The (Almost) Complete Guide to Tree Pattern Containment

Wojciech Czerwiński

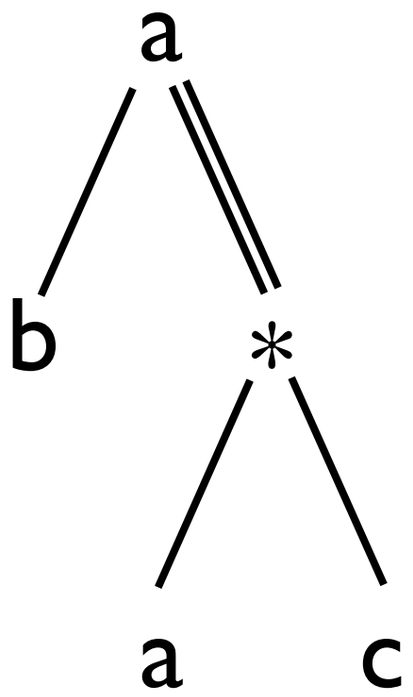
Wim Martens

Paweł Parys

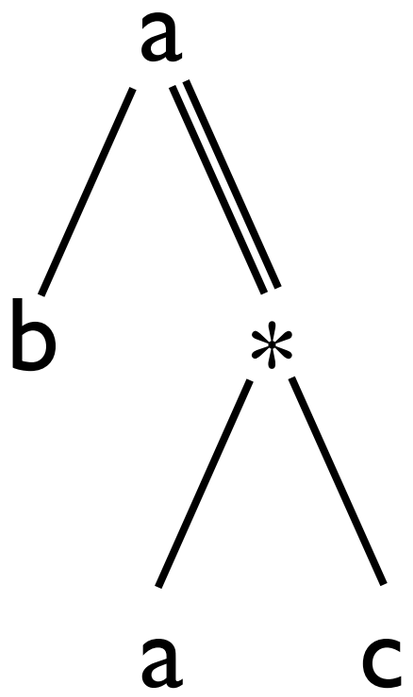
Marcin Przybyłko

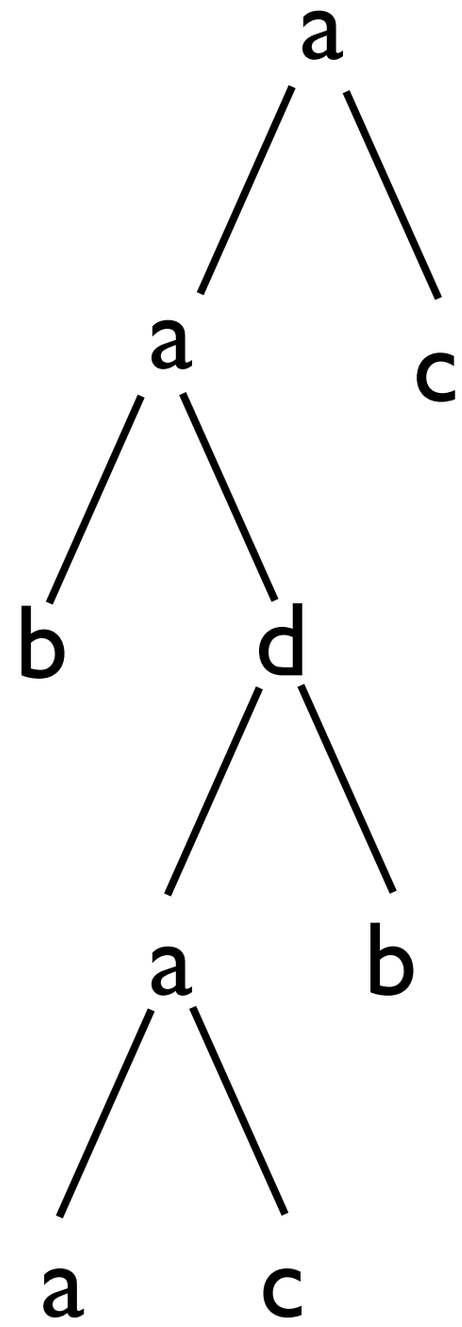
Tree Pattern

Tree Pattern

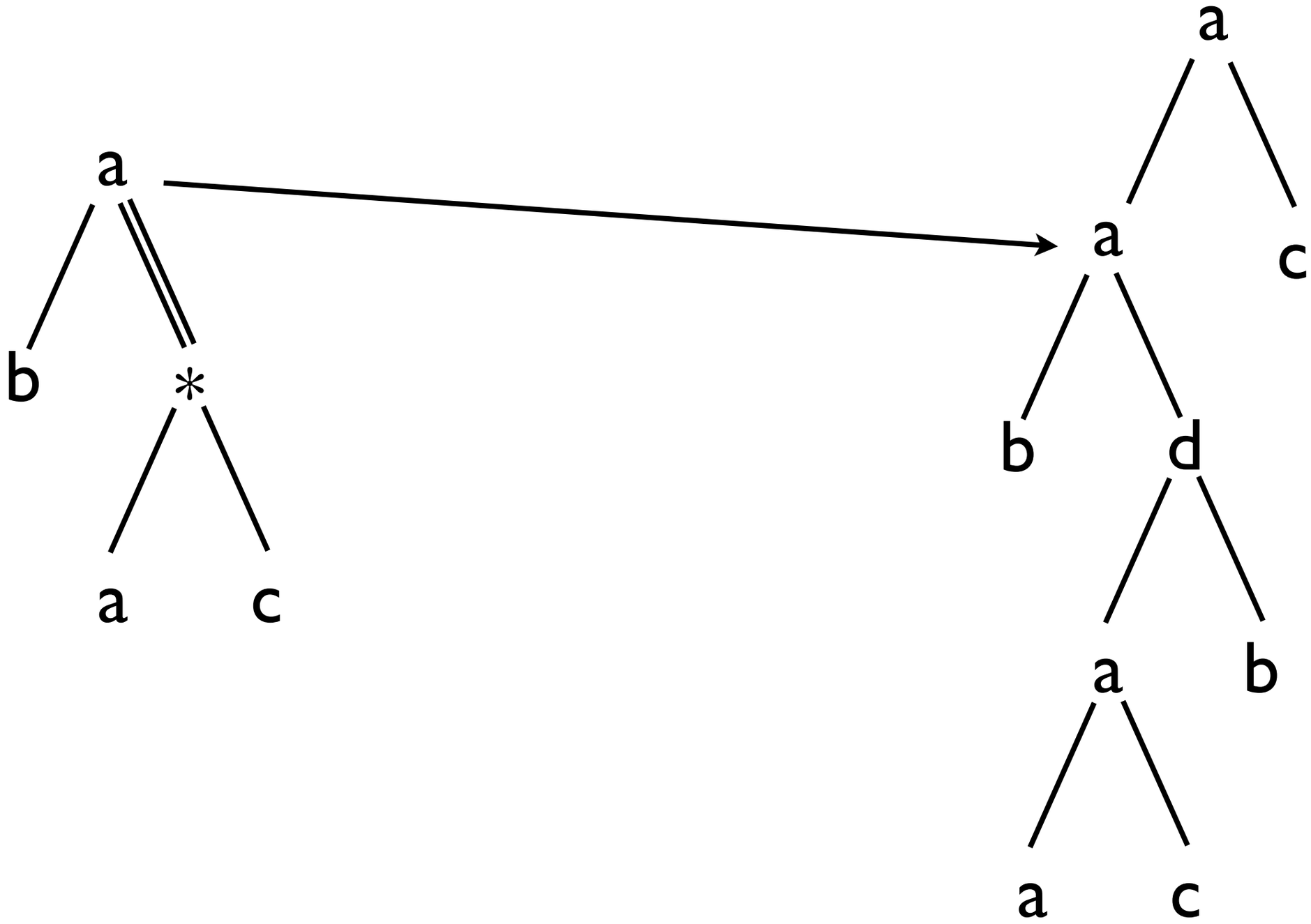


Tree Pattern

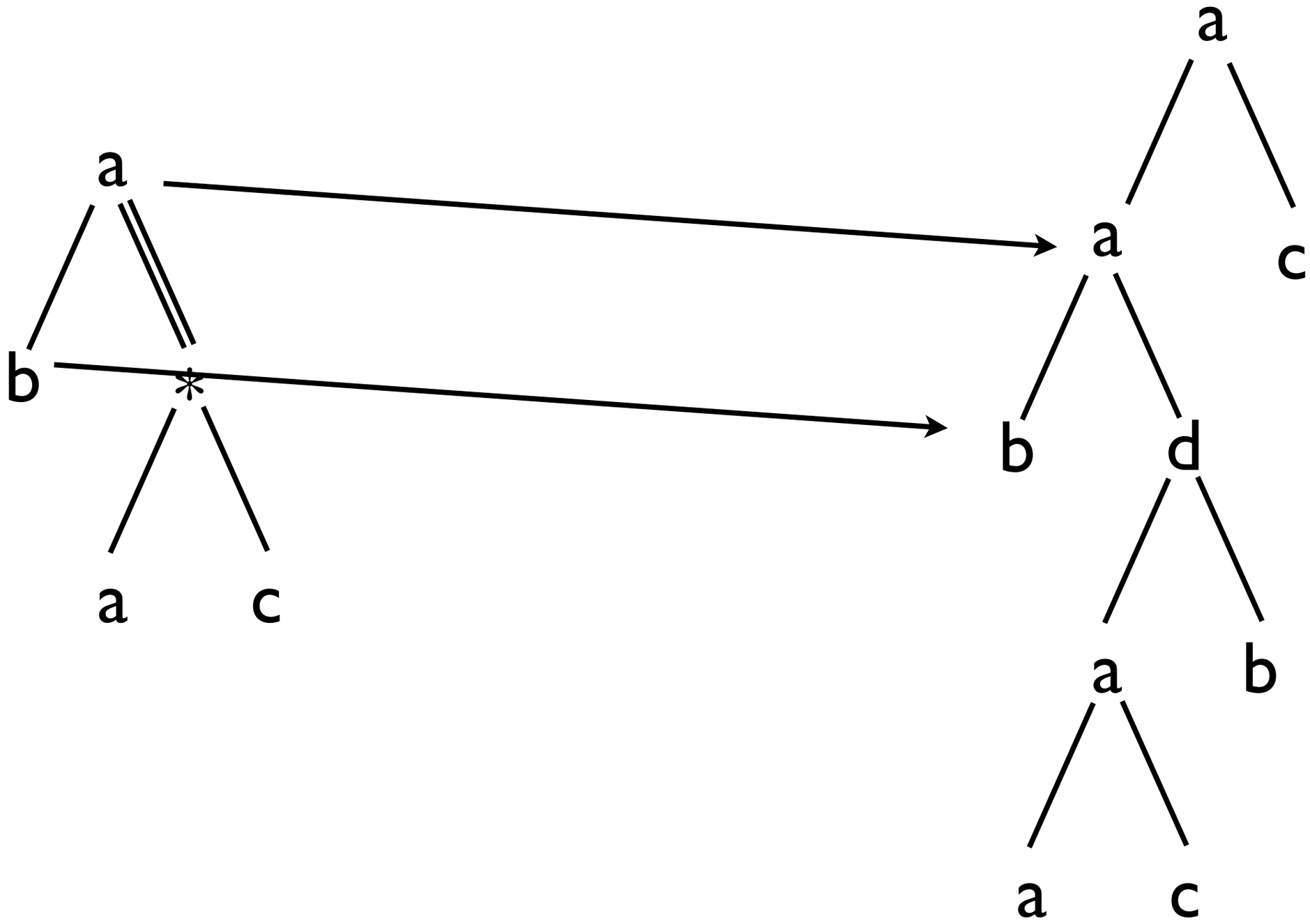




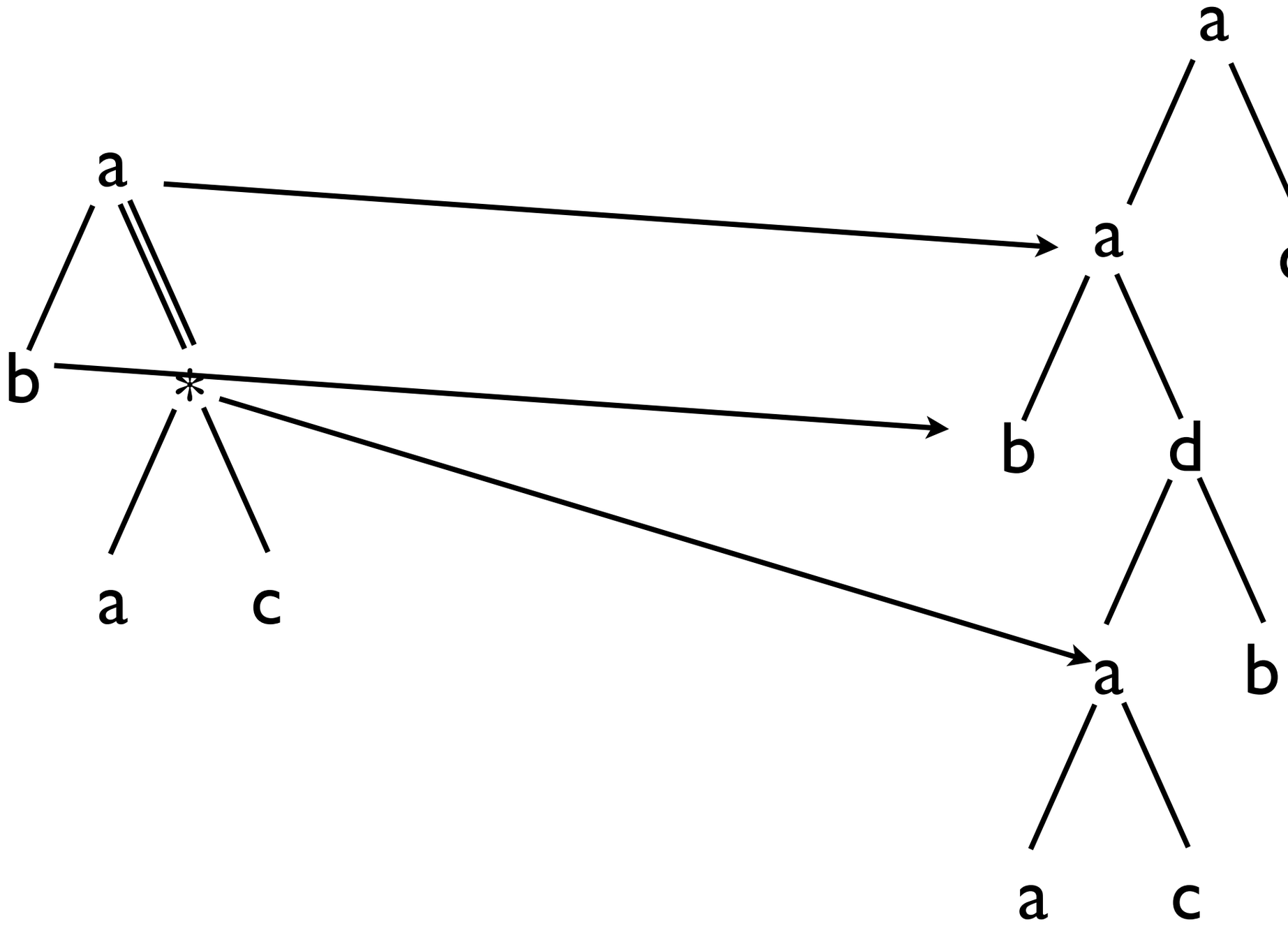
Tree Pattern



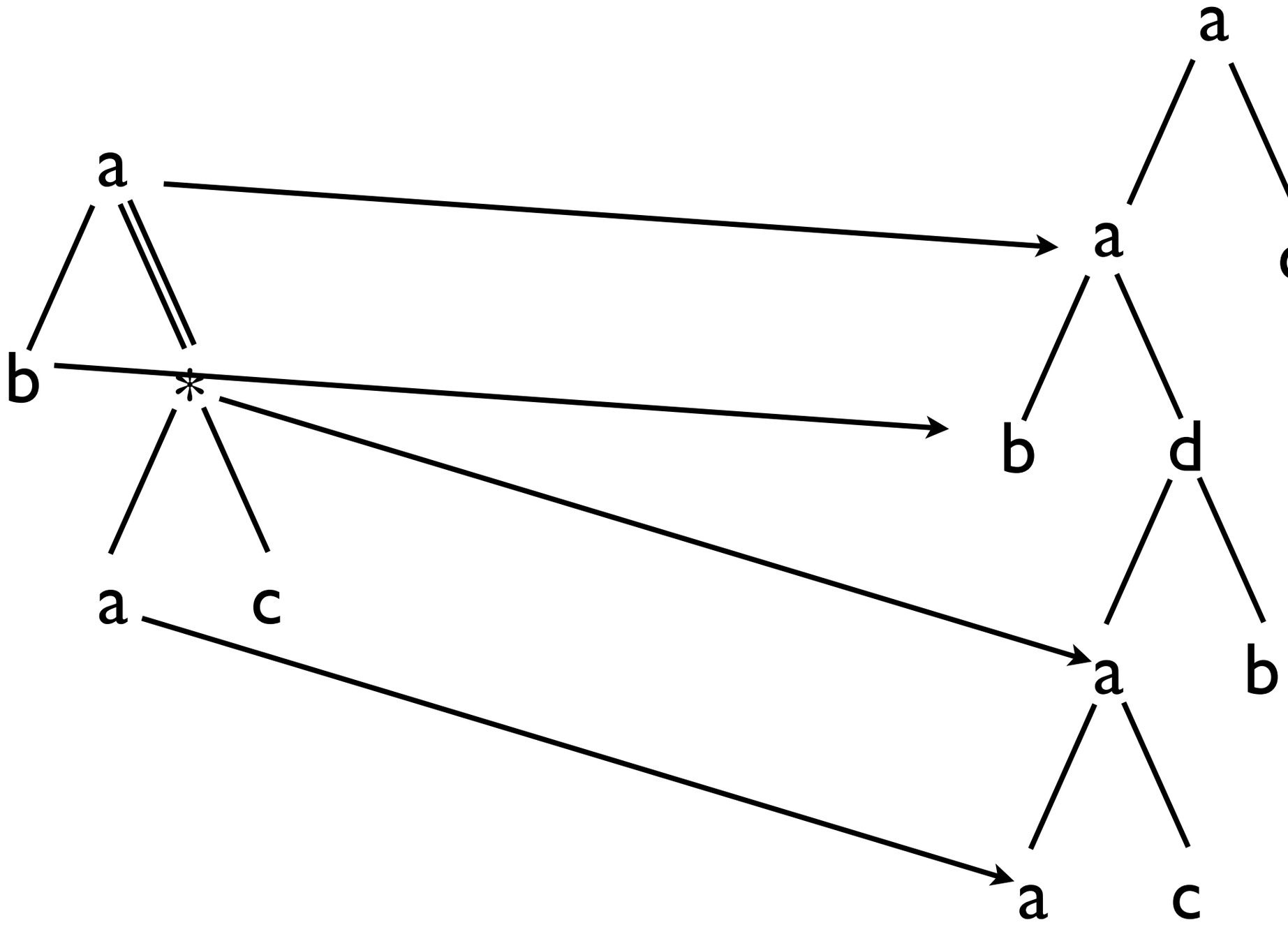
Tree Pattern



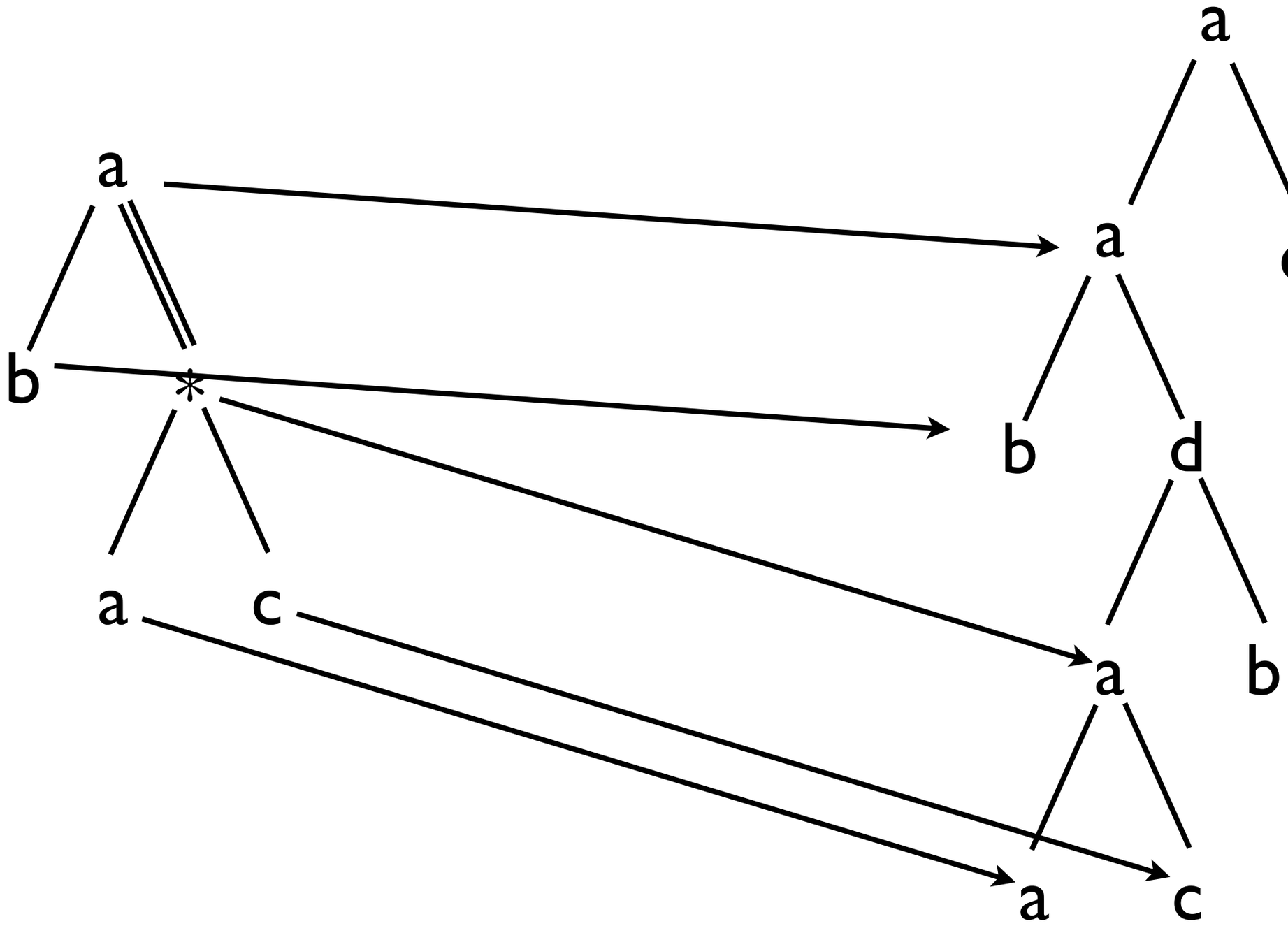
Tree Pattern



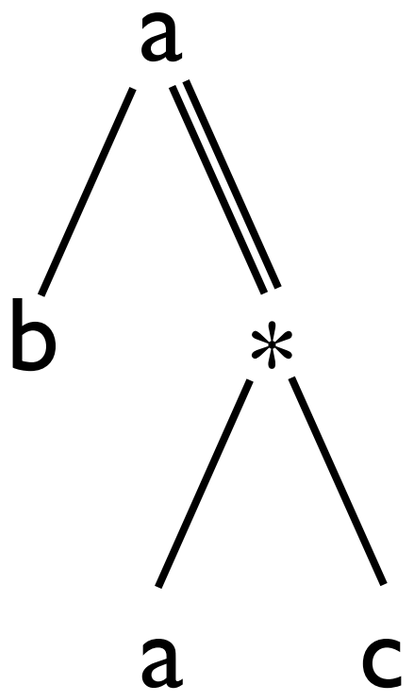
Tree Pattern

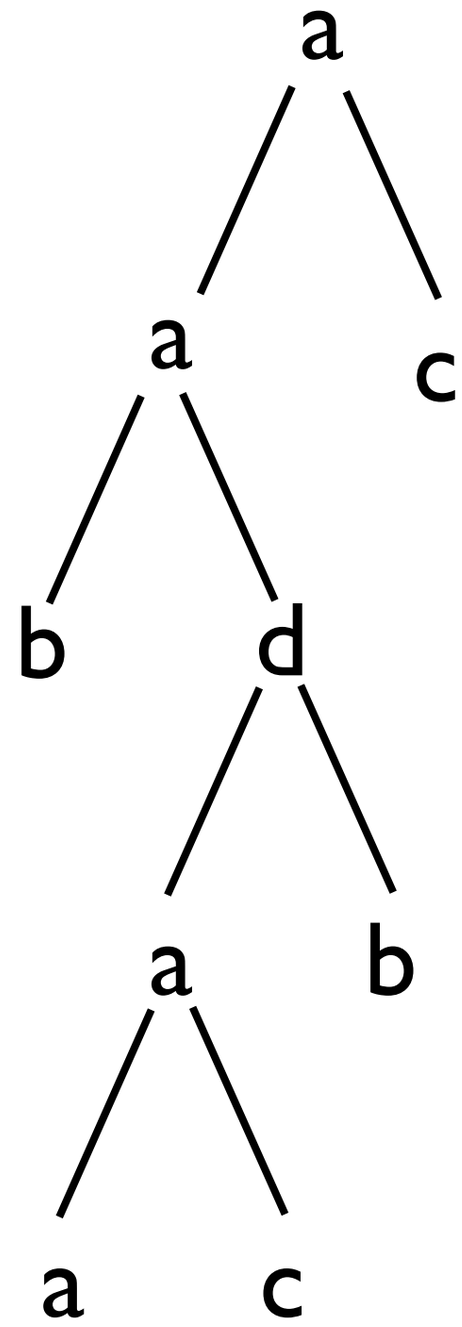


Tree Pattern

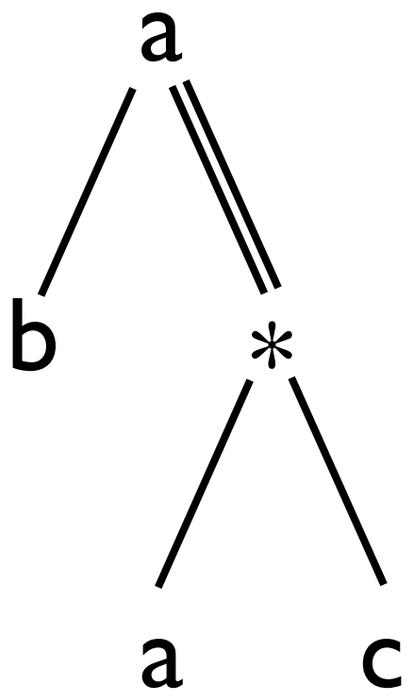


Tree Pattern

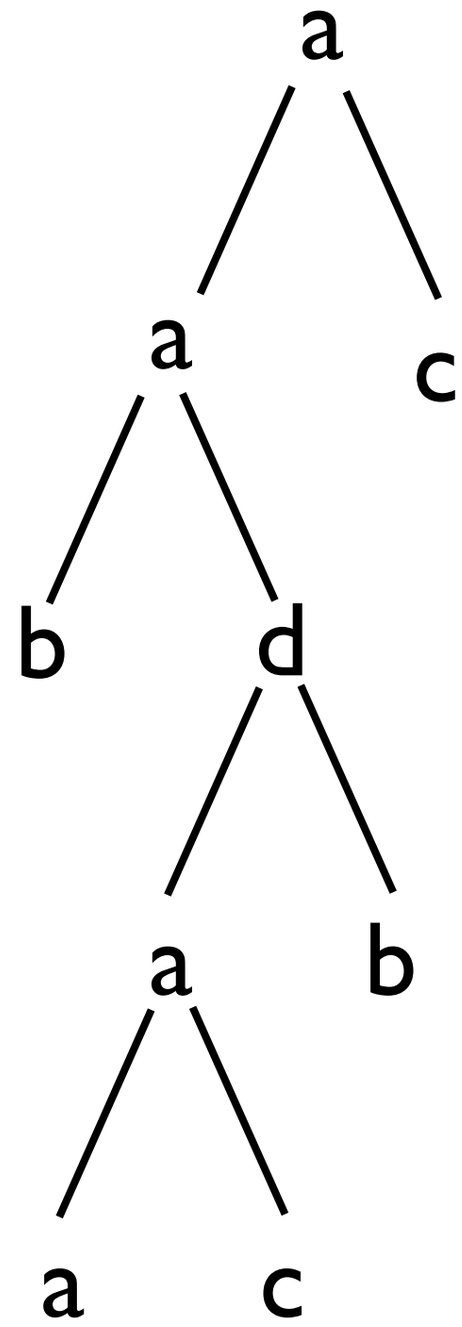




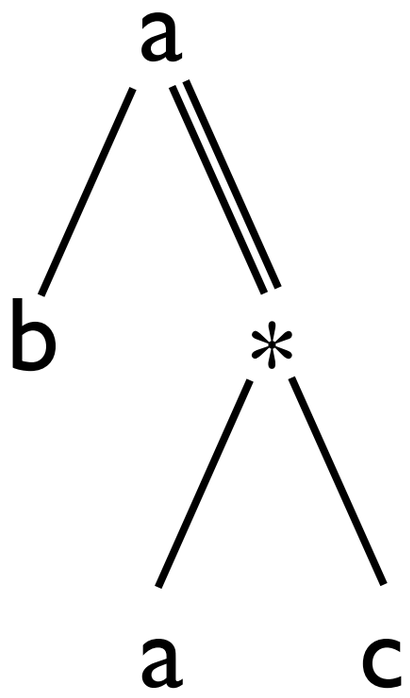
Tree Pattern

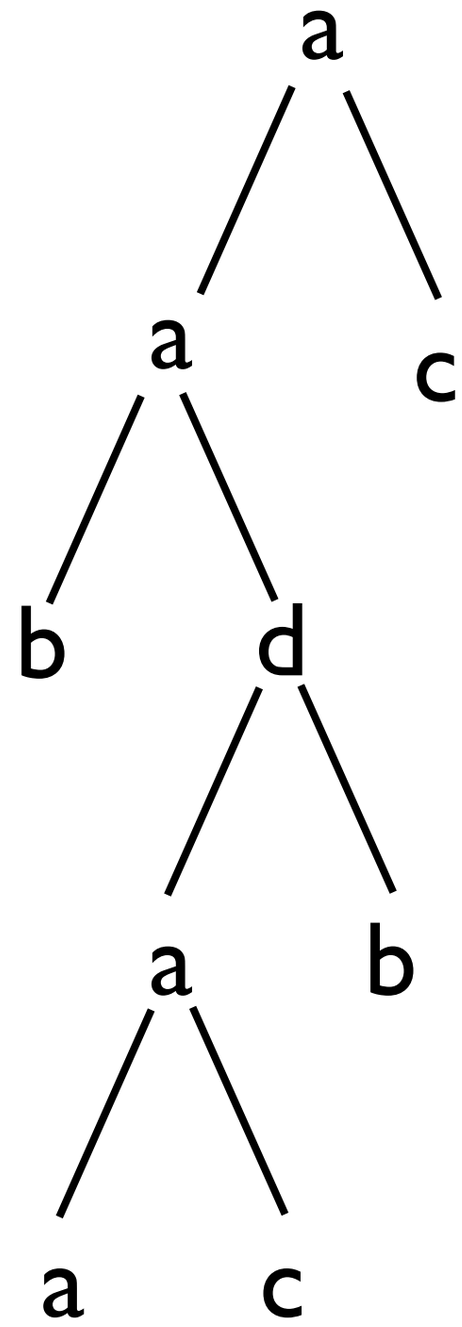


$L_w(p)$



Tree Pattern

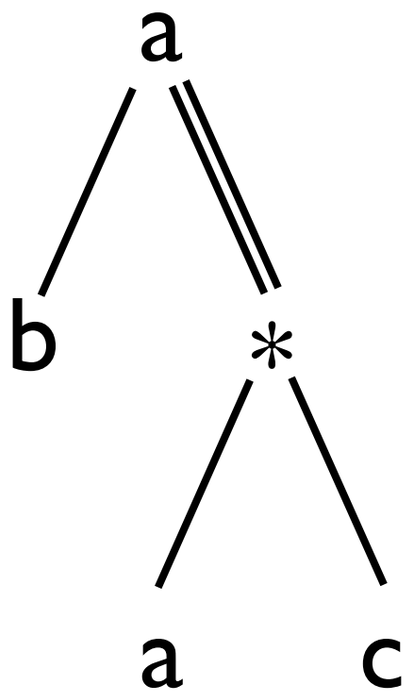


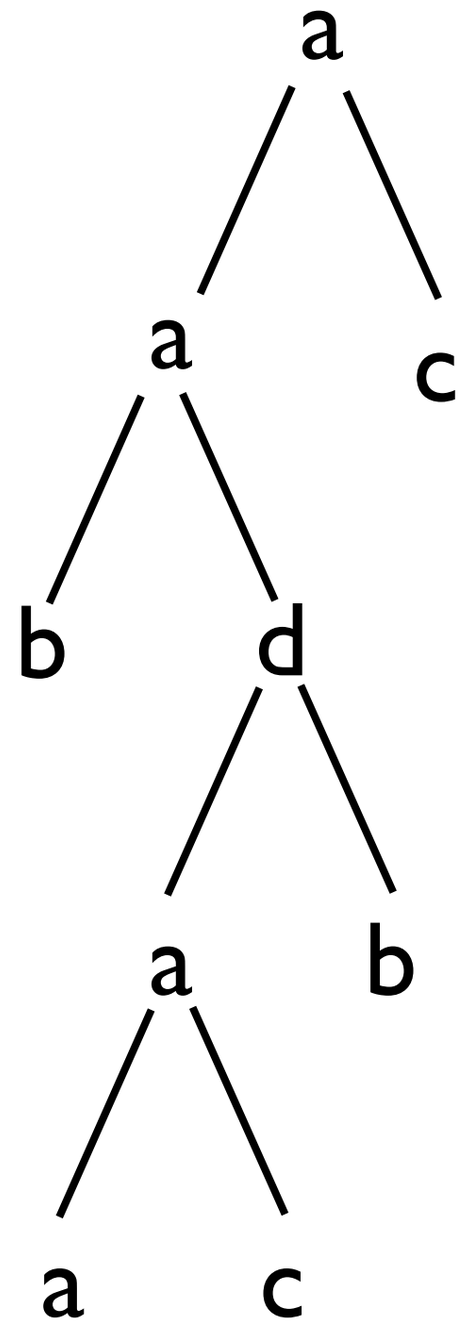


$L_w(p)$ = set of trees which embed

weakly into p

Tree Pattern





$L_s(p)$ = set of trees which embed

strongly into p

Tree Pattern Containment

Tree Pattern Containment

p includes weakly in q:

$L_w(p) \subseteq L_w(q)$

Tree Pattern Containment

p includes weakly in q:

$L_w(p) \subseteq L_w(q)$

does p enforce q?

Tree Pattern Containment

p includes weakly in q:

$L_w(p) \subseteq L_w(q)$

does p enforce q?

p includes weakly in q wrt DTD d:

$L_w(p) \cap L(d) \subseteq L_w(q)$

Tree Pattern Containment

p includes weakly in q:

$L_w(p) \subseteq L_w(q)$

does p enforce q?

defines a tree language

p includes weakly in q wrt DTD d:

$L_w(p) \cap L(d) \subseteq L_w(q)$

Tree Pattern Containment

p includes weakly in q:

$L_w(p) \subseteq L_w(q)$

does p enforce q?

p includes weakly in q wrt DTD d:

$L_w(p) \cap L(d) \subseteq L_w(q)$

Tree Pattern Containment

p includes weakly in q:

$L_w(p) \subseteq L_w(q)$

does p enforce q?

p includes weakly in q wrt DTD d:

$L_w(p) \cap L(d) \subseteq L_w(q)$

does p enforce q among trees in L(d)?

Problem

Problem

Given: two patterns p in F_1, q in F_2, (sometimes) DTD d

Problem

Given: two patterns p in F_1, q in F_2, (sometimes) DTD d

Question: is p included in q wrt schema d

Problem

Given: two patterns p in F_1, q in F_2, (sometimes) DTD d

Question: is p included in q wrt schema d

Complexity may differ depending on:

Problem

Given: two patterns p in F_1, q in F_2, (sometimes) DTD d

Question: is p included in q wrt schema d

Complexity may differ depending on:

- fragments F_1, F_2

Problem

Given: two patterns p in F_1, q in F_2, (sometimes) DTD d

Question: is p included in q wrt schema d

Complexity may differ depending on:

- fragments F_1, F_2
- weak/strong version

Problem

Given: two patterns p in F_1, q in F_2, (sometimes) DTD d

Question: is p included in q wrt schema d

Complexity may differ depending on:

- fragments F_1, F_2
- weak/strong version
- presence of DTD

Why should we care?

Why should we care?

- fundamental problem for trees

Why should we care?

- fundamental problem for trees
- tree data still with us - JSON

Why should we care?

- fundamental problem for trees
- tree data still with us - JSON
- XML with incomplete information (Barcelo et al. `10)

Why should we care?

- fundamental problem for trees
- tree data still with us - JSON
- XML with incomplete information (Barcelo et al. `10)
- (some) results transfer to graph structured data

Many cases

Fragment:

Many cases

Fragment: wildcard / no wildcard

Many cases

Fragment: wildcard / no wildcard

child / descendant / both

Many cases

Fragment: wildcard / no wildcard

child / descendant / both

brachning / no branching

Many cases

Fragment: wildcard / no wildcard

child / descendant / both

brachning / no branching

Two sides: $12^2 = 144$ options

Many cases

Fragment: wildcard / no wildcard

child / descendant / both

brachning / no branching

Two sides: $12^2 = 144$ options

Weak / strong, no DTD / DTD / fixed DTD

Many cases

Fragment: wildcard / no wildcard

child / descendant / both

brachning / no branching

Two sides: $12^2 = 144$ options

Weak / strong, no DTD / DTD / fixed DTD

Together: 144 x 6 = 864

Many cases

Fragment: wildcard / no wildcard

child / descendant / both

brachning / no branching

Two sides: $12^2 = 144$ options

Weak / strong, no DTD / DTD / fixed DTD

Together: 144 x 6 = 864

Plus satisfiability and validity

Many cases

Fragment: wildcard / no wildcard

child / descendant / both

brachning / no branching

Two sides: $12^2 = 144$ options

Weak / strong, no DTD / DTD / fixed DTD

Together: 144 x 6 = 864

Plus satisfiability and validity

Goal: uniform overview, solve hard cases,
track the PTIME borderline

What was known?

What was known?

- a lot!

What was known?

- a lot!
  - Miklau & Suciu *Containment and equivalence for a fragment of XPath* `04

What was known?

- a lot!
  - Miklau & Suciu *Containment and equivalence for a fragment of XPath* `04
  - Benedikt, Fan & Geerts *XPath satisfiability in the presence of DTD* `08

What was known?

- a lot!
  - Miklau & Suciu *Containment and equivalence for a fragment of XPath* `04
  - Benedikt, Fan & Geerts *XPath satisfiability in the presence of DTD* `08
  - Neven & Schwentick *On the complexity of XPath containment in the presence of disjuntion, DTDs and variables* `06

What was known?

- a lot!
  - Miklau & Suciu *Containment and equivalence for a fragment of XPath* `04
  - Benedikt, Fan & Geerts *XPath satisfiability in the presence of DTD* `08
  - Neven & Schwentick *On the complexity of XPath containment in the presence of disjuntion, DTDs and variables* `06
- inclusion with DTD: EXPTIME-comp.

What was known?

- a lot!
  - Miklau & Suciu *Containment and equivalence for a fragment of XPath* `04
  - Benedikt, Fan & Geerts *XPath satisfiability in the presence of DTD* `08
  - Neven & Schwentick *On the complexity of XPath containment in the presence of disjuntion, DTDs and variables* `06
- inclusion with DTD: EXPTIME-comp.
- inclusion without DTD: coNP-comp.

What was known?

- a lot!
  - Miklau & Suciu *Containment and equivalence for a fragment of XPath* `04
  - Benedikt, Fan & Geerts *XPath satisfiability in the presence of DTD* `08
  - Neven & Schwentick *On the complexity of XPath containment in the presence of disjuntion, DTDs and variables* `06
- inclusion with DTD: EXPTIME-comp.
- inclusion without DTD: coNP-comp.
- a lot of mess in the landscape

Our contribution

Our contribution

- solved most of the cases

Our contribution

- solved most of the cases
- in not so many theorems

Our contribution

- solved most of the cases
- in not so many theorems
- only around five crucial ideas

Our contribution

- solved most of the cases
- in not so many theorems
- only around five crucial ideas
- some open problems solved

Our contribution

- solved most of the cases
- in not so many theorems
- only around five crucial ideas
- some open problems solved
- uniform overview

This talk

This talk

- a few most interesting results

This talk

- a few most interesting results
- a few observations after looking at the table

This talk

- a few most interesting results
- a few observations after looking at the table
- open problem

Notation

Notation

- TPQ(/, //, *) = TPQ - all tree pattern queries

Notation

- TPQ(/, //, *) = TPQ - all tree pattern queries
- TPQ(//, *) - child edge disallowed

Notation

- TPQ(/, //, *) = TPQ - all tree pattern queries
- TPQ(//, *) - child edge disallowed
- PQ(/, //, *) = PQ - all Path Queries

Notation

- TPQ(/, //, *) = TPQ - all tree pattern queries
- TPQ(//, *) - child edge disallowed
- PQ(/, //, *) = PQ - all Path Queries
- PQ(/, //) - no wildcard

Interesting results (negative)

Interesting results (negative)

Inclusion of TPQ(/) in PQ(/) with fixed DTD is coNP-hard

Interesting results (negative)

Inclusion of TPQ(/) in PQ(/) with fixed DTD is coNP-hard

Message: trees on the left are hard

Interesting results (negative)

Inclusion of TPQ(/) in PQ(/) with fixed DTD is coNP-hard

Message: trees on the left are hard

Technique: reduction from 3-partition

Interesting results (negative)

Inclusion of TPQ(/) in PQ(/) with fixed DTD is coNP-hard

Message: trees on the left are hard

Technique: reduction from 3-partition

Weak inclusion of PQ(/) in PQ(/,*) with fixed DTD is EXPTIME-hard

Interesting results (negative)

Inclusion of TPQ(/) in PQ(/) with fixed DTD is coNP-hard

Message: trees on the left are hard

Technique: reduction from 3-partition

Weak inclusion of PQ(/) in PQ(/,*) with fixed DTD is EXPTIME-hard

Message: a*****b on the right provides all the hardness

Interesting results (negative)

Inclusion of TPQ(/) in PQ(/) with fixed DTD is coNP-hard

Message: trees on the left are hard

Technique: reduction from 3-partition

Weak inclusion of PQ(/) in PQ(/,*) with fixed DTD is EXPTIME-hard

Message: a*****b on the right provides all the hardness

Technique: new triomino-tiling method

Triomino tiling

Triomino tiling

Input:

Triomino tiling

Input: set of tiles T

Triomino tiling

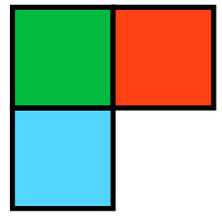
Input: set of tiles T

set of allowed triominos in T^3

Triomino tiling

Input: set of tiles T

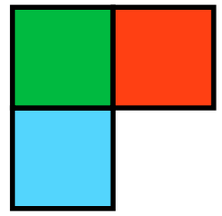
set of allowed triominos in T^3



Triomino tiling

Input: set of tiles T

set of allowed triominos in T^3

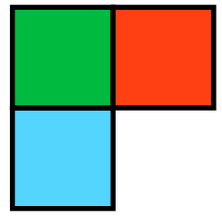


initial row, final tile

Triomino tiling

Input: set of tiles T

set of allowed triominos in T^3



initial row, final tile

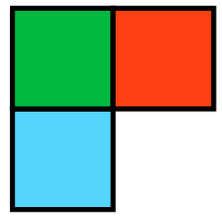
Question:

Triomino tiling

Input: set of tiles T

set of allowed triominos in T^3

initial row, final tile

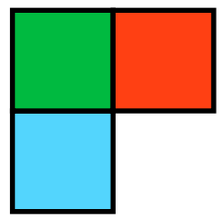


Question: can we fill a rectangle?

Triomino tiling

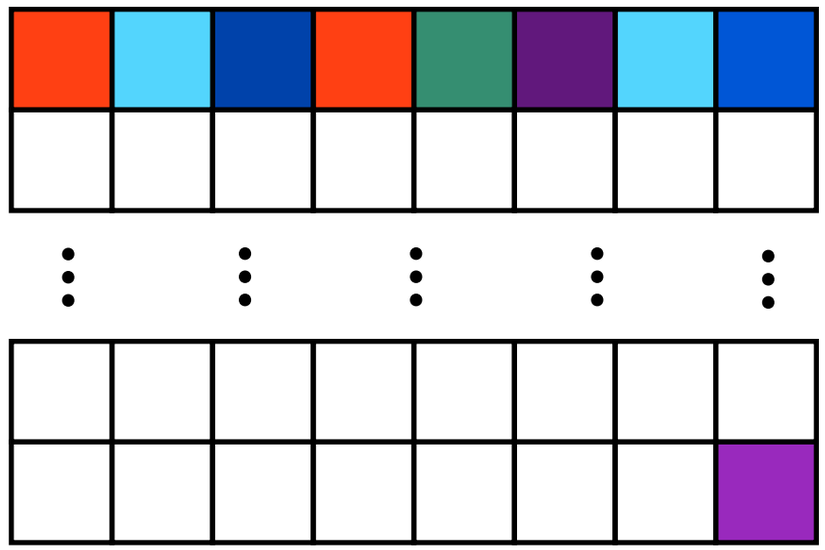
Input: set of tiles T

set of allowed triominos in T^3



initial row, final tile

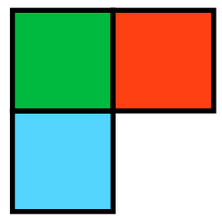
Question: can we fill a rectangle?



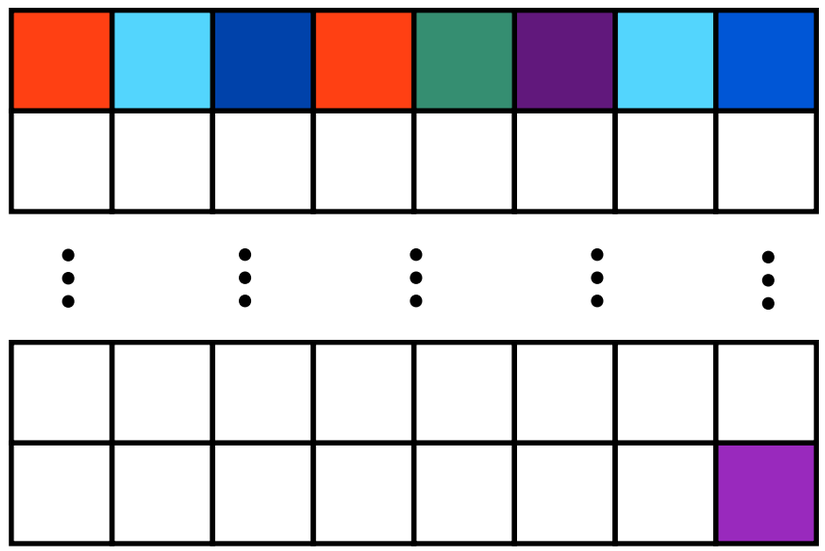
Triomino tiling

Input: set of tiles T

set of allowed triominos in T^3



initial row, final tile



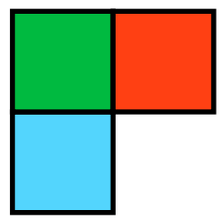
Question: can we fill a rectangle?

PSPACE-complete

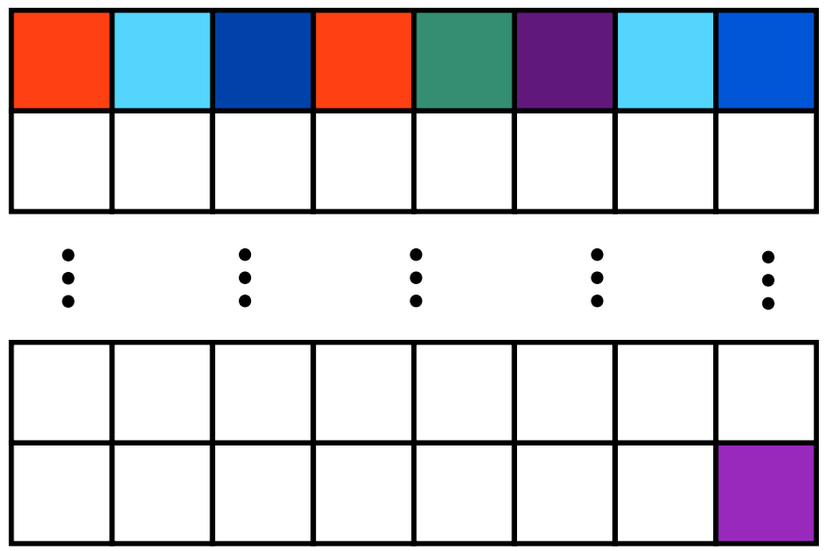
Triomino tiling

Input: set of tiles T

set of allowed triominos in T^3



initial row, final tile



Question: can we fill a rectangle?

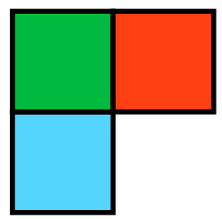
PSPACE-complete

game: EXPTIME-complete

Triomino tiling

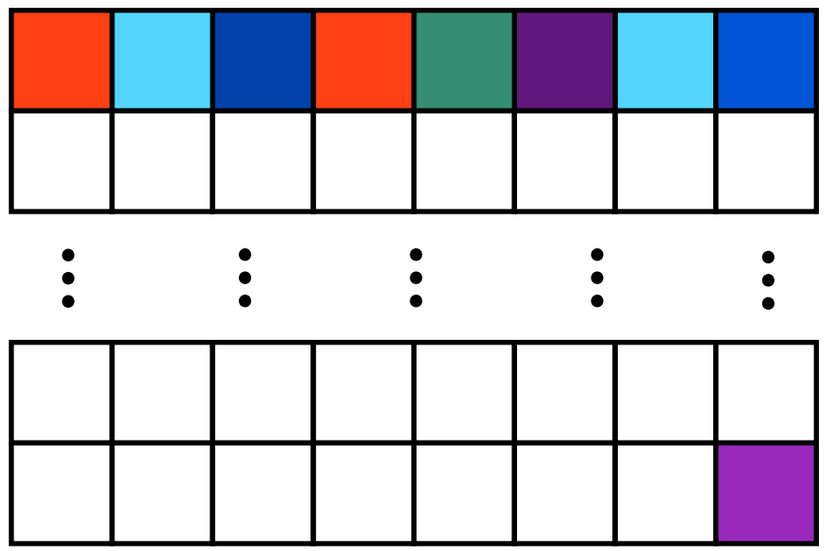
Input: set of tiles T

set of allowed triominos in T^3



initial row, final tile

Question: can we fill a rectangle?



PSPACE-complete

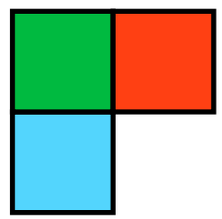
game: EXPTIME-complete

Left pattern: initial row

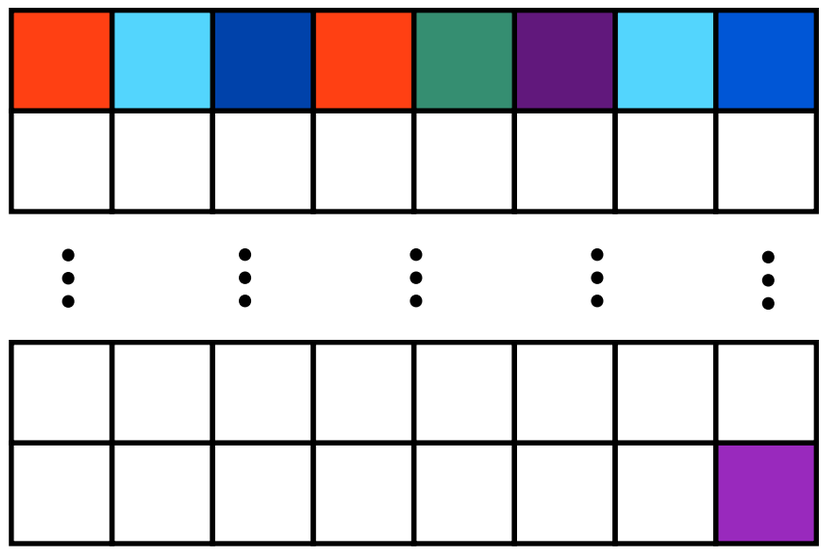
Triomino tiling

Input: set of tiles T

set of allowed triominos in T^3



initial row, final tile



Question: can we fill a rectangle?

PSPACE-complete

game: EXPTIME-complete

Left pattern: initial row

Right pattern and DTD: implement forbiden triominos via forbiding a******b

Interesting results (positive)

Interesting results (positive)

Inclusion of PQ in TPQ without DTD is in PTIME

Interesting results (positive)

Inclusion of PQ in TPQ without DTD is in PTIME

Message: without DTD words on the left are easy

Interesting results (positive)

Inclusion of PQ in TPQ without DTD is in PTIME

Message: without DTD words on the left are easy

Technique: involved combinatorics

Interesting results (positive)

Inclusion of PQ in TPQ without DTD is in PTIME

Message: without DTD words on the left are easy

Technique: involved combinatorics

open for a decade - now complete classification without DTD

Interesting results (positive)

Inclusion of PQ in TPQ without DTD is in PTIME

Message: without DTD words on the left are easy

Technique: involved combinatorics

Interesting results (positive)

Inclusion of PQ in TPQ without DTD is in PTIME

Message: without DTD words on the left are easy

Technique: involved combinatorics

Strong inclusion of PQ in (TPQ minus sth) with DTD is in PTIME

Interesting results (positive)

Inclusion of PQ in TPQ without DTD is in
PTIME

Message: without DTD words on the left are easy

Technique: involved combinatorics

Strong inclusion of PQ in (TPQ minus sth) with DTD is in
PTIME

Technique: easy - construct automaton for
trees not embedding the right pattern

Interesting results (positive)

Inclusion of PQ in TPQ without DTD is in
PTIME

Message: without DTD words on the left are easy

Technique: involved combinatorics

Strong inclusion of PQ in (TPQ minus sth) with DTD is in
PTIME

Technique: easy - construct automaton for
trees not embedding the right pattern

Message: such a uniform technique works in all easy cases

Intriguing observations

Intriguing observations

- results for fixed DTD and DTD a part of an input does not differ

Intriguing observations

- results for fixed DTD and DTD a part of an input does not differ
- only three possible answers: PTIME, coNP-complete, EXPTIME-complete

Intriguing observations

- results for fixed DTD and DTD a part of an input does not differ
- only three possible answers: PTIME, coNP-complete, EXPTIME-complete
- for simple reasons

Open question

Open question

- inclusion of PQ in TPQ(/,//) with DTD

Open question

- inclusion of PQ in TPQ(/,//) with DTD
- PTIME or not?

Open question

- inclusion of PQ in TPQ(/,//) with DTD
- PTIME or not?
- probably hard to solve

Thank you!