dica: a função *constructSierp* será um hilomorfismo de listas, cujo anamorfismo $carpets :: Int \to [[Square]]$ constrói, recebendo como parâmetro a profundidade n, a lista com todos os tapetes de profundidade $1..n$, e o catamorfismo $present :: [[Square]] \to \mathsf{IO}\ [()]$ percorre a lista desenhando os tapetes e esperando 1 segundo de intervalo.

Problema 4

Este ano ocorrerá a vigésima segunda edição do Campeonato do Mundo de Futebol, organizado pela Federação Internacional de Futebol (FIFA), a decorrer no Qatar e com o jogo inaugural a 20 de Novembro.

Uma casa de apostas pretende calcular, com base numa aproximação dos *rankings*[5] das seleções, a probabilidade de cada seleção vencer a competição.

Para isso, o diretor da casa de apostas contratou o Departamento de Informática da Universidade do Minho, que atribuiu o projeto à equipa formada pelos alunos e pelos docentes de Cálculo de Programas.

[5] Os *rankings* obtidos aqui foram escalados e arredondados.

Para resolver este problema de forma simples, ele será abordado por duas fases:
1. versão académica sem probabilidades, em que se sabe à partida, num jogo, quem o vai vencer;
2. versão realista com probabilidades usando o mónade *Dist* (distribuições probabilísticas) que vem descrito no anexo C.

A primeira versão, mais simples, deverá ajudar a construir a segunda.

Descrição do problema

Uma vez garantida a qualificação (já ocorrida), o campeonato consta de duas fases consecutivas no tempo:

1. fase de grupos;
2. fase eliminatória (ou "mata-mata", como é habitual dizer-se no Brasil).

Para a fase de grupos, é feito um sorteio das 32 seleções (o qual já ocorreu para esta competição) que as coloca em 8 grupos, 4 seleções em cada grupo. Assim, cada grupo é uma lista de seleções.

Os grupos para o campeonato deste ano são:

$$
\begin{aligned}
&\textbf{type } \mathit{Team} = \mathit{String}\\
&\textbf{type } \mathit{Group} = [\mathit{Team}]\\
&\mathit{groups} :: [\mathit{Group}]\\
&\mathit{groups} = [[\texttt{"Qatar"}, \texttt{"Ecuador"}, \texttt{"Senegal"}, \texttt{"Netherlands"}],\\
&\quad [\texttt{"England"}, \texttt{"Iran"}, \texttt{"USA"}, \texttt{"Wales"}],\\
&\quad [\texttt{"Argentina"}, \texttt{"Saudi Arabia"}, \texttt{"Mexico"}, \texttt{"Poland"}],\\
&\quad [\texttt{"France"}, \texttt{"Denmark"}, \texttt{"Tunisia"}, \texttt{"Australia"}],\\
&\quad [\texttt{"Spain"}, \texttt{"Germany"}, \texttt{"Japan"}, \texttt{"Costa Rica"}],\\
&\quad [\texttt{"Belgium"}, \texttt{"Canada"}, \texttt{"Morocco"}, \texttt{"Croatia"}],\\
&\quad [\texttt{"Brazil"}, \texttt{"Serbia"}, \texttt{"Switzerland"}, \texttt{"Cameroon"}],\\
&\quad [\texttt{"Portugal"}, \texttt{"Ghana"}, \texttt{"Uruguay"}, \texttt{"Korea Republic"}]]
\end{aligned}
$$

Deste modo, *groups !! 0* corresponde ao grupo A, *groups !! 1* ao grupo B, e assim sucessivamente. Nesta fase, cada seleção de cada grupo vai defrontar (uma vez) as outras do seu grupo.

Passam para o "mata-mata" as duas seleções que mais pontuarem em cada grupo, obtendo pontos, por cada jogo da fase grupos, da seguinte forma:

- vitória — 3 pontos;
- empate — 1 ponto;
- derrota — 0 pontos.

Como se disse, a posição final no grupo irá determinar se uma seleção avança para o "mata-mata" e, se avançar, que possíveis jogos terá pela frente, uma vez que a disposição das seleções está desde o início definida para esta última fase, conforme se pode ver na figura 5.

Assim, é necessário calcular os vencedores dos grupos sob uma distribuição probabilística. Uma vez calculadas as distribuições dos vencedores, é necessário colocá-las nas folhas de uma *LTree* de forma a fazer um *match* com a figura 5, entrando assim na fase final da competição, o tão esperado "mata-mata". Para avançar nesta fase final da competição (i.e. subir na árvore), é preciso ganhar, quem perder é automaticamente eliminado ("mata-mata"). Quando uma seleção vence um jogo, sobe na árvore, quando perde, fica pelo caminho. Isto significa que a seleção vencedora é aquela que vence todos os jogos do "mata-mata".

Arquitetura proposta

A visão composicional da equipa permitiu-lhe perceber desde logo que o problema podia ser dividido, independentemente da versão, probabilística ou não, em duas partes independentes — a da fase de grupos e a do "mata-mata". Assim, duas sub-equipas poderiam trabalhar em paralelo, desde que se

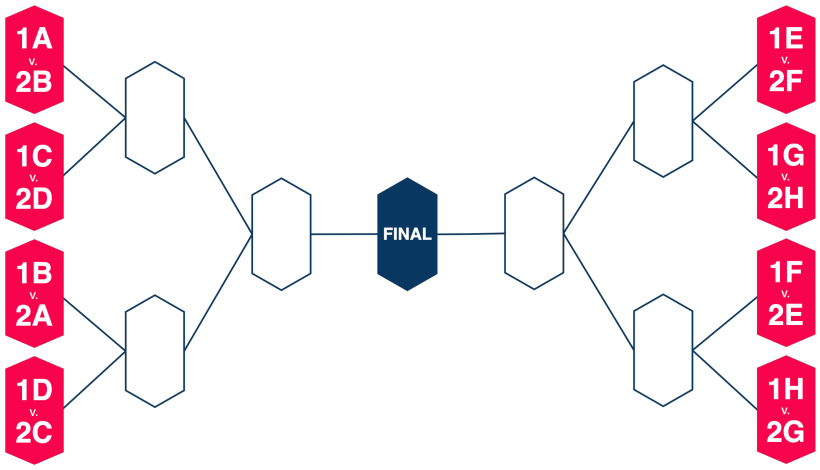


Figura 5: O “mata-mata”

garantisse a composicionalidade das partes. Decidiu-se que os alunos desenvolveriam a parte da fase de grupos e os docentes a do “mata-mata”.

Versão não probabilística

O resultado final (não probabilístico) é dado pela seguinte função:

$$
\begin{aligned}
&winner :: Team \\
&winner = wcup\ groups \\
&wcup = knockoutStage \cdot groupStage
\end{aligned}
$$

A sub-equipa dos docentes já entregou a sua parte:

$$knockoutStage = (\!|\, [id, koCriteria] \,|\!)$$

Considere-se agora a proposta do *team leader* da sub-equipa dos alunos para o desenvolvimento da fase de grupos:

> *Vamos dividir o processo em 3 partes:*
> - *gerar os jogos,*
> - *simular os jogos,*
> - *preparar o “mata-mata” gerando a árvore de jogos dessa fase (fig. 5).*
>
> *Assim:*
>
> $$
> \begin{aligned}
> &groupStage :: [Group] \to LTree\ Team \\
> &groupStage = initKnockoutStage \cdot simulateGroupStage \cdot genGroupStageMatches
> \end{aligned}
> $$
>
> *Comecemos então por definir a função* $genGroupStageMatches$ *que gera os jogos da fase de grupos:*
>
> $$
> \begin{aligned}
> &genGroupStageMatches :: [Group] \to [[Match]] \\
> &genGroupStageMatches = \mathsf{map}\ \ generateMatches
> \end{aligned}
> $$
>
> *onde*
>
> $$\mathbf{type}\ Match = (Team, Team)$$
>
> *Ora, sabemos que nos foi dada a função*
>
> $$gsCriteria :: Match \to Maybe\ Team$$
>
> *que, mediante um certo critério, calcula o resultado de um jogo, retornando* $Nothing$ *em caso de empate, ou a equipa vencedora (sob o construtor* $Just$*). Assim, precisamos de definir a função*

> $simulateGroupStage :: [[Match]] \to [[Team]]$
> $simulateGroupStage = \text{map}\ \ (groupWinners\ gsCriteria)$
>
> *que simula a fase de grupos e dá como resultado a lista dos vencedores, recorrendo à função* $groupWinners$*:*
>
> $groupWinners\ criteria = best\ 2 \cdot consolidate \cdot (\gg\!= matchResult\ criteria)$
>
> *Aqui está apenas em falta a definição da função* $matchResult$*.*
> *Por fim, teremos a função* $initKnockoutStage$ *que produzirá a LTree que a sub-equipa do "mata-mata" precisa, com as devidas posições. Esta será a composição de duas funções:*
>
> $initKnockoutStage = [\![\ glt\ ]\!] \cdot arrangement$

Trabalho a fazer:

1. Definir uma alternativa à função genérica $consolidate$ que seja um catamorfismo de listas:

   $consolidate' :: (Eq\ a, Num\ b) \Rightarrow [(a,b)] \to [(a,b)]$
   $consolidate' = (\!|cgene|\!)$

2. Definir a função $matchResult :: (Match \to Maybe\ Team) \to Match \to [(Team, Int)]$ que apura os pontos das equipas de um dado jogo.
3. Definir a função genérica $pairup :: Eq\ b \Rightarrow [b] \to [(b,b)]$ em que $generateMatches$ se baseia.
4. Definir o gene glt.

Versão probabilística

Nesta versão, mais realista, $gsCriteria :: Match \to (Maybe\ Team)$ dá lugar a

$$pgsCriteria :: Match \to \text{Dist}\ (Maybe\ Team)$$

que dá, para cada jogo, a probabilidade de cada equipa vencer ou haver um empate. Por exemplo, há 50% de probabilidades de Portugal empatar com a Inglaterra,

```
pgsCriteria("Portugal","England")'
        Nothing  50.0%'
 Just "England"  26.7%'
Just "Portugal"  23.3%'
```

etc.

O que é Dist? É o mónade que trata de distribuições probabilísticas e que é descrito no anexo C, página 11 e seguintes. O que há a fazer? Eis o que diz o vosso *team leader*:

> *O que há a fazer nesta versão é, antes de mais, avaliar qual é o impacto de* $gsCriteria$ *virar monádica (em* Dist*) na arquitetura geral da versão anterior. Há que reduzir esse impacto ao mínimo, escrevendo-se tão pouco código quanto possível!*

Todos relembraram algo que tinham aprendido nas aulas teóricas a respeito da "monadificação" do código: há que reutilizar o código da versão anterior, monadificando-o.

Para distinguir as duas versões decidiu-se afixar o prefixo 'p' para identificar uma função que passou a ser monádica.

A sub-equipa dos docentes fez entretanto a monadificação da sua parte:

$pwinner :: \text{Dist}\ Team$
$pwinner = pwcup\ groups$
$pwcup = pknockoutStage \bullet pgroupStage$

E entregou ainda a versão probabilística do "mata-mata":

$$
\begin{aligned}
&pknockoutStage = mcataLTree'\ [return, pkoCriteria]\\
&mcataLTree'\ g = k\ \textbf{where}\\
&\quad k\ (Leaf\ a) = g1\ a\\
&\quad k\ (Fork\ (x,y)) = mmbin\ g2\ (k\ x, k\ y)\\
&\quad g1 = g \cdot i_1\\
&\quad g2 = g \cdot i_2
\end{aligned}
$$

A sub-equipa dos alunos também já adiantou trabalho,

$$pgroupStage = pinitKnockoutStage \bullet psimulateGroupStage \cdot genGroupStageMatches$$

mas faltam ainda *pinitKnockoutStage* e *pgroupWinners*, esta usada em *psimulateGroupStage*, que é dada em anexo.

Trabalho a fazer:

- Definir as funções que ainda não estão implementadas nesta versão.
- **Valorização**: experimentar com outros critérios de "ranking" das equipas.

> **Importante**: (a) código adicional terá que ser colocado no anexo E, obrigatoriamente; (b) todo o código que é dado não pode ser alterado.

Anexos

A Documentação para realizar o trabalho

Para cumprir de forma integrada os objectivos do trabalho vamos recorrer a uma técnica de programação dita "literária" [**?**], cujo princípio base é o seguinte:

> *Um programa e a sua documentação devem coincidir.*

Por outras palavras, o código fonte e a documentação de um programa deverão estar no mesmo ficheiro.

O ficheiro `cp2223t.pdf` que está a ler é já um exemplo de programação literária: foi gerado a partir do texto fonte `cp2223t.lhs`[6] que encontrará no material pedagógico desta disciplina descompactando o ficheiro `cp2223t.zip` e executando:

```
$ lhs2TeX cp2223t.lhs > cp2223t.tex
$ pdflatex cp2223t
```

em que `lhs2tex` é um pré-processador que faz "pretty printing" de código Haskell em LaTeX e que deve desde já instalar utilizando o utiliário cabal disponível em haskell.org.

Por outro lado, o mesmo ficheiro `cp2223t.lhs` é executável e contém o "kit" básico, escrito em Haskell, para realizar o trabalho. Basta executar

```
$ ghci cp2223t.lhs
```

Abra o ficheiro `cp2223t.lhs` no seu editor de texto preferido e verifique que assim é: todo o texto que se encontra dentro do ambiente

```
\begin{code}
...
\end{code}
```

é seleccionado pelo GHCi para ser executado.

[6] O sufixo 'lhs' quer dizer *literate Haskell*.

A.1 Como realizar o trabalho

Este trabalho teórico-prático deve ser realizado por grupos de 3 (ou 4) alunos. Os detalhes da avaliação (datas para submissão do relatório e sua defesa oral) são os que forem publicados na página da disciplina na *internet*.

Recomenda-se uma abordagem participativa dos membros do grupo em todos os exercícios do trabalho, para assim poderem responder a qualquer questão colocada na *defesa oral* do relatório.

Em que consiste, então, o *relatório* a que se refere o parágrafo anterior? É a edição do texto que está a ser lido, preenchendo o anexo E com as respostas. O relatório deverá conter ainda a identificação dos membros do grupo de trabalho, no local respectivo da folha de rosto.

Para gerar o PDF integral do relatório deve-se ainda correr os comando seguintes, que actualizam a bibliografia (com BibTEX) e o índice remissivo (com `makeindex`),

```
$ bibtex cp2223t.aux
$ makeindex cp2223t.idx
```

e recompilar o texto como acima se indicou.

No anexo D, disponibiliza-se algum código Haskell relativo aos problemas apresentados. Esse anexo deverá ser consultado e analisado à medida que isso for necessário.

A.2 Como exprimir cálculos e diagramas em LaTeX/lhs2tex

Como primeiro exemplo, estudar o texto fonte deste trabalho para obter o efeito:[7]

$$
\begin{array}{ll}
 & id = \langle f, g\rangle \\
\equiv & \quad \{\ \text{universal property}\ \} \\
 & \left\{ \begin{array}{l} \pi_1 \cdot id = f \\ \pi_2 \cdot id = g \end{array} \right. \\
\equiv & \quad \{\ \text{identity}\ \} \\
 & \left\{ \begin{array}{l} \pi_1 = f \\ \pi_2 = g \end{array} \right. \\
\Box &
\end{array}
$$

Os diagramas podem ser produzidos recorrendo à *package* LaTeX xymatrix, por exemplo:

$$
\begin{array}{ccc}
\mathbb{N}_0 & \xleftarrow{\text{in}} & 1 + \mathbb{N}_0 \\
{\scriptstyle (\!|g|\!)}\downarrow & & \downarrow{\scriptstyle id + (\!|g|\!)} \\
B & \xleftarrow[g]{} & 1 + B
\end{array}
$$

B Regra prática para a recursividade mútua em $\mathbb{N}_0$

Nesta disciplina estudou-se como fazer programação dinâmica por cálculo, recorrendo à lei de recursividade mútua.[8]

Para o caso de funções sobre os números naturais ($\mathbb{N}_0$, com functor F $X = 1 + X$) é fácil derivar-se da lei que foi estudada uma *regra de algibeira* que se pode ensinar a programadores que não tenham estudado Cálculo de Programas. Apresenta-se de seguida essa regra, tomando como exemplo o cálculo do ciclo-for que implementa a função de Fibonacci, recordar o sistema:

$$
\begin{array}{l}
fib\ 0 = 1 \\
fib\ (n+1) = f\ n
\end{array}
$$

[7] Exemplos tirados de [**?**].
[8] Lei (3.95) em [**?**], página 112.

$$
\begin{aligned}
&f\ 0 = 1\\
&f\ (n+1) = fib\ n + f\ n
\end{aligned}
$$

Obter-se-á de imediato

$$
\begin{aligned}
&fib' = \pi_1 \cdot \text{for}\ loop\ init\ \textbf{where}\\
&\quad loop\ (fib, f) = (f, fib + f)\\
&\quad init = (1,1)
\end{aligned}
$$

usando as regras seguintes:

- O corpo do ciclo *loop* terá tantos argumentos quanto o número de funções mutuamente recursivas.
- Para as variáveis escolhem-se os próprios nomes das funções, pela ordem que se achar conveniente.[9]
- Para os resultados vão-se buscar as expressões respectivas, retirando a variável n.
- Em *init* coleccionam-se os resultados dos casos de base das funções, pela mesma ordem.

Mais um exemplo, envolvendo polinómios do segundo grau $ax^2 + bx + c$ em $\mathbb{N}_0$. Seguindo o método estudado nas aulas[10], de $f\ x = ax^2 + bx + c$ derivam-se duas funções mutuamente recursivas:

$$
\begin{aligned}
&f\ 0 = c\\
&f\ (n+1) = f\ n + k\ n\\
&k\ 0 = a + b\\
&k\ (n+1) = k\ n + 2\ a
\end{aligned}
$$

Seguindo a regra acima, calcula-se de imediato a seguinte implementação, em Haskell:

$$
\begin{aligned}
&f'\ a\ b\ c = \pi_1 \cdot \text{for}\ loop\ init\ \textbf{where}\\
&\quad loop\ (f, k) = (f + k, k + 2 * a)\\
&\quad init = (c, a + b)
\end{aligned}
$$

C O mónade das distribuições probabilísticas

Mónades são functores com propriedades adicionais que nos permitem obter efeitos especiais em programação. Por exemplo, a biblioteca Probability oferece um mónade para abordar problemas de probabilidades. Nesta biblioteca, o conceito de distribuição estatística é captado pelo tipo

$$
\textbf{newtype}\ \text{Dist}\ a = D\ \{unD :: [(a, ProbRep)]\} \tag{1}
$$

em que *ProbRep* é um real de 0 a 1, equivalente a uma escala de 0 a 100%.

Cada par (a,p) numa distribuição $d ::$ Dist a indica que a probabilidade de a é p, devendo ser garantida a propriedade de que todas as probabilidades de d somam 100%. Por exemplo, a seguinte distribuição de classificações por escalões de A a E,

será representada pela distribuição

$$
\begin{aligned}
&d_1 :: \text{Dist}\ Char\\
&d_1 = D\ [(\texttt{'A'}, 0.02), (\texttt{'B'}, 0.12), (\texttt{'C'}, 0.29), (\texttt{'D'}, 0.35), (\texttt{'E'}, 0.22)]
\end{aligned}
$$

que o GHCi mostrará assim:

[9] Podem obviamente usar-se outros símbolos, mas numa primeira leitura dá jeito usarem-se tais nomes.
[10] Secção 3.17 de [**?**] e tópico Recursividade mútua nos vídeos de apoio às aulas teóricas.

```
'D'  35.0%
'C'  29.0%
'E'  22.0%
'B'  12.0%
'A'   2.0%
```

É possível definir geradores de distribuições, por exemplo distribuições *uniformes*,

$$d_2 = uniform\ (words\ \texttt{"Uma frase de cinco palavras"})$$

isto é

```
    "Uma"  20.0%
  "cinco"  20.0%
     "de"  20.0%
  "frase"  20.0%
"palavras"  20.0%
```

distribuição *normais*, eg.

$$d_3 = normal\ [10..20]$$

etc.[11] Dist forma um **mónade** cuja unidade é $return\ a = D\ [(a,1)]$ e cuja composição de Kleisli é (simplificando a notação)

$$(f \bullet g)\ a = [(y, q*p) \mid (x,p) \leftarrow g\ a, (y,q) \leftarrow f\ x]$$

em que $g : A \to$ Dist B e $f : B \to$ Dist C são funções **monádicas** que representam *computações probabilísticas*.

Este mónade é adequado à resolução de problemas de *probabilidades e estatística* usando programação funcional, de forma elegante e como caso particular da programação monádica.

D Código fornecido

Problema 1

Alguns testes para se validar a solução encontrada:

$$test\ a\ b\ c = \text{map}\ \ (fbl\ a\ b\ c)\ x \equiv \text{map}\ \ (f\ a\ b\ c)\ x\ \textbf{where}\ x = [1..20]$$
$$test1 = test\ 1\ 2\ 3$$
$$test2 = test\ (-2)\ 1\ 5$$

Problema 2

Verificação: a árvore de tipo Exp gerada por

$$acm\_tree = tax\ acm\_ccs$$

deverá verificar as propriedades seguintes:

- $expDepth\ acm\_tree \equiv 7$ (profundidade da árvore);
- $length\ (expOps\ acm\_tree) \equiv 432$ (número de nós da árvore);
- $length\ (expLeaves\ acm\_tree) \equiv 1682$ (número de folhas da árvore).[12]

O resultado final

$$acm\_xls = post\ acm\_tree$$

não deverá ter tamanho inferior ao total de nodos e folhas da árvore.

[11] Para mais detalhes ver o código fonte de Probability, que é uma adaptação da biblioteca PHP ("Probabilistic Functional Programming"). Para quem quiser saber mais recomenda-se a leitura do artigo [**?**].

[12] Quer dizer, o número total de nodos e folhas é 2114, o número de linhas do texto dado.

Problema 3

Função para visualização em SVG:

$$drawSq\ x = picd''\ [Svg.scale\ 0.44\ (0,0)\ (x \gg\!= sq2svg)]$$
$$sq2svg\ (p,l) = (color\ \texttt{"\#67AB9F"} \cdot polyg)\ [p, p .+ (0,l), p .+ (l,l), p .+ (l,0)]$$

Para efeitos de temporização:

$$await = threadDelay\ 1000000$$

Problema 4

Rankings:

```
rankings = [
  ("Argentina",4.8),
  ("Australia",4.0),
  ("Belgium",5.0),
  ("Brazil",5.0),
  ("Cameroon",4.0),
  ("Canada",4.0),
  ("Costa Rica",4.1),
  ("Croatia",4.4),
  ("Denmark",4.5),
  ("Ecuador",4.0),
  ("England",4.7),
  ("France",4.8),
  ("Germany",4.5),
  ("Ghana",3.8),
  ("Iran",4.2),
  ("Japan",4.2),
  ("Korea Republic",4.2),
  ("Mexico",4.5),
  ("Morocco",4.2),
  ("Netherlands",4.6),
  ("Poland",4.2),
  ("Portugal",4.6),
  ("Qatar",3.9),
  ("Saudi Arabia",3.9),
  ("Senegal",4.3),
  ("Serbia",4.2),
  ("Spain",4.7),
  ("Switzerland",4.4),
  ("Tunisia",4.1),
  ("USA",4.4),
  ("Uruguay",4.5),
  ("Wales",4.3)]
```

Geração dos jogos da fase de grupos:

$$generateMatches = pairup$$

Preparação da árvore do "mata-mata":

$$arrangement = (\gg\!= swapTeams) \cdot chunksOf\ 4\ \textbf{where}$$
$$swapTeams\ [[a_1,a_2],[b_1,b_2],[c_1,c_2],[d_1,d_2]] = [a_1,b_2,c_1,d_2,b_1,a_2,d_1,c_2]$$

Função proposta para se obter o *ranking* de cada equipa:

$$rank\ x = 4 ** (pap\ rankings\ x - 3.8)$$

Critério para a simulação não probabilística dos jogos da fase de grupos:

$$\begin{aligned}
&gsCriteria = s \cdot \langle id \times id, rank \times rank \rangle \textbf{ where} \\
&\quad s\ ((s_1,s_2),(r_1,r_2)) = \textbf{let } d = r_1 - r_2 \textbf{ in} \\
&\qquad \textbf{if } d > 0.5 \textbf{ then } Just\ s_1 \\
&\qquad\quad \textbf{else if } d < -0.5 \textbf{ then } Just\ s_2 \\
&\qquad\qquad \textbf{else } Nothing
\end{aligned}$$

Critério para a simulação não probabilística dos jogos do mata-mata:

$$\begin{aligned}
&koCriteria = s \cdot \langle id \times id, rank \times rank \rangle \textbf{ where} \\
&\quad s\ ((s_1,s_2),(r_1,r_2)) = \textbf{let } d = r_1 - r_2 \textbf{ in} \\
&\qquad \textbf{if } d \equiv 0 \textbf{ then } s_1 \\
&\qquad\quad \textbf{else if } d > 0 \textbf{ then } s_1 \textbf{ else } s_2
\end{aligned}$$

Critério para a simulação probabilística dos jogos da fase de grupos:

$$\begin{aligned}
&pgsCriteria = s \cdot \langle id \times id, rank \times rank \rangle \textbf{ where} \\
&\quad s\ ((s_1,s_2),(r_1,r_2)) = \\
&\qquad \textbf{if } abs\ (r_1 - r_2) > 0.5 \textbf{ then } \mathsf{fmap}\ Just\ (pkoCriteria\ (s_1,s_2)) \textbf{ else } f\ (s_1,s_2) \\
&\quad f = D \cdot ((Nothing, 0.5) :) \cdot \mathsf{map}\ \ (Just \times (/2)) \cdot unD \cdot pkoCriteria
\end{aligned}$$

Critério para a simulação probabilística dos jogos do mata-mata:

$$\begin{aligned}
&pkoCriteria\ (e_1,e_2) = D\ [(e_1, 1 - r_2\ /\ (r_1 + r_2)), (e_2, 1 - r_1\ /\ (r_1 + r_2))] \textbf{ where} \\
&\quad r_1 = rank\ e_1 \\
&\quad r_2 = rank\ e_2
\end{aligned}$$

Versão probabilística da simulação da fase de grupos:[13]

$$\begin{aligned}
&psimulateGroupStage = trim \cdot \mathsf{map}\ \ (pgroupWinners\ pgsCriteria) \\
&trim = top\ 5 \cdot sequence \cdot \mathsf{map}\ \ (filterP \cdot norm) \textbf{ where} \\
&\quad filterP\ (D\ x) = D\ [(a,p) \mid (a,p) \leftarrow x, p > 0.0001] \\
&\quad top\ n = vec2Dist \cdot take\ n \cdot reverse \cdot presort\ \pi_2 \cdot unD \\
&\quad vec2Dist\ x = D\ [(a, n\ /\ t) \mid (a,n) \leftarrow x] \textbf{ where } t = sum\ (\mathsf{map}\ \ \pi_2\ x)
\end{aligned}$$

Versão mais eficiente da *pwinner* dada no texto principal, para diminuir o tempo de cada simulação:

$$\begin{aligned}
&pwinner :: \mathsf{Dist}\ Team \\
&pwinner = mbin\ f\ x \gg\!= pknockoutStage \textbf{ where} \\
&\quad f\ (x,y) = initKnockoutStage\ (x +\!\!+ y) \\
&\quad x = \langle g \cdot take\ 4, g \cdot drop\ 4 \rangle\ groups \\
&\quad g = psimulateGroupStage \cdot genGroupStageMatches
\end{aligned}$$

Auxiliares:

$$\begin{aligned}
&best\ n = \mathsf{map}\ \ \pi_1 \cdot take\ n \cdot reverse \cdot presort\ \pi_2 \\
&consolidate :: (Num\ d, Eq\ d, Eq\ b) \Rightarrow [(b,d)] \to [(b,d)] \\
&consolidate = \mathsf{map}\ \ (id \times sum) \cdot collect \\
&collect :: (Eq\ a, Eq\ b) \Rightarrow [(a,b)] \to [(a,[b])] \\
&collect\ x = nub\ [k \mapsto [d' \mid (k',d') \leftarrow x, k' \equiv k] \mid (k,d) \leftarrow x]
\end{aligned}$$

Função binária monádica f:

$$\begin{aligned}
&mmbin :: Monad\ m \Rightarrow ((a,b) \to m\ c) \to (m\ a, m\ b) \to m\ c \\
&mmbin\ f\ (a,b) = \textbf{do } \{x \leftarrow a; y \leftarrow b; f\ (x,y)\}
\end{aligned}$$

Monadificação de uma função binária f:

$$\begin{aligned}
&mbin :: Monad\ m \Rightarrow ((a,b) \to c) \to (m\ a, m\ b) \to m\ c \\
&mbin = mmbin \cdot (return \cdot)
\end{aligned}$$

Outras funções que podem ser úteis:

$$\begin{aligned}
&(f\ `is`\ v)\ x = (f\ x) \equiv v \\
&rcons\ (x,a) = x +\!\!+ [a]
\end{aligned}$$

[13] Faz-se "trimming" das distribuições para reduzir o tempo de simulação.

E Soluções dos alunos

Os alunos devem colocar neste anexo as suas soluções para os exercícios propostos, de acordo com o "layout" que se fornece. Não podem ser alterados os nomes ou tipos das funções dadas, mas pode ser adicionado texto, diagramas e/ou outras funções auxiliares que sejam necessárias.

Valoriza-se a escrita de *pouco* código que corresponda a soluções simples e elegantes.

Problema 1

Inicialmente temos a funcao f.

$$
\begin{aligned}
&f :: Int \to Int \to Int \to Int \to Int\\
&f\ a\ b\ c\ 0 = 0\\
&f\ a\ b\ c\ 1 = 1\\
&f\ a\ b\ c\ 2 = 1\\
&f\ a\ b\ c\ n = a * (f\ a\ b\ c\ (n-1)) + b * (f\ a\ b\ c\ (n-2)) + c * (f\ a\ b\ c\ (n-3))
\end{aligned}
$$

Esta pode ser dividida em tres funcoes diferentes, todas em ordem a n.

$$
\begin{aligned}
&g :: Int \to Int \to Int \to Int \to Int\\
&g\ a\ b\ c\ 0 = 1\\
&g\ a\ b\ c\ n = a * (g\ a\ b\ c\ (n-1)) + b * (h\ a\ b\ c\ (n-1)) + c * (i\ a\ b\ c\ (n-1))\\
&h :: Int \to Int \to Int \to Int \to Int\\
&h\ a\ b\ c\ 0 = 1\\
&h\ a\ b\ c\ n = (g\ a\ b\ c\ (n-1))\\
&i :: Int \to Int \to Int \to Int \to Int\\
&i\ a\ b\ c\ 0 = 0\\
&i\ a\ b\ c\ n = (h\ a\ b\ c\ (n-1))
\end{aligned}
$$

Apos as 3 estarem definidas podemos fazer a funcao fbl, que chama ordenadamente as funcoes num loop. Esta funcao recebe 3 inteiros, sendo eles os 3 primeiros elementos da lista que serao modificados segundo a logica das funcoes desenvolvidas anteriormente, armazenando o resultado para o n atual no segundo elemento do tupulo, o elemento seguinte no segundo elemento do primeiro elemento do tupulo e, finalmente, o n+2 no primeiro elemento do primeiro elemento do tupulo. Isto permite que os elementos da lista sejam calculados em tempo linear gracas ao recurso a um loop.

$$
\begin{aligned}
&fbl\ a\ b\ c = \pi_2 \cdot \text{for}\ loop\ a\ b\ c\ initial\ \textbf{where}\\
&\quad loop\ a\ b\ c\ ((g,h),i) = ((a*g + b*h + c*i, g), h)\\
&\quad initial = ((1,1),0)
\end{aligned}
$$

Funções auxiliares pedidas:

$$
\begin{aligned}
&loop\ a\ b\ c\ ((g,h),i) = ((a*g + b*h + c*i, g), h)\\
&initial = ((1,1),0)\\
&wrap = \pi_2
\end{aligned}
$$

Problema 2

O gene vai ser representado por (id -$-id \times gene2) \cdot out$, *onde gene2 será a fun*ção *que gera uma lista de sub* $-$ *listas dividida pela sub* $-$ *cabe*ças *da mesma, de modo a entregar uma* $[[String]]$ *para executar* map *tax de forma conveniente* $\cdot$ *Antes de entregar a* $[[String]]$ *a* map *tax, esta fun*ção *retira ainda* 4 *espa*ços *a cada elemento para que a nota*ção *seja correta em ca*
*Gene de*tax:λ*begin* $\{code\}$ *espacos* $::$ *String* $\to$ *Int espacos* $[\,] = 0$ *espacos* $(a:t)$ a == ' ' = 1 + espacos(t)
otherwise $= 0$ *removeEspacoAux* $::$ *String* $\to$ *Int* $\to$ *String removeEspacoAux a* 4 $=$ *a removeEspacoAux* $(x:t)$ *n* x == '
' = removeEspacoAux t (n+1) *otherwise* $= (x:t)$ *addLast* $::$ $[[String]] \to String \to [[String]]$ *addLast* $(x:xs)$ *b* xs ==
[] = [x ++ [b]] *otherwise* $=$ *addLast xs b* -- caso tenha 4 espaços no inicio vai ser uma sub-cabeça então eu crio uma n
espacos b == 4 = auxGene t (c ++ [[(removeEspacoAux b 1)]]) *otherwise* $=$ *auxGene t* $(addLast\ c\ (removeEspacoAux\ b\ 1))$ *auxG*
c gene2 $::$ $[String] \to [[String]]$ *gene2* $(a:t) =$ *auxGene t* $[[\,]]$ *gene2* $[\,] = [[\,]]$ *gene* $= (id$-- id ¿¡ gene2).out

Para esta função vamos usar duas funções auxiliares dadas na biblioteca exp, deco que em cada nodo da árvore adiciona o nivel da árvore onde nos encontramos, e a função lnks que cria pares de todos os elementos pai e filho da árvore. Por fim só precisamos de verificar os pares que têm pai e filho iguais e criar a lista com o devido resultado. Função de pós-processamento:

$$post = aux1 \cdot aux \cdot lnks \cdot deco$$

```
-- aux é uma função que retorna os pares ordenados para serem posteriormente escritos
aux :: [((Int, String), (Int, String))] → [(String, String)]
aux [] = []
aux ((a, b) : t)
  | π1 (a) ≡ (π1 (b) + 1) = ((π2 (a), π2 (b)) : (aux t))
  | otherwise = aux t
aux1 :: [(String, String)] → [[String]]
aux1 a = aux2 a [[]]
aux2 :: [(String, String)] → [[String]] → [[String]]
aux2 [(a, b)] _ = [[a, b]]
aux2 ((a, b) : t) ((h : j))
  | a ≡ last (h) = aux2 t ((h ++ [a]) : j)
  | otherwise = aux2 t ((h : j) ++ [[a]])
aux2 [] _ = []
```

Problema 3

Neste problema, inicialmente foram definidos os tipos de dados necessários para a resolução do problema, sendo eles:

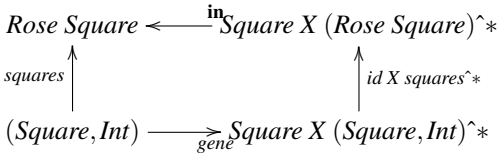
```
type Square = (Int, Int)
type Side = Double
type Point = (Double, Double)
```

Em seguida, foram definidas as funções necessárias para a resolução do problema, sendo elas:

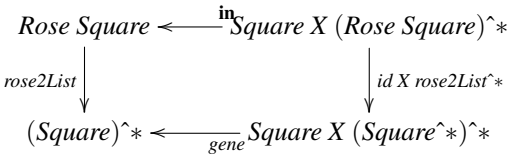
```
squares :: (Square, Int) → Rose Square
rose2List :: Rose a → [a]
sierpinski :: (Square, Int) → [Square]
sierpinski = ⟦gr2l, gsq⟧R
```

Posteriormente, fotam feitos os diagramas das funções pedidas:

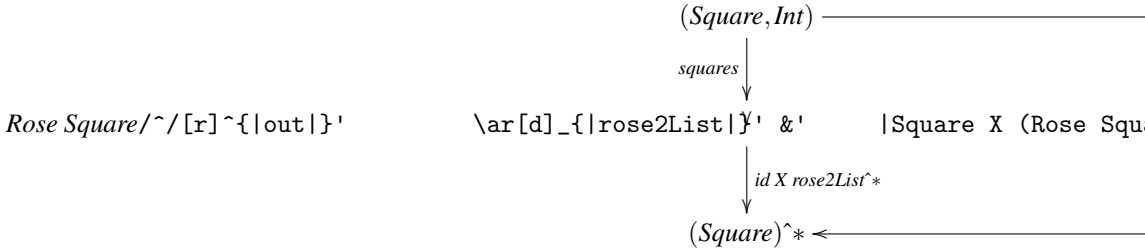
Anamorfismo Squares:



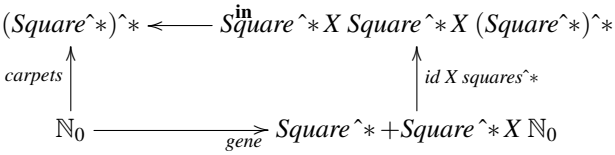
Catamorfismo RoseToList:



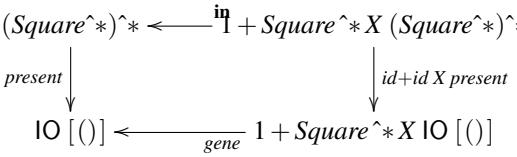
Hilomorfismo:



Carpets:



Present:



Em seguida, foram definidas as funções pedidas:

```
sierpinski :: (Square, Int) → [Square]
sierpinski (((x,y),t),p) = ⟦gr2l, gsq⟧R (((x+t,y+t),t),p)

geneCarpets :: Int → [Square] + ([Square], Int)
geneCarpets 0 = i1 (sierpinski (((0,0),32),0))
geneCarpets p = i2 (sierpinski (((0,0),32),p−1),p−1)

carpets :: Int → [[Square]]
carpets = ⟦geneCarpets⟧

genePresent :: () + ([Square], IO [()]) → IO [()]
genePresent g = case g of
  i1 () → return []
  i2 (s,io) → do x ← io
     drawSq s
     await
     return (() : x)

present :: [[Square]] → IO [()]
present = ⦇genePresent⦈

constructSierp :: Int → IO [()]
constructSierp = present · carpets

gr2l :: (a, [[a]]) → [a]
gr2l (s, []) = [s]
gr2l (s, l) = s : concat l

gsq :: (Square, Int) → (Square, [(Square, Int)])
gsq (s,0) = (s, [])
gsq (s,p) = (s, map (λx → (x,p−1)) (carpet s))

carpet :: Square → [Square]
carpet (p,l) = [(p .+ (−2∗l/3,l/3),l/3),(p .+ (−2∗l/3,4∗l/3),l/3),
  (p .+ (−2∗l/3,−2∗l/3),l/3),(p .+ (l/3,4∗l/3),l/3),
  (p .+ (l/3,−2∗l/3),l/3),(p .+ (4∗l/3,4∗l/3),l/3),
  (p .+ (4∗l/3,l/3),l/3),(p .+ (4∗l/3,     −2∗l/3),l/3)]

rose2List :: Rose Square → [Square]
rose2List = ⦇gr2l⦈R
```

Para a construção do algoritmo de Sierpinski foi necessário definir um hilomorfismo para gerar uma lista de quadrados a partir de um quadrado e um inteiro. Primeiramente o gene gsq obtem os sub-quadrados de um quadrado inicial e de um dado nível de recursão. O gene gr2l concatena a lista de quadrados obtida pelo gene gsq com o quadrado inicial. Sendo assim a função sierpinski cria uma lista de quadrados dum dado nível. Por fim, a função constructSierp cria os carpetes até um dado nível de recursão e desenha os quadrados, recorrendo ás funções carpets e present.

Problema 4

Versão não probabilística

Para a versão não probabilistica deste problema foram propostos quatro problemas, sendo o primeiro destes a definicão do gene do catamorfismo que caracteriza a funcão consolidate'. O 'cgene' e responsavel por dada uma lista de tupulos juntar todos os que comecem com elementos iguais somando o segundo elemento de cada tupulo. Esta ultima caracteristica surge de aplicar a funcao collect' a lista de tupulos.

Gene de $consolidate'$:

$$
\begin{aligned}
&cgene\ (i_1\ ()) = [\,]\\
&cgene\ (i_2\ (h,t)) = \mathsf{map}\ \ (id \times sum)\ (collect'\ c)\\
&\quad \textbf{where}\ c = h : t\\
&collect' :: (Eq\ a, Num\ b) \Rightarrow [(a,b)] \to [(a,[b])]\\
&collect'\ [\,] = [\,]\\
&collect'\ l@((x,y):xs) = (x, (\mathsf{map}\ \ \pi_2\ (filter\ (\lambda(a,b) \to a \equiv x)\ l))) : collect'\ (filter\ (\lambda(a,b) \to a \not\equiv x)\ xs)
\end{aligned}
$$

A segunda parte do problema consiste na implementação da função *matchResult* que, como o nome indica, dada uma função de comparação de equipas e um jogo, devolve uma lista de tuplos com o nome da equipa e o número de pontos que esta obteve no jogo. Esta funcao foi implementada tendo em conta que o resultado da funcao de comparacao ser Monádico, isto é, nem sempre existe um resultado para um dado jogo, podendo nao ser retornada nenhuma das equipas como vencedora. Dado tudo isto a funcao para alem de atribuir os pontos segundo a equipa vencedora, verifica a possibilidade do resultado ser *Nothing*, atribuindo nesses casos um ponto a cada uma delas.

$$
\begin{aligned}
&matchResult\ f\ (x,y)\\
&\quad \mid w \equiv Nothing = [(x,1),(y,1)]\\
&\quad \mid w \equiv Just\ x = [(x,3),(y,0)]\\
&\quad \mid w \equiv Just\ y = [(x,0),(y,3)]\\
&\quad \textbf{where}\ w = f\ (x,y)
\end{aligned}
$$

De seguida o desafio foi a implementação da funcao responsavel pela gera dos jogos da fase de grupos. Para tal foi definida a funcao pairup que gera todas as possiveis combinações de equipas num dado grupo. Com esta função torna-se possivel gerar todos os grupos recorrendo a funcao *genGroupStageMatches*.

$$
\begin{aligned}
&pairup\ [\,] = [\,]\\
&pairup\ [x] = [\,]\\
&pairup\ (x:xs) = (\mathsf{map}\ \ (\lambda a \to (x,a))\ xs) +\!\!+ pairup\ xs
\end{aligned}
$$

Para finalizar a versao nao probabilistica do problema 4 falta ainda a definicão do gene do anamorfismo que atua sobre o resultado da funcao *arrangement* de modo a dar inicio a fase de grupos. A funcao arrangement recebe como argumento uma lista de listas com as duas equipas vencedoras de cada grupo, retornando uma lista de equipas. A lista fornecida esta ordenada de modo a que cada dois elementos formem um dos jogos e, na vista de arvore de jogos que levariam a uma final, a divisao recursiva da lista em duas listas do mesmo tamanho seja correspondente a uma das fases eliminatórias. Isto significa que, tendo a lista original $[a,b,c,d]$:

- a primeira metade da lista corresponde a uma das semi-finais $[a,b]$
- a segunda metade da lista corresponde a outra das semi-finais $[c,d]$
- a lista resultante seria $[[a,b],[c,d]]$ e esta corresponderia a uma final

Esta logica e o objetivo que se pretende implementar com o catamorfismo de gene glt. O gene glt e entao responsavel pela aplicação destal logica de modo a criar a "arvore"de jogos que levaria a uma final. Caso a lista de equipas seja de tamanho 1 significa que chegamos a uma das equipas que ira disputar um dado jogo, sendo assim devolivido o proprio elemento, na vista de arvore o equivalente a uma folha, neste caso especifico e utilizado um $\cdot + \cdot$, sendo entao $i_1\ equipa$. Caso contrario,

a lista de equipas e dividida em duas partes que, no futuro, disputarao um dado jogo. Ao contrario do caso singular, o resultado da divisao e devolvido como $i_2\ (lista1, lista2)$, ou, numa vista de arvore, $Fork\ (lista1, lista2)$, representando entao o jogo. Olhando para o exemplo anterior, a lista $[a,b,c,d]$ e dividida em duas partes, $[a,b]$ e $[c,d]$, que representam os dois jogos que levariam a uma final. Na representacao em Haskell, a mesma seria representada como $i_2\ (i_2\ (i_1\ a, i_1\ b), i_2\ (i_1\ c, i_1\ d))$.

$$
\begin{aligned}
&glt\ [a] = i_1\ a\\
&glt\ l = i_2\ (halfList\ l)\\
&halfList :: [a] \to ([a],[a])\\
&halfList\ l = ((take\ h\ l),(drop\ h\ l))\\
&\quad \textbf{where}\ h = (length\ l)\ `div`\ 2
\end{aligned}
$$

Versão probabilística

A versão probabilística do problema 4 consiste na implementação de uma estrategia probabilistica para a previsao da probabilidade de cada uma das equipas conseguir ganhar o Mundial de Seleções. Para que o mesmo seja possivel e necessario a adaptação da versao nao probabilistica do problema 4 de modo a suportar a utilização de probabilidades. Estas serao representadas por uma distribuição de probabilidades representada pelo Monad Dist referido no anexo C.

A primeira alteracao logica sobre o problema foi o calculo dos resultados de cada jogo. Estes deixaram de ser calculados de forma deterministica e, a cima de tudo, deixaram de fornecer um resultado unico, passando a fornecer uma distribuição de probabilidades. Esta distribuição pode, para alem de conter a probabilidade de cada equipa ganhar o jogo, conter a probabilidade de ocorrer um empate. Para tal foi definida a funcao *pmatchResult* que recebe como argumento uma funcao que calcula a probabilidade de uma equipa ganhar um jogo e um jogo, retornando uma distribuição de probabilidades com as probabilidades de cada equipa ganhar o jogo ou existir empate. Apos aplicar a funcao de probabilidades ao jogo verifica-se a existencia de uma probabilidade de existir um empate. Finalmente substitui-se cada uma das equipas na distribuição de probabilidades pelos pontos que seriam atribuidos a cada equipa caso esse fosse o resultado.

Por exemplo, para a distribuição de probabilidades, $D\ [(Nothing, 0.3), (\texttt{'A'}, 0.6), (\texttt{'B'}, 0.1)]$, deveriamos fazer as seguintes substituições:

- *Nothing* por $((\texttt{'A'},1),(\texttt{'B'},1))$
- `'A'` por $((\texttt{'A'},3),(\texttt{'B'},0))$
- `'B'` por $((\texttt{'A'},1),(\texttt{'B'},3))$

O resultado final seria, portanto, a distribuição de probabilidades $D\ [((\texttt{'A'},1),(\texttt{'B'},1),0.3),((\texttt{'A'},3),(\texttt{'B'},0),0.6),((\texttt{'A}$

$$
\begin{aligned}
&pmatchResult\ f\ (x,y)\\
&\quad \mid \pi_1\ (di\ !!\ 0) \equiv Nothing = D\ [(emp, \pi_2\ (di\ !!\ 0)),(vx, \pi_2\ (di\ !!\ 1)),(vy, \pi_2\ (di\ !!\ 2))]\\
&\quad \mid otherwise = D\ [(vx, \pi_2\ (di\ !!\ 0)),(vy, \pi_2\ (di\ !!\ 1))]\\
&\quad \textbf{where}\ D\ di = f\ (x,y)\\
&\qquad emp = ((x,1),(y,1))\\
&\qquad vx = ((x,3),(y,0))\\
&\qquad vy = ((x,0),(y,3))
\end{aligned}
$$

O proximo passo foi a implementação da função *pgroupWinners* que recebe como argumento uma função que calcula a probabilidade de duas dadas equipas ficarem em primeiro e segundo lugar, respectivamente, de um dado grupo. Esta função recebe como argumento uma lista de jogos, calculando entao, para cada um deles, a probabilidade de cada equipa ganhar 'n' pontos no final do mesmo, dependo do resultado. Apos a obtenção das varias distribuições de probabilidades devemos fazer uma combinação das mesma com todos os possiveis resultados,i.e., todas as combinações possiveis de resultados de cada um dos jogos. Para tal, foi utilizada a função *sequenceA* que recebe como argumento uma lista de distribuições de probabilidades e retorna uma distribuição de probabilidades com todas as combinações possiveis dos resultados de cada uma das distribuições. A distribuição de todos os possiveis resultados e entao obtida da seguinte forma:

```
distMatches = sequenceA $ map (pmatchResult pgsCriteria)
```

Com a distribuição de todos os possiveis resultados, devemos agora, para cada um deles, calcular a equipa que ficaria em primeiro e segundo lugar. Assim, o primeiro passo foi alterar a estrutura do interior da distribuição de probabilidades, passando a ter a lista de tupulos $(equipa, pontos)$. Estes sao entao "consolidados", obtendo as pontuacoes finais de cada equipa para esse "set"de resultados. Uma vez que apenas queremos as duas equipas com maior pontuacao, utilizamos a função $best\ 2$ para obter as duas melhores equipas. Finalmente, para cada um dos resultados, obtemos a equipa que ficaria em primeiro e segundo lugar, e a probabilidade de esse resultado ocorrer, voltando a obter a distribuição de probabilidades desejada.

```
pgroupWinners :: (Match -> Dist (Maybe Team)) -> [Match] -> Dist [Team]
pgroupWinners pgsCriteria m = D (map (λ(a,b) -> ((best 2 $ consolidate $ concat $ map (λ(c,d) -> [c,d]) a),b)) x)
  where D x = sequenceA $ map (pmatchResult pgsCriteria) m
```

Finalmente, para preparar a fase eliminatória podemos recorrer a funcao *initKnockoutStage* definida anteriormente, devendo apenas ter em conta a monadificação da mesma.

```
pinitKnockoutStage d = return $ (initKnockoutStage d)
```

Para testar o programa desenvolvido, devemos entao compilar e testar a função 'pwinner' de modo a obter as probabilidades de cada uma das equipas ganhar o Mundial de Futebol. No nosso caso, quando corremos o programa a solução obtida e a seguinte:

```
    "Belgium" 16.3 %
     "Brazil" 15.3 %
     "France"  9.4 %
"Argentina"    7.6 %
      "Spain" 7.6 %
    "England"  7.5 %
"Netherlands"  5.2 %
    "Denmark"  4.8 %
  "Portugal"   4.7 %
     "Mexico"  4.3 %
    "Uruguay"  4.0 %
    "Germany"  2.9 %
    "Senegal"  2.5 %
"Switzerland"  2.3 %
      "Wales" 1.7 %
        "USA" 1.1 %
    "Morocco"  1.0 %
    "Croatia"  0.6 %
     "Poland"  0.4 %
     "Serbia"  0.3 %
      "Japan" 0.2 %
```

Esta, apesar de corretamente calculada nao e muito justa a realidade dos eventos, concluindo assim que o ranking nem sempre pode ser adequado, devendo quando possivel utilizar metricas diversar para obter uma visao mais realista das chances de cada equipa.

Valorização

Apos termos todas as funcoes definidas, decidimos exprimentar outras metricas de ranking, a fim de verificar o funcionamento do programa com diferentes parametros. Para tal, adicionamos o seguinte ranking, onde as equipas estao associadas ao numero medio de golos em cada partida. Esta metrica seria, potencialmente, mais realista, uma vez que tem um maior impacto no resultado do jogo.

```
rankingGolos = [
  ("Argentina",2.14),
```

```
("Australia",1.0),
("Belgium",1.29),
("Brazil",2.01),
("Cameroon",1.3),
("Canada",0.6),
("Costa Rica",1.0),
("Croatia",1.28),
("Denmark",1.11),
("Ecuador",1.33),
("England",1.29),
("France",1.7),
("Germany",1.47),
("Ghana",1.6),
("Iran",0.8),
("Japan",1.40),
("Korea Republic",1.2),
("Mexico",0.66),
("Morocco",1.7),
("Netherlands",1.63),
("Poland",0.64),
("Portugal",1.40),
("Qatar",0.2),
("Saudi Arabia",1),
("Senegal",1.15),
("Serbia",1.57),
("Spain",2.25),
("Switzerland",0.62),
("Tunisia",0.33),
("USA",0.66),
("Uruguay",1.12),
("Wales",1.67)]
```

Para testar entao a nova medida de ranking foi necessario alterar a funcao que cuida do mesmo. I.e., alteramos a funcao 'rank' de modo a obter o ranking baseado na media de golos por jogo de cada equipa..5pc Ainda alem da lista de ranks, foram alterados os parametros que geravam o rank final, sendo o valor minimo reduzido para '0.2' que corresponde ao "Qatar"e o impacto reduzido de 4 para 2. Esta medida foi tomada dada a imprevisibilidade dos jogos, sendo a diferenca de golos, apesar de significativa, algo que nao caracteriza uma equipa ao mesmo nivel do que o seu ranking da FIFA, por exemplo.

code rank x = 2 ** (pap rankingsGolos x - 0.2)

Ao correr a nova versao, os resultados obtidos foram os seguintes:

```
  "Argentina"  11.0%
      "Spain" 10.8%
     "France"  7.6%
    "Morocco"  7.1%
      "Wales"  7.0%
"Netherlands"  7.0%
     "Brazil"  6.8%
   "Portugal"  6.5%
      "Japan"  5.3%
     "Serbia"  4.8%
    "Croatia"  4.1%
"Saudi Arabia"  3.7%
    "Uruguay"  3.5%
    "Denmark"  3.3%
    "Ecuador"  3.2%
    "England"  1.9%
```

"Senegal"	1.7 %
"USA"	1.7 %
"Cameroon"	1.1 %
"Korea Republic"	0.7 %
"Australia"	0.6 %
"Switzerland"	0.5 %

Que, comparados a versao anterior, e dado ja ter sido realizado o Mundial de Futebol, uma representacao mais proxima a realidade, apesar de ainda assim ter alguns ”outliars”.

Índice