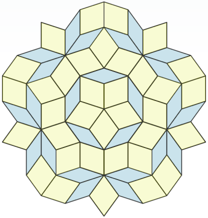


Структуры над словами: образцы и уравнения

Летняя практика, Переславль-Залесский

4–6 июля, 2022 г.

Проектирование структур с образцами

- Вопрос достижимости образца:

```
f (A : x) = Expr1
f [] = Expr2
f [A] = Expr3
```

- Вопрос накрытия образцами:

```
f ((x : y) : z) = Expr1
f [] = Expr2
```

- Вопрос перестановочности образцов:

```
f (x : (A : y)) = Expr1
f (A : (y : z)) = Expr2
```

...а с отказом от свободы и единственности вхождений переменных эти вопросы становятся намного сложнее.

Проектирование структур с образцами

- Вопрос достижимости образца:

```
f {x t t y = Expr1 }
f {x 'A' t z1 'A' y = Expr2 }
f {x 'A' y 'A' z = Expr3 }
```

- Вопрос накрытия образцами:

```
f {x1 (z) x2 t x3  = Expr1}
f {x (z)  = Expr2}
f {t x  = Expr3}
```

- Вопрос перестановочности образцов:

```
f {x, x 'A' : 'A' x = Expr1}
f (x, x 'AB' : 'BA' x = Expr2}
```

Необходимо определить выразительную силу образцов — языки, которые они описывают, и свойства этих языков.

Обозначения для подстановки

В курсе алгебры результат применения подстановки σ к терму t обозначают обычно $\sigma(t)$. У логиков (и в CS) принята постфиксная нотация: $t\sigma$. Мы будем использовать её.

Почему так?

Во многих классических работах по логике (Тарский, Карри) подстановка сразу записывалась в квадратных скобках $[t/A]$. В этом случае было более естественно приписывать её в конец выражения-аргумента: например, $F(t,t)[t/A]$.
В более современных работах приняты обозначения $[t := A]$ и $[t \mapsto A]$. В последнем случае форма стрелки существенна: $\to$ используется для выделения области определения и значений, а не подстановок.

Базовые определения

$V_{\mathcal{T}}$ — множество переменных типа $\mathcal{T}$, $V = \bigcup^{\mathcal{T}} V_{\mathcal{T}}$.
Рассматриваем e-переменные (типа строка/выражение) и t-переменные (типа терм). Т.е. $V = V_{\mathsf{e}} \cup V_{\mathsf{t}}$.

Кратность терма T в образце P обозначаем $|\mathsf{P}|_{\mathsf{T}}$.

Σ — по умолчанию **неограниченный** алфавит констант. $\mathcal{B}[S]$ — множество скобочных структур над строками в алфавите S.

Плоский образец P — строка в алфавите $V_{\mathsf{e}} \cup \mathcal{B}[\Sigma \cup V_{\mathsf{t}}]$. Образец P линеен, если $\forall x \in V_{\mathsf{e}}\ (|\mathsf{P}|_x = 1)$. Подстановка в образец — гомоморфизм, сохраняющий константы (т.е. для всех $\mathbf{A} \in \Sigma$ $\mathbf{A}\sigma = \mathbf{A}$).

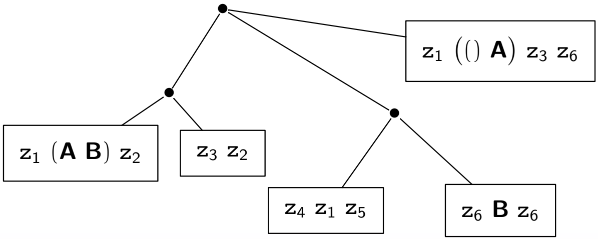
Образец допускает плоское разбиение, если он плоский, либо имеет вид $(\mathsf{P}_1)\ (\mathsf{P}_2) \dots\ (\mathsf{P}_n)\ \mathsf{P}_{n+1}$, где все P_i допускают плоское разбиение. Максимальные плоские подобразцы такого образца называем фрагментами плоского разбиения (ФПР).

Плоские разбиения и деревья

Рассмотрим следующий образец:

$$\Big(\big(z_1\ (\mathbf{A}\ \mathbf{B})\ z_2\big)\ z_3\ z_2\Big)\Big(\big(z_4\ z_1\ z_5\big)\ z_6\ \mathbf{B}\ z_6\Big)\ z_1\ \big(()\ \mathbf{A}\big)\ z_3\ z_6$$

Структура его ФПР приведена ниже.



Поскольку скобочные структуры могут возникнуть только сразу справа от открывающей скобки, то ФПР образуют древесные структуры, аналогичные АТД.

Пример образца, не разбиваемого на ФПР:

$$\big(x_1\ (\mathbf{A}\ x_2)\ x_1\ x_2\big)\ x_1$$

Языки, распознаваемые образцами

Определение

Языком $\mathscr{L}(\mathsf{P})$*, распознаваемым образцом* P*, назовем множество элементов* $\Phi \in \mathcal{B}[\Sigma]$*, для которых существует подстановка* σ*:* $\mathsf{P}\sigma = \Phi$*. Образец* P_1 *сводится к образцу* P_2*, если* $\mathscr{L}(\mathsf{P}_1) \subseteq \mathscr{L}(\mathsf{P}_2)$*.*

Подстановка $x\sigma = \varepsilon$ допустима! В терминологии pattern languages — рассматриваются E-pattern languages (EPL, сокращение от Erasing Pattern Languages, языки стирающих образцов).

- Язык, распознаваемый образцом-строкой $\mathsf{P} \in \Sigma^*$, есть $\{\mathsf{P}\}$.
- Язык, распознаваемый образцом $\mathsf{P} = x_1\, x_2\, x_1$, есть всё множество $\mathcal{B}[\Sigma]$.

Языки, распознаваемые образцами

Определение

Языком $\mathscr{L}(\mathsf{P})$*, распознаваемым образцом* P, *назовем множество элементов* $\Phi \in \mathcal{B}[\Sigma]$, *для которых существует подстановка* σ: $\mathsf{P}\sigma = \Phi$. *Образец* P_1 *сводится к образцу* P_2, *если* $\mathscr{L}(\mathsf{P}_1) \subseteq \mathscr{L}(\mathsf{P}_2)$.

- Язык, распознаваемый образцом-строкой $\mathsf{P} \in \Sigma^*$, есть $\{\mathsf{P}\}$.
- Язык, распознаваемый образцом $\mathsf{P} = \mathtt{x}_1\ \mathtt{x}_2\ \mathtt{x}_1$, есть всё множество $\mathcal{B}[\Sigma]$.

С точки зрения семантики сопоставления, образец $\mathtt{x}_1\ \mathtt{x}_2\ \mathtt{x}_1$ также неудачный: $\mathtt{x}_1$ всегда успешно сопоставляется с ε. Бывает и иначе: хотя $\mathscr{L}(\mathtt{z}_1\ \mathtt{z}_2\ \mathtt{z}_2) = \mathscr{L}(\mathtt{x}_1\ \mathtt{x}_2\ \mathtt{x}_1) = \mathcal{B}[\Sigma]^*$ из-за существования тривиальной подстановки $\mathtt{z}_2 := \varepsilon$, но ленивое сопоставление строки **ABB** с $\mathtt{z}_1\ \mathtt{z}_2\ \mathtt{z}_2$ построит подстановку $\mathtt{z}_2 :=$ **B**, а вовсе не $\mathtt{z}_2 := \varepsilon$.

Сводимость и эквивалентность

Если P_1, P_2 оба из $(V_e \cup \mathcal{B}[\Sigma])^*$, то:

- P_1 сводится к P_2 $\Leftrightarrow$ существует подстановка σ такая, что $P_2\sigma = P_1$;
- если P_2 линеен, тогда вычислительная сложность проверки сводимости образца P_1 к образцу P_2 линейна от суммы длин P_1 и P_2.

Из-за того, что образцы стирающие (определяют EPL), двухсторонняя сводимость не эквивалентна наличию переименовки: вспомним те же $\mathtt{x}_1\ \mathtt{x}_2\ \mathtt{x}_1$ и $\mathtt{z}_1\ \mathtt{z}_2\ \mathtt{z}_2$.

Сводимость и эквивалентность

Если P_1, P_2 оба из $(V_e \cup \mathcal{B}[\Sigma])^*$, то:

- P_1 сводится к P_2 $\Leftrightarrow$ существует подстановка σ такая, что $P_2\sigma = P_1$;
- если P_2 линеен, тогда вычислительная сложность проверки сводимости образца P_1 к образцу P_2 линейна от суммы длин P_1 и P_2.

Если рассматривать только линейные образцы без идущих подряд переменных из V_e, тогда уже для образцов без переменных из V_t выполняется утверждение

$$\mathscr{L}(P_1) = \mathscr{L}(P_2) \Leftrightarrow \exists\sigma(P_1\sigma = P_2 \,\&\, \forall x \in V_e(x\sigma \in V_e))$$

Краткие и избыточные образцы

Определение

Образец P_1 *называется кратким, если любой образец* P_2 *такой, что* $\mathscr{L}(P_1) = \mathscr{L}(P_2)$*, имеет длину, не меньшую, чем* P_1*.*
Иначе P_1 *называется избыточным.*

Пример

Образец $P = \boxed{x_1\ x_2}\ \mathbf{A}\ x_3\ \mathbf{B}\ \boxed{x_1\ x_2}$ избыточен.

Образец $P' = x_1\ x_2\ x_2\ x_1$ является кратким.

Алгебраисты также говорят, что избыточные образцы определяются нетривиальными неподвижными точками морфизмов над образцами (т.е. существует нетривиальная подстановка, переводящая избыточный образец в себя). Например, для P: $x_1 \mapsto \varepsilon$, $x_2 \mapsto x_1\ x_2$.

Критерий избыточности образца (Reidenbach, 2004)

Образец P избыточен, если существует представление $P = Q_0 \, R_1 \, Q_1 \ldots R_n \, Q_n$, $Q_i \in \{\mathcal{B}[\Sigma] \cup V_e\}^*$, $R_i \in {V_e}^+{V_e}^+$, такое, что:

- множества переменных образцов Q_i и R_j не пересекаются;
- в каждом слове R_i найдется имеющая единственное вхождение в R_i переменная x_i *(выделенная)* такая, что

$$\forall j(| \, R_j \, |_{x_i} > 0 \Rightarrow R_j = R_i).$$

Этот критерий является необходимым и достаточным условием при рассмотрении плоских образцов в $\{\mathcal{B}[\Sigma] \cup V_e\}^*$[a].

[a] У Рейденбаха он доказан для образцов в V_e^*. Для образцов над $\mathcal{B}[\Sigma]$ доказательство где-то в моих старых тетрадях — здесь существенно, что скобки порождают бесконечный «алфавит констант».

Критерий Рейденбаха под лупой

Пусть искомое разбиение образца P существует. Тогда по каждому блоку R_i построим подстановку σ_{fix} так: $x_i\sigma_{\mathrm{fix}} = R_i$, а образы прочих переменных из R_i равны ε (они могут встречаться и в сочетании с другой выделенной переменной x_k в прочих R-блоках). Очевидно, $P\sigma_{\mathrm{fix}} = P$.

Образец $P \in V_e^*$ всегда допускает хотя бы две разные нестирающие подстановки $\Leftrightarrow$ образец P избыточен по Рейденбаху.

Критерий Рейденбаха под лупой

Образец $P \in V_e^*$ всегда допускает хотя бы две разные нестирающие подстановки $\Leftrightarrow$ образец P избыточен по Рейденбаху.

- Для образцов, содержащих константные фрагменты, это не верно: $\mathtt{x}_1$ **A** $\mathtt{x}_2$ допускает много подстановок в $\mathbf{A}^n$, хотя является кратким. Однако это верно для фрагментов таких образцов, не содержащих констант.
- Стирающих подстановок может быть и несколько: например, $\mathtt{x}_1\ \mathtt{x}_2\ \mathtt{x}_2\ \mathtt{x}_1$ допускает две подстановки в **A B A B**. Однако поиск возможных подстановок можно экспоненциально ускорить, если пользоваться критерием Рейденбаха.
- Иногда структура слова слишком однородна, чтобы критерий Рейденбаха гарантировал единственность подстановки: см. $\mathbf{A}^n$ и любой краткий образец без констант, содержащий как минимум две различные переменные.

Добавление переменных типа терм

За увеличение выразительной силы образцов приходится платить усложнением теоретических конструкций.

- $\mathscr{L}(\mathsf{P}_1) \subseteq \mathscr{L}(\mathsf{P}_2)$ уже не определяется подстановкой.

  $\mathsf{P}_1 = \mathbf{A}\ \mathtt{x}_1$, $\mathsf{P}_2 = \mathtt{x}_2\ \mathtt{t}$. Язык P_1 вкладывается в язык P_2, а подстановки нет.

- Нет (пока ещё) исследованного понятия избыточного и краткого образца. Более того, образцы для одного и того же языка не образуют нижнюю полурешётку.

  Образцы $\mathtt{x}\ \mathtt{t}$ и $\mathtt{t}\ \mathtt{x}$ оба краткие.

Плавающие t-переменные

Определение

*Назовем переменную t_i в плоском линейном образце P **якорной**, если*

- t_i *имеет кратность, не меньшую 2;*
- *или в P существует подслово α, не содержащее переменных из V_e, такое, что $\alpha = \alpha_1\ t_i\ \alpha_2$, причем α_1 и α_2 оба содержат хотя бы один символ или t-переменную, имеющую кратность в P не меньше 2.*

*В противном случае назовем t_i **плавающей**.*

Пример

Рассмотрим образец $t_1\ t_2\ x_1\ t_3\ t_4\ t_2\ x_2\ t_5$. Якорными переменными являются t_2 и t_1.

Плавающие переменные и языки образцов

Плавающая переменная в образце не подсказывает, какая конкретно буква должна быть подставлена вместо неё, а только указывает на то, что подстановка не пустая. Фрагменты образца, содержащие только e-переменные и плавающие переменные — аналог «нестираемых» фрагментов.

Плавающий сегмент линейного образца P — максимальное подслово P, содержащее только плавающие t-переменные и переменные из V_e.

Образец, в котором все e-переменные входят в плавающие сегменты — аналог нестирающего (non-erasing) образца. Хуже всего, если есть и стирающие, и нестирающие фрагменты.

Плавающие переменные и языки образцов

Плавающий сегмент линейного образца P — максимальное подслово P, содержащее только плавающие t-переменные и переменные из V_e.

Образец, в котором все e-переменные входят в плавающие сегменты — аналог нестирающего (non-erasing) образца. Хуже всего, если есть и стирающие, и нестирающие фрагменты.

Язык образца $P_1 = \mathbf{BBA}\,\mathtt{x}\,\mathbf{ABCDA}$ вкладывается в язык образца $P_2 = \mathtt{z}_0\,\mathtt{t}\,\mathtt{t}\,\mathtt{z}_1\,\mathtt{t}_1\,\mathtt{t}_2\,\mathtt{t}_3\,\mathtt{t}\,\mathtt{z}_2$. Чтобы это доказать, приходится перебирать два случая: пустоты и непустоты подставляемого в x значения.

Multi-pattern languages (Kari, Salomaa)

Определение

Языком $\mathscr{L}(\mathrm{P})$*, распознаваемым множеством образцов* P_{i} *(англ. — multi-pattern language, сокращенно MPL), назовем множество элементов* $\Phi \in \mathcal{B}[\Sigma]^*$*, для которых существует* $\mathtt{i} \in \mathbb{N}$ *и подстановка* σ*:* $\sigma(\mathrm{P}_{\mathtt{i}}) = \Phi$*.*

Множество MPL–объединений стирающих образцов совпадает с множеством MPL–объединений нестирающих образцов. Образец с плавающими $\mathtt{t}$-переменными тоже определяет мультиобразец, и здесь уже смешивание стирающих и нестирающих фрагментов может быть разрешено.

Однако переход от стирающих образцов к нестирающим порождает экспоненциальное разрастание описания MPL.

Пример «плавающего» MPL

Пусть $P_1 = x_1$ **A A C** x_2 **C A B** x_3 **B B C**,
$P_2 = z_1\ t_n\ t_n\ z_2\ t_{m1}\ z_3\ t_{m2}\ z_4\ t_{m3}\ z_5\ t_n\ z_6$.

Множество нестирающих образцов, объединению которых равен P_2:

P_2^1	$t_n\ t_n\ y_1\ y_2\ y_3\ t_n$
P_2^2	$t_n\ t_n\ y_1\ y_2\ y_3\ t_n\ y_4$
P_2^3	$y_0\ t_n\ t_n\ y_1\ y_2\ y_3\ t_n$
P_2^4	$y_0\ t_n\ t_n\ y_1\ y_2\ y_3\ t_n\ y_4$

Множество нестирающих образцов, объединение которых равно P_1, и обобщающие их подстановки в P_2:

A A C C A B B B C	$P_2^3\sigma_1,\ t_n\sigma_1 = \mathbf{C}$
A A C C A B x_3 **B B C**	$P_2^3\sigma_2,\ t_n\sigma_2 = \mathbf{C}$
A A C x_2 **C A B B B C**	$P_2^2\sigma_3,\ t_n\sigma_3 = \mathbf{A}$
A A C x_2 **C A B** x_3 **B B C**	$P_2^2\sigma_4,\ t_n\sigma_4 = \mathbf{A}$
x_1 **A A C C A B B B C**	$P_2^3\sigma_5,\ t_n\sigma_5 = \mathbf{C}$
x_1 **A A C C A B** x_3 **B B C**	$P_2^3\sigma_6,\ t_n\sigma_6 = \mathbf{C}$
x_1 **A A C** x_2 **C A B B B C**	$P_2^4\sigma_7,\ t_n\sigma_7 = \mathbf{A}$
x_1**A A C** x_2 **C A B** x_3 **B B C**	$P_2^4\sigma_8,\ t_n\sigma_8 = \mathbf{A}$

Размер алфавита

Все хорошие свойства образцов, позволяющие работать с ними обычными методами (поиск подстановки, разбиение Рейденбаха) — следствие того, что мы подразумеваем $|\Sigma| = O(|\Sigma_{\texttt{Prog}}|^2)$, где Σ — алфавит входных данных, $\Sigma_{\texttt{Prog}}$ — множество символов, явно входящих в образцы. Допущение реалистичное, учитывая, что «буквами» выступают и константные деревья.

Языки образцов $\mathtt{x}$ **A** **B** $\mathtt{y}$ **A** $\mathtt{z}$ и $\mathtt{x}$ **A** $\mathtt{y}$ **B** **A** $\mathtt{z}$ в алфавите {**A**, **B**, **C**} очевидно не сравнимы: первый распознаёт слово **ABCA**, второй распознаёт **ACBA**. А в алфавите {**A**, **B**} эти образцы описывают один и тот же язык[a].

[a] И поэтому, если алфавит входных данных явно присутствует в образцах, нужны другие способы проверки подстановок на однозначность.

О распознаваемых словах

Предположим, мы рассматриваем краткий образец $\mathsf{P} = \mathrm{x}_1\ \mathrm{x}_2\ \mathrm{x}_2\ \mathrm{x}_1$. Если он сопоставляется со словом $\mathbf{A}^n$, то, как уже было видно, сопоставлений может быть много. С какими ещё словами происходит такая же ситуация?

Чтобы ответить на указанный вопрос, предположим, что слово сопоставилось с P двумя разными способами. То есть нашлись x_1, x_2, z_1, z_2 такие, что $x_1 \neq z_1 \vee x_2 \neq z_2$ и при этом

$$x_1\ x_2\ x_2\ x_1 = z_1\ z_2\ z_2\ z_1$$

Что нам даёт такое равенство и как его упрощать? Ответы на этот вопрос потребуют краткое введение в теорию уравнений в словах.

Уравнения как способ описания однозначности образца

Скажем, что образец $P(x_1, \dots, x_n)$ однозначный, если уравнение $P(x_1, ..., x_n) = P(z_1, ..., z_n)$ имеет только решение $\forall i (x_i = z_i)$.

- Любой образец от одной переменной однозначен;
- Образец $(x_1\, x_2)(x_2\, x_2\, x_1)$ однозначен, поскольку $|x_2|$ определяется как разность длин строк, сопоставляемых с его ФПР $x_1\, x_2$ и $x_2\, x_2\, x_1$;
- Образец $(x_1\, x_2)(x_2\, x_3)(x_3\, x_1)$ однозначен, поскольку решение на длины переменных, входящих в него, всегда единственно, если существует;
- Образец $(x_1\, x_2)(x_2\, x_1)$ неоднозначен (см. выше).

Однозначность длин

Пусть образец P содержит константные строки только вида $(P)^i$ для некоторого простого слова P. Тогда вопрос об его однозначности можно свести к вопросу об единственности решения уравнения на длины входящих в него переменных.

Действительно, рассмотрим слова вида $(P)^{m_j}$, сопоставляемые с ФПР такого образца. По предположению о простоте P, переменные образца получат значения вида $(P)^{k_i}$. В силу уравнения коммутативности для любых двух таких значений, получаем, что существует отображение μ сопоставлений вида $P_j : (P)^{m_j}$ в диофантовы уравнения $\mu(P_j) = m_j$ над $\mathbb{N}$. Такое отображение переводит переменные в себя, конкатенацию — в сложение, а константные фрагменты $(P)^i$ — в натуральные числа i.

Матрица кратностей

Имея ФПР $P_1, P_2, \ldots, P_n$ образца $P(x_1, \ldots, x_m)$, можно построить матрицу $\mathcal{M}(P)$ кратностей переменных образца: $a_{i,j} = |P_i|_{x_j}$.

Если ранг $\mathcal{M}(P)$ равен m, то сопоставление с образцом P всегда имеет единственное решение, если существует.

Если ранг матрицы $\mathcal{M}(P)$ меньше m, то имеет смысл искать такие подмножества ФПР P_i, ранг матрицы кратности которых равен числу входящих в них различных переменных. Тогда все переменные, входящие в такие подмножества, будут сопоставлены однозначно. То есть задача поиска однозначных сопоставлений может быть аппроксимирована сверху задачей поиска ранга $\mathcal{M}(P)$.
(Для поиска целочисленных решений аналогично: ссылка на быстрый алгоритм)

Некоммутативный случай

Коммутативный случай является «худшим» с точки зрения сложности сопоставлений. Но худший случай не всегда достигается, если в образцах встречаются разнородные константные фрагменты.

Образец $\mathtt{x}_1\,\mathbf{A}\,\mathtt{x}_2\,\mathtt{x}_1\,\mathbf{B}\,\mathtt{x}_2$ однозначен (см. выше), хотя порождает уравнение на длины $2x_1 + 2x_2 + 2 = M$, имеющее много решений.

Неоднозначные сопоставления

Предположим, некоторый ФПР неоднозначен и имеет вид $\mathtt{x}_1\ \Phi_1\ \ldots \Phi_n\ \mathtt{x}_{n+1}$, причём подмножество переменных $\mathtt{x}_i$, встречающееся в других ФПР того же образца, непусто. Как эффективно организовать поиск подстановок в $\mathtt{x}_1, \ldots, \mathtt{x}_{n+1}$?

Можно заметить, что удлинение $\mathtt{x}_i$ может происходить не произвольно, а только так, чтобы предварить очередное вхождение Φ_i в сопоставляемую строку. Поэтому имеет смысл запоминать возможные следующие точки возврата по строке для каждой переменной, уже получившей значение.

Пример сопоставления

$\texttt{(x1 ABA x2 BB x3) x2 x3 BBB x4} : \texttt{(ABABBCBBABABB)BBB(AB)}^5\ \texttt{B}^{10}$

Матрица кратностей образца имеет ранг 2, а переменных в образце 4, поэтому анализ длин значений переменных не эффективен.

$$\begin{pmatrix} 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 \end{pmatrix}$$

Приходится делать перебор по вариантам сопоставлений. Выберем лидирующим первый ФПР и обозначим $\Phi_1 = \texttt{ABA}$, $\Phi_2 = \texttt{BB}$. Присваивание $\texttt{x1}\sigma = \varepsilon$ сразу же приводит к удачному сопоставлению с Φ_1. Сопоставляем остаток:
$\texttt{(x2 BB x3) x2 x3 BBB x4} : \texttt{(BBCBBABABB)BBB(AB)}^5\ \texttt{B}^{10}$

Пример сопоставления

$\mathtt{(x1\ ABA\ x2\ BB\ x3)\ x2\ x3\ BBB\ x4} : \mathtt{(ABABBCBBABABB)BBB(AB)}^5\ \mathtt{B}^{10}$

Присваивание $\mathtt{x1}\sigma = \varepsilon$ сразу же приводит к удачному сопоставлению с Φ_1. Сопоставляем остаток:
$\mathtt{(x2\ BB\ x3)\ x2\ x3\ BBB\ x4} : \mathtt{(BBCBBABABB)BBB(AB)}^5\ \mathtt{B}^{10}$
Опять получаем $\mathtt{x2}\sigma = \varepsilon$ и успех сопоставления с Φ_2.
Поскольку $\Phi_1\,\mathtt{x2}\sigma\,\Phi_2$ имеет пустое нетривиальное перекрытие с Φ_1 (не содержит Φ_1, кроме как в позиции префикса, и не кончается префиксом Φ_1), то удлинение x1 сдвигаем сразу за 5-ю позицию во входной строке: $\mathtt{x1}\sigma = \mathtt{ABABB{+}{+}z}$.
Подстановка $\mathtt{x3}\sigma = \mathtt{CBBABABB}$ определена однозначно и сразу же приводит к неуспеху сопоставления во втором ФПР.
Поскольку этот ФПР мы условно прошли до конца, можно построить массив совпадений его подстрок с Φ_i. $\Phi_2\ \mathtt{x3}\sigma$ имеет нетривиальное перекрытие с Φ_2 начиная с 4-й и 9-й позиции, и с Φ_1 на 6-й позиции. Удлиняем x2.

Пример сопоставления

(x1 ABA x2 BB x3) x2 x3 BBB x4 : (ABABBCBBABABB)BBB(AB)^5 B^{10}

Опять получаем $x2\sigma = \varepsilon$ и успех сопоставления с Φ_2.
Поскольку $\Phi_1\, x2\sigma\, \Phi_2$ имеет пустое нетривиальное перекрытие с Φ_1 (не содержит Φ_1, кроме как в позиции префикса, и не кончается префиксом Φ_1), то удлинение x1 сдвигаем сразу за 5-ю позицию во входной строке: $x1\sigma =$ ABABB++z.
Подстановка $x3\sigma =$ CBBABABB определена однозначно и сразу же приводит к неуспеху сопоставления во втором ФПР.
Поскольку этот ФПР мы условно прошли до конца, можно построить массив совпадений его подстрок с Φ_i. Φ_2 $x3\sigma$ имеет нетривиальное перекрытие с Φ_2 начиная с 4-й и 9-й позиции, и с Φ_1 на 6-й позиции. Удлиняем x2. Текущее сопоставление:
(x2 BB x3) x2 x3 BBB x4 : (BBABABB)BBB(AB)^5 B^{10}
$x2\sigma =$ BBC (удлинение на BBCBBABA++z)
$x1\sigma = \varepsilon$ (удлинение на ABABBCBB++z)

Пример сопоставления

$(x1\ ABA\ x2\ BB\ x3)\ x2\ x3\ BBB\ x4 : (ABABBCBBABABB)BBB(AB)^5\ B^{10}$

Текущее сопоставление:

$(x2\ BB\ x3)\ x2\ x3\ BBB\ x4 : (BBABABB)BBB(AB)^5\ B^{10}$

$x2\sigma = BBC$ (удлинение на BBCBBABA++z)

$x1\sigma = \varepsilon$ (удлинение на ABABBCBB++z)

Здесь можно принять нетривиальное решение сразу же проверить, сопоставится ли такое значение x2 во втором ФПР. Это сэкономит один откат. Однако мы этого не сделаем и построим очередное сопоставление вида $x3\sigma = ABABB$, после чего уже проверим сопоставление во втором ФПР. Опять неудача, требуется откат, и опять по x2.

Пример сопоставления

(x1 ABA x2 BB x3) x2 x3 BBB x4 : (ABABBCBBABABB)BBB(AB)^5 B^{10}

Здесь можно принять нетривиальное решение сразу же проверить, сопоставится ли такое значение x2 во втором ФПР. Это сэкономит один откат. Однако мы этого не сделаем и построим очередное сопоставление вида $\mathtt{x3}\sigma = \mathtt{ABABB}$, после чего уже проверим сопоставление во втором ФПР. Опять неудача, требуется откат, и опять по x2. Состояние сопоставления:

(x2 BB x3) x2 x3 BBB x4 : (BB)BBB(AB)^5 B^{10}

$\mathtt{x2}\sigma = \mathtt{BBCBBABA}$ (удлинений нет)

$\mathtt{x1}\sigma = \varepsilon$ (удлинение на ABABBCBB++z)

При удлинении x1 из массива совпадений подстрок входной строки с Φ_2 удаляются элементы, позиции которых перекрываются с новым значением $\mathtt{x1}\sigma$ Φ_1.

Пример сопоставления

$\texttt{(x1 ABA x2 BB x3) x2 x3 BBB x4} : \texttt{(ABABBCBBABABB)BBB(AB)}^5\ \texttt{B}^{10}$

Состояние сопоставления:

$\texttt{(x2 BB x3) x2 x3 BBB x4} : \texttt{(BB)BBB(AB)}^5\ \texttt{B}^{10}$

$\texttt{x2}\sigma = \texttt{BBCBBABA}$ (удлинений нет)

$\texttt{x1}\sigma = \varepsilon$ (удлинение на $\texttt{ABABBCBB++z}$)

При удлинении $\texttt{x1}$ из массива совпадений подстрок входной строки с Φ_2 удаляются элементы, позиции которых перекрываются с новым значением $\texttt{x1}\sigma$ Φ_1.

Построение подстановки $\texttt{x3}\sigma = \varepsilon$ приводит к неудаче при сопоставлении во втором ФПР, поэтому откатываемся на удлинение $\texttt{x1}$. Дальше всё однозначно ($\texttt{x2}\sigma = \texttt{x3}\sigma = \varepsilon$), сопоставление со вторым ФПР оказывается успешным.