Descriptive Patterns and Chains of Pattern Languages

Dominik D. Freydenberger

Goethe University, Frankfurt am Main

Theorietag 2013

Pattern Languages

Basic Definitions

- finite **terminal** alphabet Σ, infinite **variable** alphabet X
- **pattern**: word $\alpha \in X^+$
- **substitution**: morphism $\sigma : X^+ \to \Sigma^*$

Pattern Language

$L(\alpha) := \{\sigma(\alpha) \mid \sigma \text{ is a substitution}\}$

- also called terminal-free E-pattern languages

Example

$L(xx) = \{ww \mid w \in \Sigma^*\}$

Pattern Languages

Basic Definitions

- finite **terminal** alphabet Σ, infinite **variable** alphabet X
- **pattern**: word $\alpha \in X^+$
- **substitution**: morphism $\sigma : X^+ \to \Sigma^*$

Descriptive Pattern

A pattern $\delta \in X^+$ is **descriptive** of a language S if

1. $L(\delta) \supseteq S$, and
2. there is no $\gamma \in X^+$ with $L(\delta) \supset L(\gamma) \supseteq S$.

Pattern Language

$L(\alpha) := \{\sigma(\alpha) \mid \sigma \text{ is a substitution}\}$

- also called terminal-free E-pattern languages

Example

$L(xx) = \{ww \mid w \in \Sigma^*\}$

Pattern Languages

Basic Definitions

- finite **terminal** alphabet Σ, infinite **variable** alphabet X
- **pattern**: word $\alpha \in X^+$
- **substitution**: morphism $\sigma : X^+ \to \Sigma^*$

Descriptive Pattern

A pattern $\delta \in X^+$ is **descriptive** of a language S if

1. $L(\delta) \supseteq S$, and
2. there is no $\gamma \in X^+$ with $L(\delta) \supset L(\gamma) \supseteq S$.

Pattern Language

$L(\alpha) := \{\sigma(\alpha) \mid \sigma \text{ is a substitution}\}$

- also called terminal-free E-pattern languages

Example

$L(xx) = \{ww \mid w \in \Sigma^*\}$

- descriptive patterns are the best approximations that are possible with patterns
- a language can have more than one descriptive pattern

Learning with descriptive patterns

Descriptive Generalization [F., Reidenbach COLT 2010/JCSS 2013]

- learning from positive examples (similar to Gold style learning)
- instead of learning a target language T **exactly**, learn a pattern that is descriptive of T as **approximation**

Learning with descriptive patterns

Descriptive Generalization [F., Reidenbach COLT 2010/JCSS 2013]

- learning from positive examples (similar to Gold style learning)
- instead of learning a target language T **exactly**, learn a pattern that is descriptive of T as **approximation**

- techniques were adapted to learning regular expressions for XML schema inference (F., Kötzing ICDT 2013)

Learning with descriptive patterns

Descriptive Generalization [F., Reidenbach COLT 2010/JCSS 2013]

- learning from positive examples (similar to Gold style learning)
- instead of learning a target language T **exactly**, learn a pattern that is descriptive of T as **approximation**

- techniques were adapted to learning regular expressions for XML schema inference (F., Kötzing ICDT 2013)

Canonical Strategy

Given a finite sample S, compute a pattern that is descriptive of S.

Theorem (F., Reidenbach)

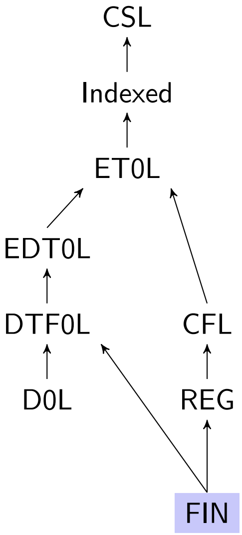
The canonical strategy identifies a language L iff there exists a pattern δ that is descriptive

- of L, and
- of a finite set $S \subseteq L$.

Existence of descriptive patterns

Previous results

- every finite language has a descriptive pattern (Jiang, Kinber, Salomaa, Salomaa, Yu)



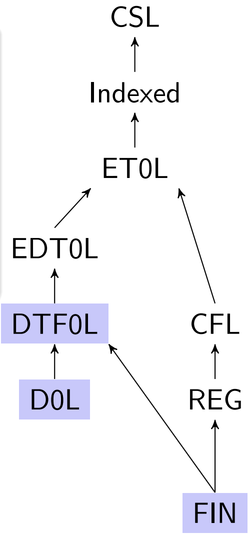
Existence of descriptive patterns

Previous results

- every finite language has a descriptive pattern (Jiang, Kinber, Salomaa, Salomaa, Yu)
- every DTF0L language has a descriptive pattern (F., Reidenbach)

L system letters

- D: morphism(s)
- E: non-terminals allowed
- F: finite set of starting words
- T: tables (multiple rule sets/morphism)



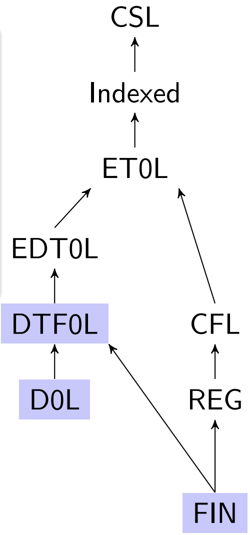
Existence of descriptive patterns

Previous results

- every finite language has a descriptive pattern (Jiang, Kinber, Salomaa, Salomaa, Yu)
- every DTF0L language has a descriptive pattern (F., Reidenbach)
- there are languages for which no pattern is descriptive (F., Reidenbach 2009) **!!!**

L system letters

- D: morphism(s)
- E: non-terminals allowed
- F: finite set of starting words
- T: tables (multiple rule sets/morphism)



Motivation

Theorem [F., Reidenbach]

There are languages for which no pattern is descriptive.

But. . .

- how complicated are those languages?
- how common are those languages?

Motivation

Theorem [F., Reidenbach]

There are languages for which no pattern is descriptive.

But...

- how complicated are those languages?
- how common are those languages?

Agenda

In order to understand descriptive generalization, we need further insights into existence of descriptive patterns

Further potential uses

insights into E-pattern languages (equivalence problem), word equation systems, regex, ...

Chains of Pattern Languages

Assume that no pattern is descriptive of L

$L(\alpha_0) \supseteq L$

Chains of Pattern Languages

Assume that no pattern is descriptive of L

$L(\alpha_0) \supset L(\alpha_1) \supseteq L$

Chains of Pattern Languages

Assume that no pattern is descriptive of L

$L(\alpha_0) \supset L(\alpha_1) \supset L(\alpha_2) \supseteq L$

Chains of Pattern Languages

Assume that no pattern is descriptive of L

$L(\alpha_0) \supset L(\alpha_1) \supset L(\alpha_2) \supset L(\alpha_3) \supseteq L$

Chains of Pattern Languages

Assume that no pattern is descriptive of L

$$L(\alpha_0) \supset L(\alpha_1) \supset L(\alpha_2) \supset L(\alpha_3) \supset \ldots \supseteq L$$

Chains of Pattern Languages

Assume that no pattern is descriptive of L

$L(\alpha_0) \supset L(\alpha_1) \supset L(\alpha_2) \supset L(\alpha_3) \supset \ldots \supseteq L$

Observation

$(L(\alpha_i))_{i=0}^{\infty}$ is an infinite descending chain

Chains of Pattern Languages

Assume that no pattern is descriptive of L

$L(\alpha_0) \supset L(\alpha_1) \supset L(\alpha_2) \supset L(\alpha_3) \supset \ldots \supseteq L$

Observation

$(L(\alpha_i))_{i=0}^{\infty}$ is an infinite descending chain

The existence of such a chain is necessary for the non-existence of a descriptive pattern. Is it also sufficient?

Chains of Pattern Languages

Assume that no pattern is descriptive of L

$L(\alpha_0) \supset L(\alpha_1) \supset L(\alpha_2) \supset L(\alpha_3) \supset \ldots \supseteq L$

Observation

$(L(\alpha_i))_{i=0}^{\infty}$ is an infinite descending chain

The existence of such a chain is necessary for the non-existence of a descriptive pattern. Is it also sufficient?

$L(x^2\, y^2)$

Chains of Pattern Languages

Assume that no pattern is descriptive of L

$L(\alpha_0) \supset L(\alpha_1) \supset L(\alpha_2) \supset L(\alpha_3) \supset \ldots \supseteq L$

Observation

$(L(\alpha_i))_{i=0}^{\infty}$ is an infinite descending chain

The existence of such a chain is necessary for the non-existence of a descriptive pattern. Is it also sufficient?

$L(x^2\, y^2) \supset L(x^2\, y^4)$

Chains of Pattern Languages

Assume that no pattern is descriptive of L

$L(\alpha_0) \supset L(\alpha_1) \supset L(\alpha_2) \supset L(\alpha_3) \supset \ldots \supseteq L$

Observation

$(L(\alpha_i))_{i=0}^{\infty}$ is an infinite descending chain

The existence of such a chain is necessary for the non-existence of a descriptive pattern. Is it also sufficient?

$L(x^2\, y^2) \supset L(x^2\, y^4) \supset L(x^2\, y^8)$

Chains of Pattern Languages

Assume that no pattern is descriptive of L

$L(\alpha_0) \supset L(\alpha_1) \supset L(\alpha_2) \supset L(\alpha_3) \supset \ldots \supseteq L$

Observation

$(L(\alpha_i))_{i=0}^{\infty}$ is an infinite descending chain

The existence of such a chain is necessary for the non-existence of a descriptive pattern. Is it also sufficient?

$$L(x^2\, y^2) \supset L(x^2\, y^4) \supset L(x^2\, y^8) \supset L(x^2\, y^{16})$$

Chains of Pattern Languages

Assume that no pattern is descriptive of L

$L(\alpha_0) \supset L(\alpha_1) \supset L(\alpha_2) \supset L(\alpha_3) \supset \ldots \supseteq L$

Observation

$(L(\alpha_i))_{i=0}^{\infty}$ is an infinite descending chain

The existence of such a chain is necessary for the non-existence of a descriptive pattern. Is it also sufficient?

$$L(x^2\,y^2) \supset L(x^2\,y^4) \supset L(x^2\,y^8) \supset L(x^2\,y^{16}) \ldots \supset L(x^2)$$

- $\lim_{n\to\infty} L(x^2\,y^{2^n}) = \bigcap_{n=1}^{\infty} L(x^2\,y^{2^n})$
- this chain converges towards a pattern language
- the y variables are **redundant**

Some Observations

Theorem [Filè; JKSSY]

$L(\alpha) \supseteq L(\beta)$ iff there is a morphism φ with $\varphi(\alpha) = \beta$

$$\begin{tikzcd}
 & w \\
\beta \arrow[ur, "\sigma"] & \\
 & \alpha \arrow[ul, "\varphi"] \arrow[uu, swap, "\sigma \circ \varphi"]
\end{tikzcd}$$

This leads to various tools for reasoning about patterns ($\Rightarrow$ morphic primitivity)

Some Observations

Theorem [Filè; JKSSY]

$L(\alpha) \supseteq L(\beta)$ iff there is a morphism φ with $\varphi(\alpha) = \beta$

$$\begin{tikzcd} & w \\ \beta \arrow[ur, "\sigma"] & \\ & \alpha \arrow[ul, "\varphi"] \arrow[uu, "\sigma \circ \varphi"'] \end{tikzcd}$$

This leads to various tools for reasoning about patterns ($\Rightarrow$ morphic primitivity)

Main idea of F., Reidenbach 2009

- define descending chain $(\alpha_i)_{i=0}^{\infty}$ such that there is an ascending chain $(\beta_i)_{i=0}^{\infty}$ with $\bigcap L(\alpha_i) = \bigcup L(\beta_i)$
- every α_i can be mapped to every β_j
- if a pattern γ can be mapped to infinitely many β_j, it can be mapped to some α_i be rewriting those morphisms

- the structures of the language and of the chain are closely related
- it seems that is always the case
- Can we exploit this?

Main Results

Main Conjecture

Let $L \subseteq \Sigma^*$. If there is a descending chain $(\alpha_i)_{i=0}^{\infty}$ such that

- $L \subseteq \bigcap L(\alpha_i)$, and
- no α_i contains redundant variables w.r.t. L,

then no pattern is descriptive of L.

Main Results

Wrong Conjecture

Let $L \subseteq \Sigma^*$. If there is a descending chain $(\alpha_i)_{i=0}^{\infty}$ such that

- $L \subseteq \bigcap L(\alpha_i)$, and
- no α_i contains redundant variables w.r.t. L,

then no pattern is descriptive of L.

Main Results

- there exists a counterexample to the conjecture

Main Results

Wrong Conjecture

Let $L \subseteq \Sigma^*$. If there is a descending chain $(\alpha_i)_{i=0}^{\infty}$ such that

- $L \subseteq \bigcap L(\alpha_i)$, and
- no α_i contains redundant variables w.r.t. L,

then no pattern is descriptive of L.

Main Results

- there exists a counterexample to the conjecture
- there exists a language that is covered by an **uncountable** number of descending chains (which are reduced and pairwise distinct)...

Main Results

Wrong Conjecture

Let $L \subseteq \Sigma^*$. If there is a descending chain $(\alpha_i)_{i=0}^{\infty}$ such that

- $L \subseteq \bigcap L(\alpha_i)$, and
- no α_i contains redundant variables w.r.t. L,

then no pattern is descriptive of L.

Main Results

- there exists a counterexample to the conjecture
- there exists a language that is covered by an **uncountable** number of descending chains (which are reduced and pairwise distinct)...
- ...and this language has a descriptive pattern

Existence of descriptive patterns (revisited)

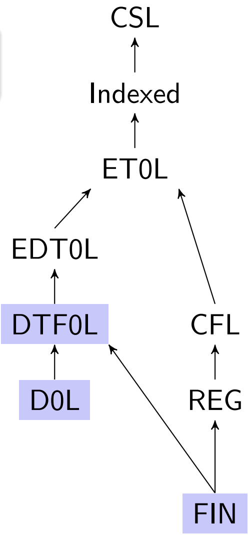
These main results use a number of example languages. . .

Existence of descriptive patterns (revisited)

These main results use a number of example languages...

Theorem

All those example languages are EDT0L languages.



L system letters

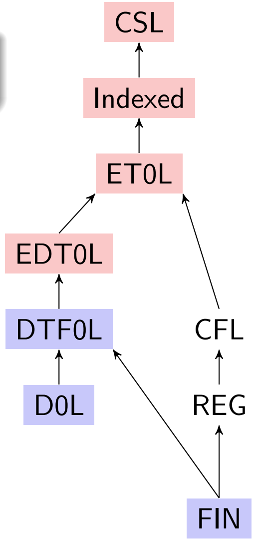
- D: morphism(s)
- E: non-terminals allowed
- F: finite set of starting words
- T: tables (multiple rule sets/morphism)

Existence of descriptive patterns (revisited)

These main results use a number of example languages...

Theorem

All those example languages are EDT0L languages.



L system letters

- D: morphism(s)
- E: non-terminals allowed
- F: finite set of starting words
- T: tables (multiple rule sets/morphism)

Main Technical Contribution

- Proofs use a generalization of the technique of F., Reidenbach 2009.

Main Technical Contribution

- Proofs use a generalization of the technique of F., Reidenbach 2009.

Chain Systems

A chain system is a pair $(\alpha, (\varphi_i)_{i=0}^{\infty})$, where

- α is a pattern
- $(\varphi_i)_{i=0}^{\infty}$ is an infinite sequence of morphisms that guarantee proper inclusion

- By applying the morphisms to α in sequence, we get a decreasing chain of patterns.
- Due to Reidenbach, Schneider 2009, such morphisms have a straightforward syntactic characterization.
- Due to a large number of Lemmas, we can use chain systems as construction kit for (counter) examples.

Conclusions

Main observations

- although languages without a descriptive pattern and the associated descending chains are connected, it is hard to exploit this connection
- a characterization of these languages is probably quite difficult
- these languages can be comparatively simple (EDT0L)
- status of REG and CFL remain open

Positive aspect

Chain systems might become a useful tool for further work in this direction