Teoria das Categorias Flechas e Mónades em Java

Mozart L. Siqueira[1], Pablo M. Parada[2]
Ciência da Computação
Centro Universitário La Salle - Unilasalle
E-mail: mozarts@unilasalle.edu.br[1], pablo.paradabol@gmail.com[2]

***Resumo*—<escrever>**
***Index Terms*—<escrever>**

I. INTRODUÇÃO

<escrever>

II. PARADIGMA FUNCIONAL

Influenciado principalmente pelo desenvolvimento do lambda calculus [1], compondo o grupo da programação declarativa, o paradigma funcional utiliza-se da ideia de expressar computações através de funções combinadas em expressões. Neste, funções determinam o que deverá ser computado, ao invés de como sera computado [2]. Programas são construídos através da composição, tal que funções triviais (building blocks) são combinadas dando origem a novas funções que descrevem computações mais complexas.

Building blocks não devem fazer uso de variáveis que dependam de estado, isso significa que a computação deve ser pura e sem efeitos colaterais (side-effects). Também destaca-se o princípio de imutabilidade, onde o valor de uma variável é determinado em sua criação, não permitindo novas atribuições posteriormente. Ao expressar um programa em uma linguagem funcional obtém-se uma maneira concisa de solucionar problemas, sendo que este constitui-se de operações e objetos atômicos e regras gerais para sua composição [3].

Tais propriedades são apreciadas nos tempos atuais, visto que a Lei de Moore nos fornece cada vez mais núcleos, não necessariamente núcleos mais rápidos [4]. Em programas não-determinísticos, múltiplas threads podem alterar os dados representados por objetos imutáveis sem ocasionar os diversos problemas já conhecidos como dead locks e race conditions. Além do thread safety oferecido pela propriedade de imutabilidade, há também o conceito de funções transparentes referencialmente, ou seja, funções que não utilizam variáveis de estado. Em ambientes distribuídos onde a execução é subdivida em diferentes threads, a transparência referencial garante sempre o mesmo retorno, dados os mesmos argumentos.

O paradigma funcional expressa programas através de composições mantendo a imutabilidade e a transparência referencial, portanto apresenta caraterísticas importantes para os tempos atuais. A complexidade intrínseca à ambientes distribuídos é reduzida. Programas completos são vistos como apenas uma aplicação de função.

III. JAVA LAMBDA EXPRESSIONS

Na última década muitos dos problemas encontrados – como enviar não só dados, mas também comandos através de redes – já foram solucionados em linguagens que suportam o paradigma funcional [5]. Assim, linguagens multi-paradigma têm adicionado suporte à estas mesmas estruturas, aumentando sua flexibilidade e ganho para com os desenvolvedores. O suporte a lambda expressions em Java não tem como objetivo apenas substituir Anonymous Inner Classes, mas também ser capaz de trazer os benefícios deste paradigma ao ponto de incrementar o ecossistema da linguagem.

A. Lambda Expressions e Anonymous Inner Classes

Ao fornecer suporte a funções de primeira classe, também chamadas de lambda expressions ou closures[1], a linguagem Java habilita a substituição de annonymous inner classes (AIC) de forma transparente. Conforme a Listagem 1, em Java a ordenação de inteiros pode ser implementada a partir de uma AIC em conjunto do método sort da classe Arrays.

```
Integer[] integers = new Integer[]{5, 4, 3, 2, 1};

Arrays.sort(integers, new Comparator<Integer>() {
    public int compare(Integer a, Integer b) {
        return a.compareTo(b);
    }
});
```

Listagem 1: Sort - Anonymous Inner Class

Neste trecho de código o método *sort* recebe como primeiro argumento um array de inteiros (já declarado na variável *integers*) e como segundo qualquer instância que satisfaça o contrato de *Comparator*. Assim, para satisfazer o segundo argumento, instância-se uma AIC através da palavra reservada **new** que implementa o método *compare* declarado no contrato.

[1]Lambda expressions ou closures são funções que não exigem vínculos de classe, como exemplo podendo ser atribuída a uma variável. Com esta caraterística, uma função atua como dado, ou seja, pode ser passada como argumento para outras funções.

Entretanto, com closures o mesmo método *sort* pode ter seus argumentos simplificados. Conforme demonstrado na Listagem 2, ao invés de instanciar uma AIC, uma lambda expression é passada como segundo argumento, removendo a necessidade de instanciação de uma classe e a implementação de um contrato imposto por *Comparator*.

Listagem 2: Sort - Lambda Expression

```
Arrays.sort(integers, (a, b) -> a.compareTo(b));
```

De fato, apesar de expressarem o mesmo comportamento a nível de código, ambas funcionalidades possuem diferentes implementações sob a Máquina Virtual Java (JVM). AIC são compiladas, ou seja, geram novos arquivos contendo declarações de classes. Além do mais, ao utilizar a palavra reservada **this** referencia-se a própria instância anônima. Como representam instâncias de uma classe, estas devem ser carregadas pelo classloader e seus construtores invocados pela máquina virtual. Ambas etapas consomem memória, tanto heap [6] para alocação de objetos quanto permgem[2].

Diferentemente de AIC, lambdas postergam a estratégia de compilação para em tempo de execução, utilizando a instrução invokedynamic [7]. Funções são traduzidas para métodos estáticos vinculados ao arquivo da classe correspondente a sua declaração, eliminando o consumo de memória. Agora, ao referir-se a **this**, a classe que delimita a lambda expression é acessada, ao contrário de AIC que acessa sua própria instância.

Dessa forma, o suporte a lambda expressions traz benefícios para os usuários da linguagem. Tal funcionalidade está além de uma mera substituição, pois acrescenta um novo paradigma no ecossistema Java.

IV. TEORIA DAS CATEGORIAS E SUAS ESTRUTURAS

A Teoria das Categorias (TC) foi inventada no início dos anos 1940 por Samuel Eilenberg e Sunders Mac Lane [**?**] como uma ponte entre os diferentes campos da topologia e álgebra [8]. Afim de demonstrar as relações entre estruturas e sistemas matemáticos [9], a TC estabelece uma linguagem formal capaz de encontrar aplicabilidade em várias áreas da ciência. Por volta dos anos 1980, computação e TC passaram a ser consideradas áreas correlatas de estudo [10].

Aplicações do modelo categorial ocorrem na composição de funções encorajada pelo paradigma funcional. Além do mais, em linguagens de programação, o estudo dos tipos pode ser representado através de categorias. Muitos modelos computacionais que fazem uso de estruturas de dados como grafos podem ser generalizados para categorias de grafos. Portanto, tais aplicações demonstram a capacidade de abstração e a importância da TC para a computação.

Tal como a Teoria dos Grupos[3] abstrai a ideia do sistema de permutações como simetrias de um objeto geométrico, a TC manifesta-se como um sistema de funções entre conjuntos de objetos [11]. Esta abstração pode ser vista a partir da Figura 1, onde os conjuntos de objetos são representado por A, B e C. Nesta mesma estrutura, f e g denotam os morfismos entre os diferentes conjuntos de objetos, tal que $f : A \to B$ e $g : B \to C$. Por fim, h expressa a ideia de composição, sendo produto da união dos morfismos f e g.

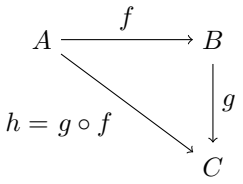

Figura 1: Funções entre coleções de objetos

A TC nasceu como uma ferramenta matemática, com propósito de estudar a relação objetos e morfismos. Contudo, tal conceito tornou-se uma ferramenta utilizada por diversas áreas da ciência. Na computação, diversos problemas são abstraídos através de estruturas da TC, pois ajudam na sua resolução.

A. Categoria – Objetos e Morfismos

Uma categoria consiste em uma coleção de coisas, todas relacionadas de algum modo. As coisas são nomeadas de objetos e as relações de morfismos [8].

Definição [8, 9]: define-se a categoria C como:

(a) uma coleção $Ob(C)$, contendo os objetos de C;

(b) para cada par $a, b \in Ob(C)$, um conjunto $Hom_C(a, b)$ chamado de morfismos de a para b;

(c) para cada objeto $a \in Ob(C)$, um morfismo de *identidade* em a denotado por $id_a \in Hom_C(x, a)$;

(d) para cada três objetos $a, b, c \in Ob(C)$, uma função de composição $\circ : Hom_C(b, c) \times Hom_C(a, b) \to Hom_C(a, c)$;

Dado os objetos $a, b \in Ob(C)$, denota-se o morfismo $f \in Hom_C(a, b)$ por $f : a \to b$; onde a é o domínio e b o contradomínio.

Estas operações em C devem satisfazer os seguintes axiomas:

(Identidade) Para todo objeto $a, b \in Ob(C)$ e todo morfismo $f : a \to b$, tem-se $id_a \circ f = f = id_b \circ f$;

(Associatividade) Sejam os objetos $a, b, c, d \in Ob(C)$ e os morfismos $f : a \to b$, $g : b \to c$ e $h : c \to d$. Então $(h \circ g) \circ f = h \circ (g \circ f)$.

Ambos axiomas podem ser representados pelos diagramas comutativos das Figura 2 e Figura 3.

[2]Área de memória limitada separada da heap chamada Permanent Generation que possui a função de armazenar objetos de geração permanente como metadados, classes e métodos.

[3]Teoria que estuda as estruturas algébricas de grupos. Um grupo é formado por um conjunto de elementos finito ou infinito associado a uma operação binária, como por exemplo a adição ou multiplicação.

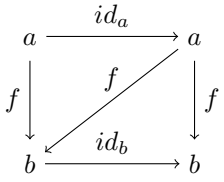

Figura 2: Identidade

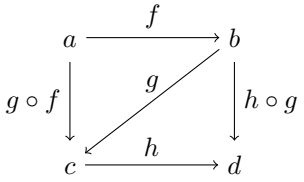

Figura 3: Associatividade

B. Funtor – Morfismos entre Categorias

Um funtor é um mapeamento entre duas categorias de tal modo que o domínio, contradomínio, objetos e morfismos são preservados [11].

Definição [8, 9]: para as categorias C e D, um funtor $F : C \to D$ com domínio C e contradomínio D consiste em duas funções relacionadas. São elas:

(a) a função de objeto F que atribui cada objeto $a \in Ob(C)$ para um objeto $F(a) \in Ob(D)$;

(b) e a função de flecha (também chamada F) que atribui cada morfismo $f : a \to b \in Hom_C(a, b)$ para um morfismo $F(f) : F(a) \to F(b) \in Hom_D(a, b)$.

Tal que os seguintes axiomas são satisfeitos:

(Identidade) Para todo objeto $a \in Ob(C)$ existe um morfismo $F(id_a) = id_{F(a)}$ que preserva a identidade;

(Associatividade) Sejam os objetos $a, b, c \in Ob(C)$ e os morfismos $f : a \to b$ e $g : b \to c$. Então $F(g \circ f) = F(g) \circ F(f)$.

Estas funções capazes de preservar as características das categorias C e D também podem ser ilustradas a partir da Figura 4.

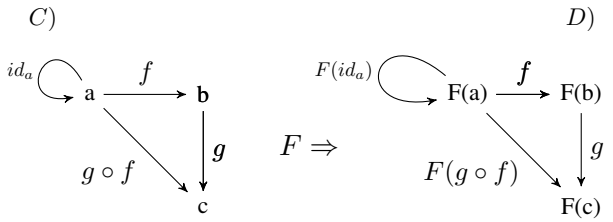

Figura 4: Funtor F

C. Mónade – Endofuntor e Transformações Naturais

Podendo ser visto como um padrão estrutural que ocorre diversos ramos da matemática [8], a construção fundamental mónade (também chamada de tripla) estrutura-se a partir de um endofuntor[4] e das transformações naturais[5] de identidade e multiplicação.

Definição [9, 11]: dada a mónade $T = (T, \eta, \mu)$ em uma categoria C consiste em:

(a) um endofuntor T, tal que $T : C \to C$;

(b) nas transformações naturais:

(*Identidade*) $\eta : id_C \to T$;

(*Multiplicação*) $\mu : T^2 \to T$, onde $T^2 = T \circ T$.

Fazendo com que os diagramas da Figura 5 e Figura 6 comutem, respeitando os axiomas de identidade (esquerda e direita) e associatividade.

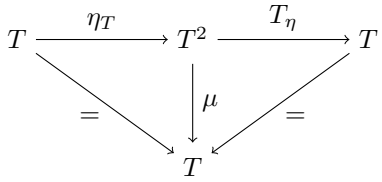

Figura 5: Identidade à Esquerda e à Direita

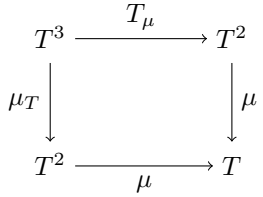

Figura 6: Associatividade

D. Flechas – Categoria de Kleisli

A partir de uma tripla obtemos a Categoria de Kleisli (CK), capaz de representar a mesma estrutura monádica através de uma sintaxe diferente [12]. Mantendo definições similares as já apresentadas em IV-A, a CK destaca-se por abstrair a estrutura de objetos e morfismos sendo capaz compor o endofuntor subjacente.

Definição [9, 12, 13]: dada a tripla (T, η, μ) sob a categoria C, então define-se C_T como:

(a) cada objeto $a \in Ob(C)$, um novo objeto a_T;

(b) cada morfismo $f : a \to T_b$, um novo morfismo $f^* : a_T \to b_T$.

Dado os morfismos $f^* : a_T \to b_T$, $g^* : b_T \to c_T$ e $h : c \to d_T$ e o operador de extensão $-^*$, então:

(*Composição*) $g^* \circ f^* = (\mu_c \circ T(g) \circ f)^*$.

Similar as outras estruturas desta seção, C_T deve obedecer as seguintes leis:

[4]Um funtor que mapeia uma categoria para ela mesma.

[5]Mapeamento entre dois funtores que possuem o mesmo domínio e contradomínio, tal que satisfaça a condição de naturalidade.

(*Identidade à Esquerda*) $f^* \circ \eta_a = f$;

(*Identidade à Direita*) $\eta_a^* \circ h = id_{T(a)} \circ h$;

(*Associatividade*) $(g^* \circ (f^* \circ h)) = (g^* \circ f)^* \circ h$.

V. APLICAÇÃO EM HASKELL

As funções do paradigma funcional podem ser vistas como morfismos na categoria dos tipos. Percebendo estas relações, Wadler [14–17] utilizou o conceito de mónade para estruturar programas puramente funcionais em Haskell. Assim, o primeiro problema pertencente ao conjunto Awkward Squad foi solucionado [18].

A introdução da mónade para E/S (entrada e saída) padronizou a maneira de executar estas ações e encapsular seus efeitos colaterais. Assim, outras mónades foram adicionadas à linguagem, estendendo as bibliotecas padrões para fornecer suporte a exceções, nulidade, concorrência, etc. Logo, mónades trouxeram utilidade para a linguagem, pois confrontaram o Awkward Squad [18].

Com a rápida adoção de mónades para a solução de problemas envolvendo E/S, constatou-se que outros conceitos da TC também poderiam ser aproveitados. Qualquer estrutura recursiva (listas, mapas, árvores, grafos) que pode ser mapeada a partir de uma função é representada por um funtor. A categoria de Haskell, chamada Hask, preocupa-se em tratar tipos como objetos e funções como morfismos, fornecendo funções de composição e identidade.

Type classes definem comportamentos genéricos que podem ser implementados por um conjunto variado de tipos. Esta funcionalidade tornou a implementação das estruturas previamente mencionadas disponíveis para qualquer tipo de dado, dependendo apenas da implementação de suas instancias [19].

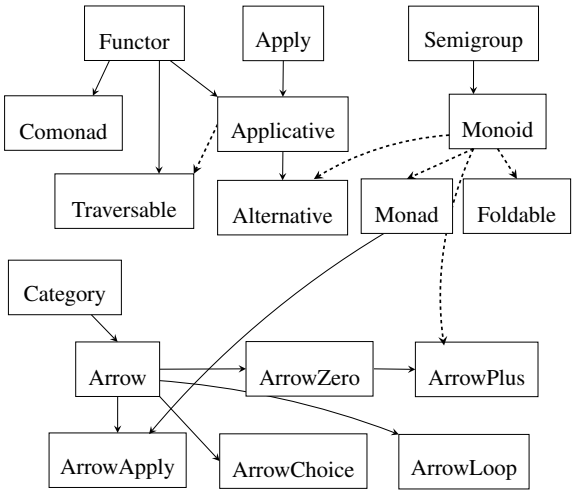

Figura 7: Hierarquia das type classes[6]

[6]Fonte: https://wiki.haskell.org/Typeclassopedia. Acessado em 17/09/2015.

Conforme a Typeclassopedia [20], a Figura 7 demonstra as relações entre as type classes, identificadas por flechas tracejadas e pontilhadas:

(a) (*Tracejadas*) Determinam relações de é-um, ou seja, se existe uma flecha de A para B, então todo B é um A;

(b) (*Pontilhadas*) Indicam algum tipo de relação, como por exemplo a equivalência entre Monoid e MonadPlus.

A implementação de uma type class pode ser visualizada na Listagem 3, onde a primeira declaração incorpora uma nova classe com nome Functor, uma variável de tipo f e um comportamento explicitado pela função fmap. Na segunda, a instanciação de Maybe fornece a implementação de fmap, tornando este tipo um Functor.

```
class Functor f where
    fmap :: (a -> b) -> f a -> f b

instance Functor Maybe where
    fmap f (Just x) = Just (f x)
    fmap f Nothing  = Nothing
```

Listagem 3: Type class para funtores[7]

[7]

A Listagem 4 demonstra a utilização da função fmap em conjunto do variante Just com valor 200 e a função de multiplicação parcialmente aplicada.

```
fmap (*2) (Just 200)
```

Listagem 4: Função fmap para Maybe[7]

<mostrar a type class de monad e explicar type constructors>

Apesar de parecidas, type classes não são iguais à interfaces, presentes na orientação a objetos. Type classes expressam sobrecarga de método (polimorfismo ad-hoc) enquanto interfaces trabalham com instâncias de classes e sobrecarga por subtipagem.

REFERÊNCIAS

[1] P. Hudak, "Conception, evolution, and application of functional programming languages," *ACM Computing Surveys (CSUR)*, vol. 21, no. 3, pp. 359–411, 1989.

[2] K. Louden *et al.*, *Programming Languages: Principles and Practices*. Cengage Learning, 2011.

[3] G. Michaelson, *An Introduction to Functional Programming Through Lambda Calculus*. Courier Corporation, 2011.

[7]Fonte: M. Lipovaca, *Learn You a Haskell for Great Good!: A Beginner's Guide*, 2011.

[4] B. Goetz *et al.*, "State of the lambda," Set. 2013, White Paper. [Online]. Available: http://cr.openjdk.java.net/~briangoetz/lambda/lambda-state-final.html
[5] R. Fischer, *Java Closures and Lambda*, 1st ed. Apress, 2015.
[6] C. Hunt and B. John, *Java Performance*. Prentice Hall Press, 2011.
[7] B. Goetz, "Translation of lambda expressions," Abr. 2012, White Paper. [Online]. Available: http://cr.openjdk.java.net/~briangoetz/lambda/lambda-translation.html
[8] D. I. Spivak, *Category Theory for the Sciences*. The MIT Press, 2014, no. 1.
[9] S. Mac Lane, *Categories for the Working Mathematician*, ser. Graduate Texts in Mathematics. Springer-Verlag, 1971, no. 5.
[10] P. B. Menezes and C. E. Ramisch, "Características da teoria das categorias e sua importância para a ciência da computação," in *STC 2006 - I Simpósio de Teoria das Categorias*. Universidade Federal do Rio Grande do Sul, Mar. 2006.
[11] S. Awodey, *Category Theory - Oxford Logic Guides*, 2nd ed. Oxford University Press, 2010, vol. 52.
[12] J. M. Hill and K. Clarke, "An introduction to category theory, category theory monads, and their relationship to functional programming," Technical Report QMW-DCS-681, Department of Computer Science, Queen Mary and Westfield College, Tech. Rep., 1994.
[13] M. C. Pedicchio and W. Tholen, *Categorical Foundations Special Topics in Order, Topology, Algebra, and Sheaf Theory*. Cambridge University Press, 2004, no. 97.
[14] S. L. Peyton Jones and P. Wadler, "Imperative functional programming," in *Proceedings of the 20th ACM SIGPLAN-SIGACT symposium on Principles of programming languages*. ACM, 1993, pp. 71–84.
[15] P. Wadler, "Monads for functional programming," in *Advanced Functional Programming*. Springer, 1995, pp. 24–52.
[16] ——, "Comprehending monads," *Mathematical structures in computer science*, vol. 2, no. 04, pp. 461–493, 1992.
[17] ——, "The essence of functional programming," in *Proceedings of the 19th ACM SIGPLAN-SIGACT symposium on Principles of programming languages*. ACM, 1992, pp. 1–14.
[18] S. P. Jones, "Tackling the awkward squad: monadic input/output, concurrency, exceptions, and foreign-language calls in haskell," in *Engineering Theories of Software Construction*. Press, 2001, pp. 47–96.
[19] B. O'Sullivan, J. Goerzen and D. B. Stewart, *Real world haskell: Code you can believe in*, 1st ed. O'Reilly Media, Inc., 2008.
[20] B. Yorgey, "Typeclassopedia," https://wiki.haskell.org/Typeclassopedia, Mar. 2009, Acessado em 17/09/2015.