Sistemas com dados versionados

Thiago Rafael Becker

Dezembro de 2011

Resumo

A semântica de um sistema com dados versionados é estudada a partir de um mapeamento de um monóide de transformação para um grafo direto reflexivo. O monóide de transformação é composto pelas operações adição de símbolo e remoção de símbolo sobre o conjuntos gerado pelo fecho de Kleene $\Sigma_\perp^\star$. O funtor de mapeamento deste grupo para um grafo reflexivo direto apresenta a estrutura e a semântica de um VCS.

Introdução

Os sistemas com dados versionados (VCS) são sistemas que armazenam um histórico de modificações dos dados que eles contém. Estes sistemas tem diversas funções tais como gerenciamento de documentos, versões de sistemas e configurações destes sistemas para fins de manter histórico de modificações e *rollback* para correção de erros.

O objeto de estudo deste artigo é o cerne da semântica destes sistema, a transição de estados. Para tal, inicialmente vamos definir como são os objetos manipulados por um destes sistemas como um monóide sobre um alfabeto de símbolos, $\Sigma_\perp^\star$. Com essa definição, vamos definir duas operações de adição e remoção de caracteres destes objetos, e provar que estes formam um monóide de transformação, $T_{\Sigma_\perp^\star}$. Por fim vamos definir um funtor **Commit** que mapeia da categoria livremente gerada por $T_{\Sigma_\perp^\star}$ para `RGr`, assim forncendo a estrutura de um grafo para estes sistemas.

Objetos maniulados

Um SCM gerencia estados de objetos. O conteúdo destes objetos é uma sequência de dados pertencentes ao alfabeto $\Sigma_\perp = \Sigma \cup \perp$. Todos os estados de um objeto são definidos pelo conjnto gerado pelo fecho de Kleene sobre o alfabeto $\Sigma_\perp$, $\Sigma_\perp^\star = \langle \Sigma_\perp, \circ, \varepsilon \rangle$. $\perp$ é um elemento especial que indica indefinição das funções parciais sobre o conjunto $\Sigma_\perp^\star$.

Algumas definições

Definição 1 *Uma palavra em $\Sigma_\perp^\star$ é uma cadeia de símbolos $\sigma = \{\sigma_1\sigma_2 \dots \sigma_n\}$ de tamanho n em Σ ou $\perp$.*

Definição 2 *O comprimento de uma cadeia de símbolos σ de $\Sigma_\perp^\star$ é dado por $len(\sigma)$, $len(\varepsilon) = len(\perp) = 0$.*

Adição e remoção de caracteres

Dois funcionais são definidos sobre $\Sigma_\perp^\star$: *adição de caracter* e *remoção de caracter*. A adição de caracter

$$\mathbf{add\_char}_c^k : \mathbb{N} \to \Sigma \to (\Sigma_\perp^\star \to \Sigma_\perp^\star) \tag{1}$$

definida como

$$\mathbf{add\_char}_c^k(\sigma) = \begin{cases} \perp & \text{se } len(\sigma) < k, \\ \perp & \text{se } \sigma = \perp, \\ \{\sigma_1 \dots \sigma_{k-1} c \sigma_k \dots \sigma_n\} & \end{cases} \tag{2}$$

insere o símbolo c na posição k, e desloca todos os caracteres a direita de k uma posição à direita, ou retorna $\perp$ se $k > len(\sigma)$. A remoção de caractere

$$\mathbf{rem\_char}_c^k : \mathbb{N} \to \Sigma \to (\Sigma_\perp^\star \to \Sigma_\perp^\star) \tag{3}$$

definida como

$$\mathbf{rem\_char}_c^k(\sigma) = \begin{cases} \perp & \text{se } len(\sigma) < k, \\ \perp & \text{se } \sigma_k \neq c, \\ \perp & \text{se } \sigma = \perp, \\ \{\sigma_1 \dots \sigma_{k-1}\varepsilon\sigma_{k+1} \dots \sigma_n\} = \{\sigma_1 \dots \sigma_{k-1}\sigma_{k+1} \dots \sigma_n\} & \end{cases} \tag{4}$$

remove o símbolo c da posição k, e move todos os caracteres à sua direita uma poisção para a esquerda, ou retorna $\perp$ se

- $k > len(\sigma)$
- $\sigma_k \neq c$.

Além disso, a identidade $id : \Sigma_\perp^\star \to \Sigma_\perp^\star$ é definida tal que $id(\sigma) = \sigma$.

Observação 2.1 *Note que as operações* $\mathbf{add\_char}_c^k$ *e* $\mathbf{rem\_char}_c^k$ *não são inversas em* $\Sigma_\perp^\star$ *se* $\mathbf{add\_char}_c^k(\sigma) = \perp$ *ou* $\mathbf{rem\_char}_c^k(\sigma) = \perp$*para* $\perp \neq \sigma \in \Sigma_\perp^\star$*, uma vez que* $(\mathbf{add\_char}_c^k \circ \mathbf{rem\_char}_c^k)(\sigma) = \perp$*. Para todos os outros valores, as operações são inversas,* $(\mathbf{add\_char}_c^k \circ \mathbf{rem\_char}_c^k)(\sigma) = \sigma$ *e* $(\mathbf{rem\_char}_c^k \circ \mathbf{add\_char}_c^k)(\sigma) = \sigma$*.*

Observação 2.2 *Exemplos de indefinição. Dado o alfabeto* $\Sigma = \{a, b\}$ *e a palavra* $\sigma = a^3b^2$ *(*$len(\sigma) = 5$*, as seguintes operações são indefinidas:* $\mathbf{add\_char}_7^a(\sigma)$ *e* $\mathbf{rem\_char}_7^a(\sigma)$ *(pois* $7 > len(\sigma)$*),* $(\mathbf{rem\_char}_3^a \circ \mathbf{rem\_char}_3^a)(\sigma)$ *(pois após a primeira operação a posição 3 de* $\sigma\prime$ *é ocupada pelo símbolo b).*

As funções geradas pelos funcionais $\mathbf{add\_char}_c^k$ e $\mathbf{rem\_char}_c^k$ ($ger\{\mathbf{add\_char}_c^k\}$, $ger\{\mathbf{rem\_char}_c^k\}$) formam um monóide de transformação $T_{\Sigma_\perp^\star} = \langle ger\{\mathbf{add\_char}_c^k\} \cup ger\{\mathbf{rem\_char}_c^k\}, \circ, id\rangle$. Para provarmos as propriedades deste monóide, definimos o mapeamento $\Sigma_\perp^\star \to \mathbb{N}_\perp$

$$\Gamma = \{\perp_{\Sigma_\perp} \mapsto \perp_{\mathbb{N}\cup\perp}, \varepsilon_{\Sigma_\perp} \mapsto 0_{\mathbb{N}\cup\perp}, c_\Sigma \mapsto n_\mathbb{N}\} \tag{5}$$

tal que o mapeamento é uma bijeção. O mapeamento das operações $\mathbf{add\_char}_c^k$ é

$$\Gamma(\mathbf{add\_char}_c^k(\sigma)) = \Gamma(\sigma)p_k^{\Gamma(c)}, p_k \in \mathbb{P} \tag{6}$$

e $\mathbf{rem\_char}_c^k$ é

$$\Gamma(\mathbf{rem\_char}_c^k(\sigma)) = \Gamma(\sigma)p_k^{-\Gamma(c)}, p_k \in \mathbb{P}, \tag{7}$$

e caso exista $n > 0$ tal que $\Gamma(\mathbf{rem\_char}_c^k(\sigma))p_k^{-n} \in \mathbb{N}$, então $\mathbf{rem\_char}_c^k(\sigma) = \perp$. Além disso, se $\Gamma(\mathbf{rem\_char}_c^k(\sigma)) \notin \mathbb{N}$, então $\mathbf{rem\_char}_c^k(\sigma) = \perp$.

O mapeamento de $\Gamma(\sigma)$, $\sigma \in \Sigma_\perp^\star$ é tal que

$$\Gamma(\sigma) = \prod_{i\geq 1} p_i^{\Gamma(\sigma_i)}, p_i \in \mathbb{P}, \sigma_i \in \Sigma_\perp^\star \tag{8}$$

Este mapeamento é uma bijeção nas classes de equvalência de $\mathbb{N}$, tal que $\Gamma^{-1}(n) = \Gamma^{-1}(n\prime), n \neq n\prime, n, n\prime \in \mathbb{N}_\perp$, sendo Γ^{-1} o mapeamento inverso a Γ, isto é, o mapeamento de um número para $\Sigma_\perp^\star$. Este mapeamento permite que existam inúmeros números primos entre p_k e p_{k+1} usados para repesentar σ.

Com isso é possível verificar que $\Gamma(\mathbf{add\_char}_c^k)$ e $\Gamma(\mathbf{rem\_char}_c^k)$ formam um monóide de transformação sobre $\mathbb{N} \cup \{\perp\}$, provando a existencia de $T_{\Sigma_\perp^\star}$.

Representação como grafo

Com as informações acima, é possível criar um funtor que implementa a operação *commit*, **Commit**. O funtor é tal que

$$\mathbf{Commit}(\Sigma_{\perp}^{\star}) = \sigma_1, \sigma_1 \subset \Sigma_{\perp}^{\star} \tag{9}$$

$$\mathbf{Commit}(ger\{\mathbf{add\_char}_c^k\} \cup ger\{\mathbf{rem\_char}_c^k\} \cup \{id\}) = f : \sigma_1 \to \sigma_1\prime, \sigma_1, \sigma_1\prime \in \Sigma_{\perp}^{\star} \tag{10}$$

em que σ_1 é um conjunto unário. **Commit** constrói a estrutura do repositório, e é inversível. Intuitivamente, este funtor toma um conjunto σ_1 e uma composição de morfismos em $ger\{\mathbf{add\_char}_c^k\} \cup ger\{\mathbf{rem\_char}_c^k\} \cup \{id\}$ que transforma o estado do repositório.

O grafo que representa um dado repositório é um subgrafo *Repo* do grafo direto reflexivo

$$\langle \mathbf{Commit}(\Sigma_{\perp}^{\star}), \mathbf{Commit}(ger\{\mathbf{add\_char}_c^k\} \cup ger\{\mathbf{rem\_char}_c^k\} \cup \{id\}) \rangle,$$

e é simples verificar as propriedades compositiva e associativa das setas $\mathbf{Commit}(ger\{\mathbf{add\_char}_c^k\} \cup ger\{\mathbf{rem\_char}_c^k\} \cup \{id\}$ pelo mapeamento Γ acima. Desta forma, a categoria `Repo`

$$\langle \mathbf{Commit}(\Sigma_{\perp}^{\star}), \mathbf{Commit}(ger\{\mathbf{add\_char}_c^k\} \cup ger\{\mathbf{rem\_char}_c^k\} \cup \{id\}), \delta_0, \delta_1, id, \circ \rangle$$

é a categoria livremente gerada pelo grafo *Repo*. O objeto inicial de toda a categoria `Repo` é a palavra vazia ε, que representa um repositório vazio.

Conclusões e desenvolvimentos futuros

O funtor **Commit** da categoria livremente gerada de $T_{\Sigma_{\perp}^{\star}}$ para `Repo` representa o cerne da semântica dos sistema com dados versionados. O trabalho é incompleto, entretanto: não trata de sistemas que possuam *merging* e *branching*, nem de metadados comumente presentes nestes sistemas.

É possível perceber que a categoria `Repo` apresenta uma estrutura semelhante ao de um reticulado. Analogamente aos conjuntos parcialmente ordenados vistos como categorias, *branching* poderia ser visto como um produto de dois objetos, e *merging* como o coproduto de dois objetos, mas isto são coisas a se estudar.

Metadados podem ser estudados como endomorfismos etiquetados, porém a construção dos funtores para a criação e movimento destes metadados pode ser complexa. Isso talvez possa ser possível através de uma gramática de grafos sobre *Repo*.

Além disso, o estudo pode se estender para o estudos de todos os sistemas versionados e seus funtores, para verificar que tipo de estrutura tem estes.