

The (Almost) Complete Guide to Tree Pattern Containment

Wojciech Czerwiński

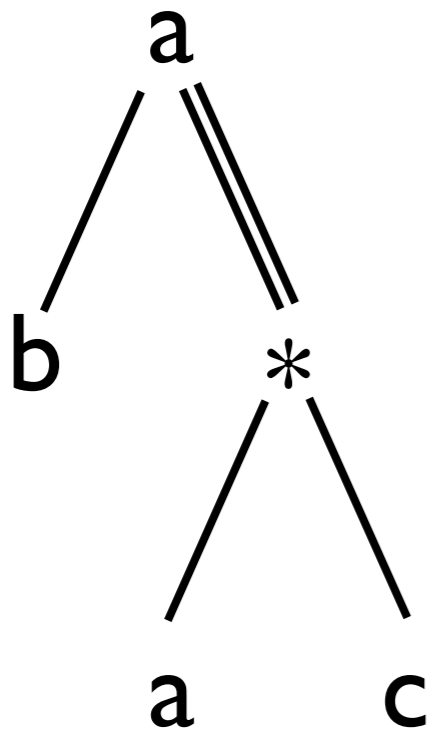
Wim Martens

Paweł Parys

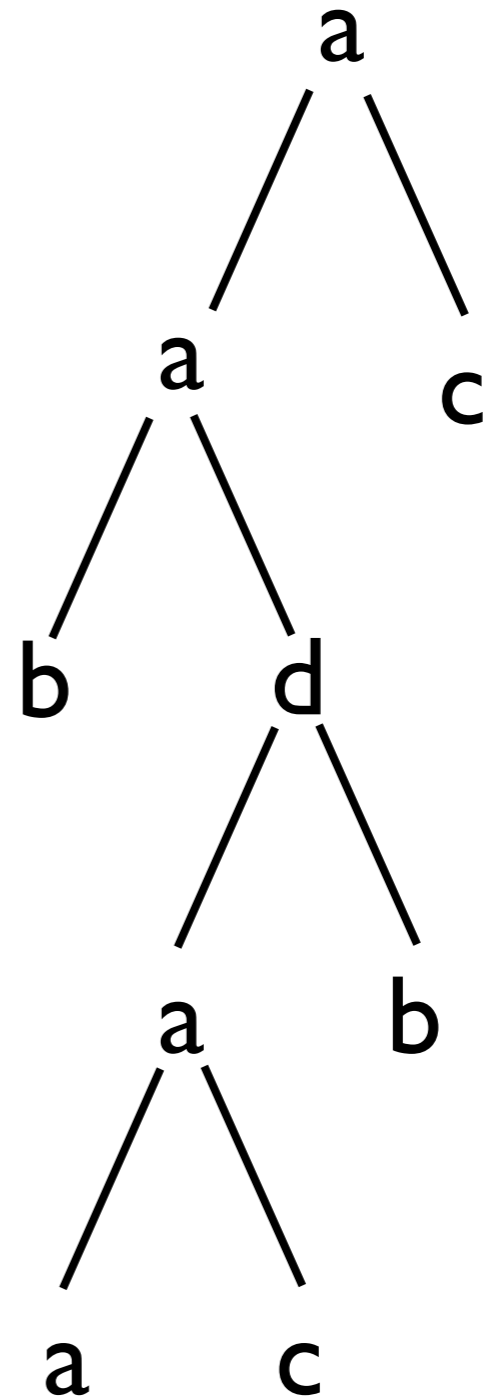
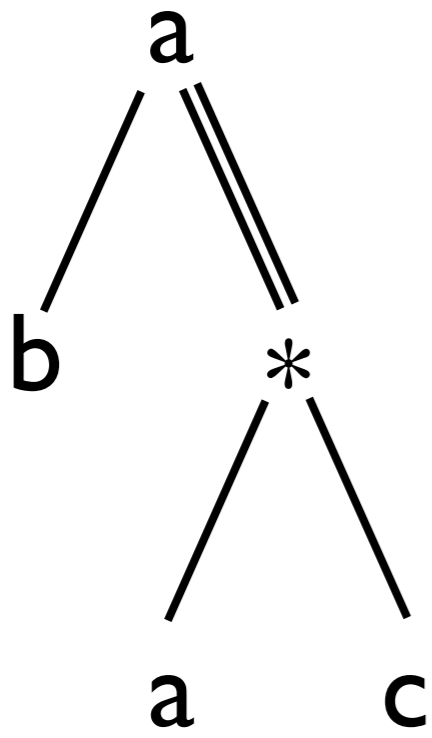
Marcin Przybyłko

Tree Pattern

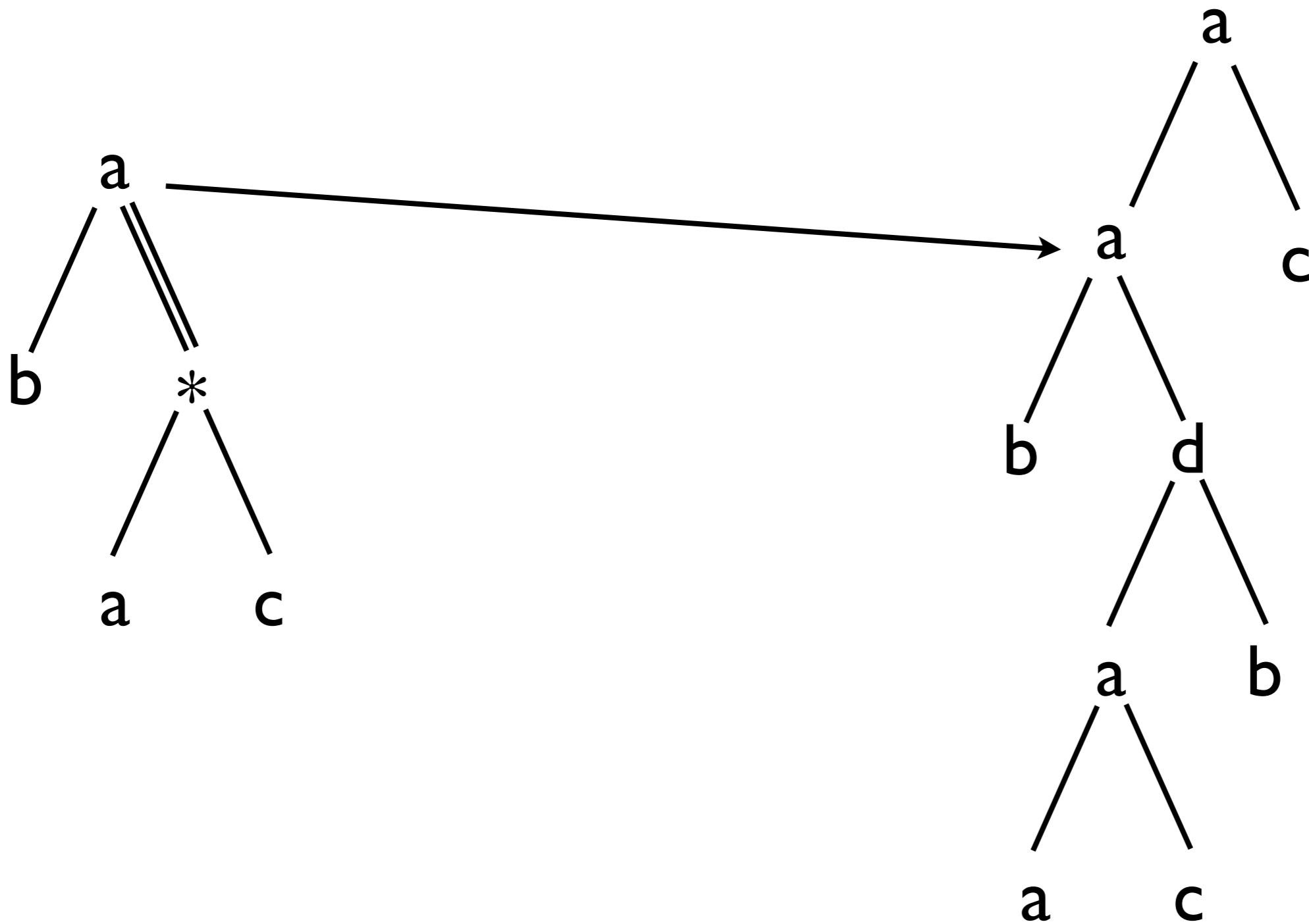
Tree Pattern



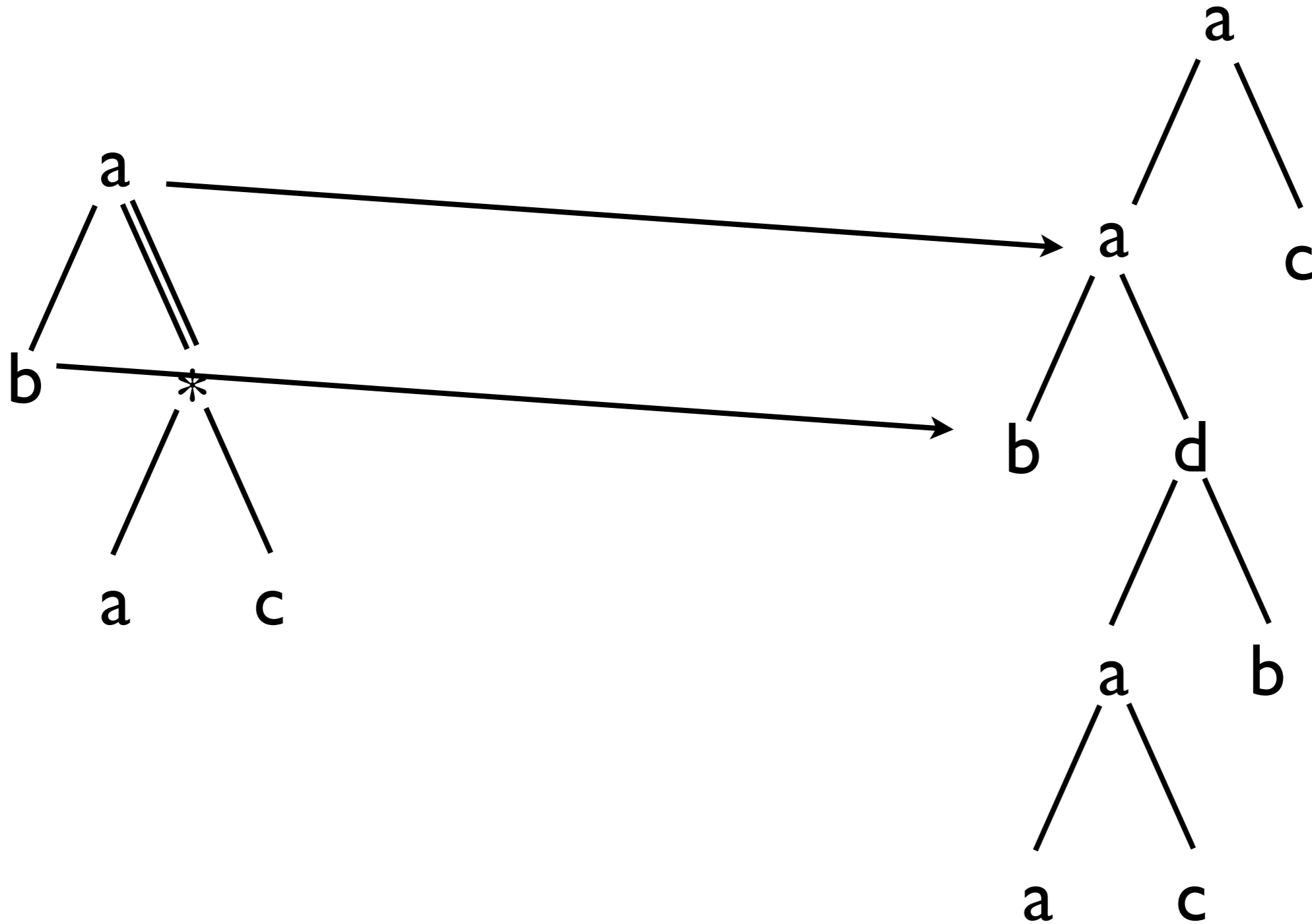
Tree Pattern



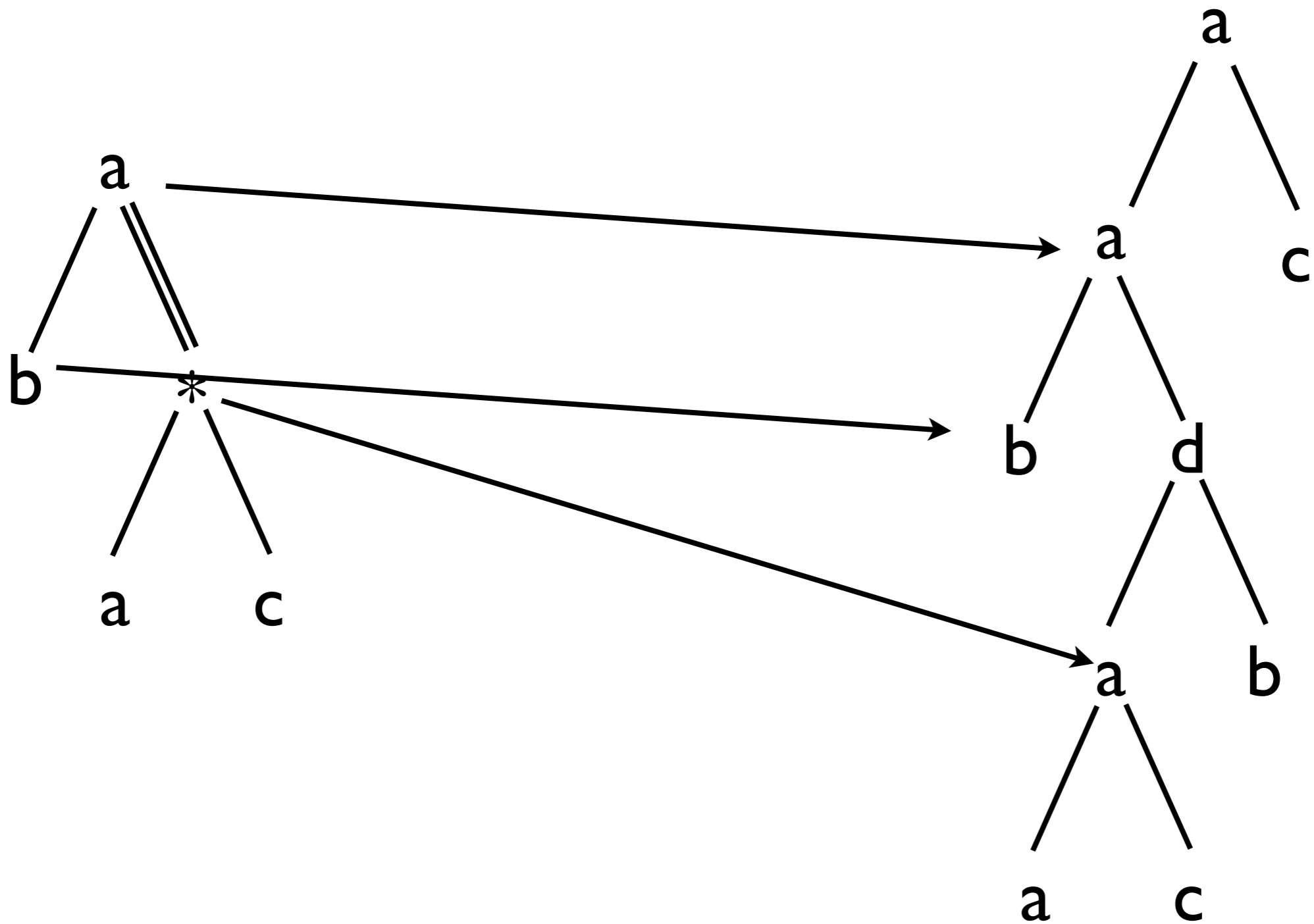
Tree Pattern



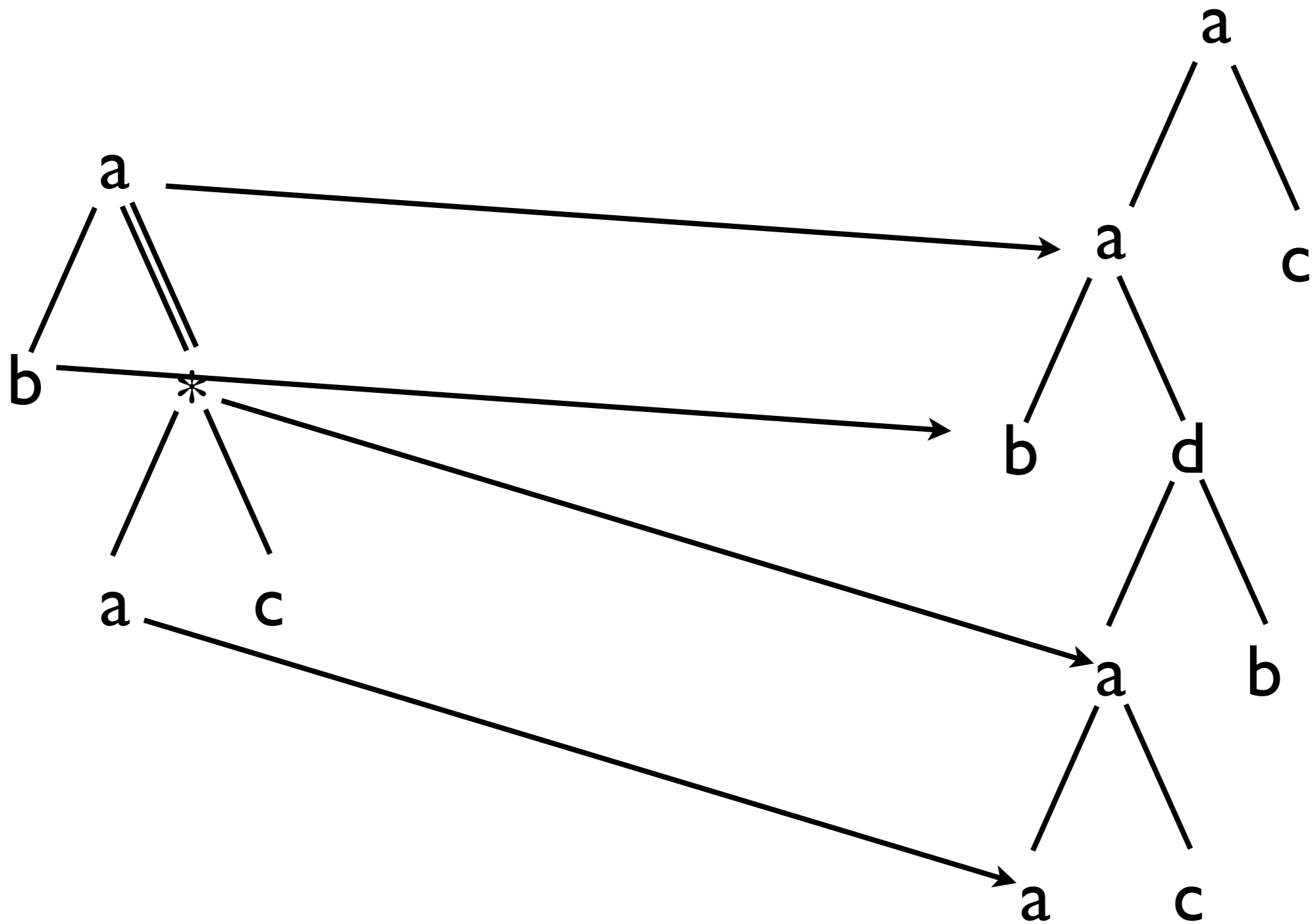
Tree Pattern



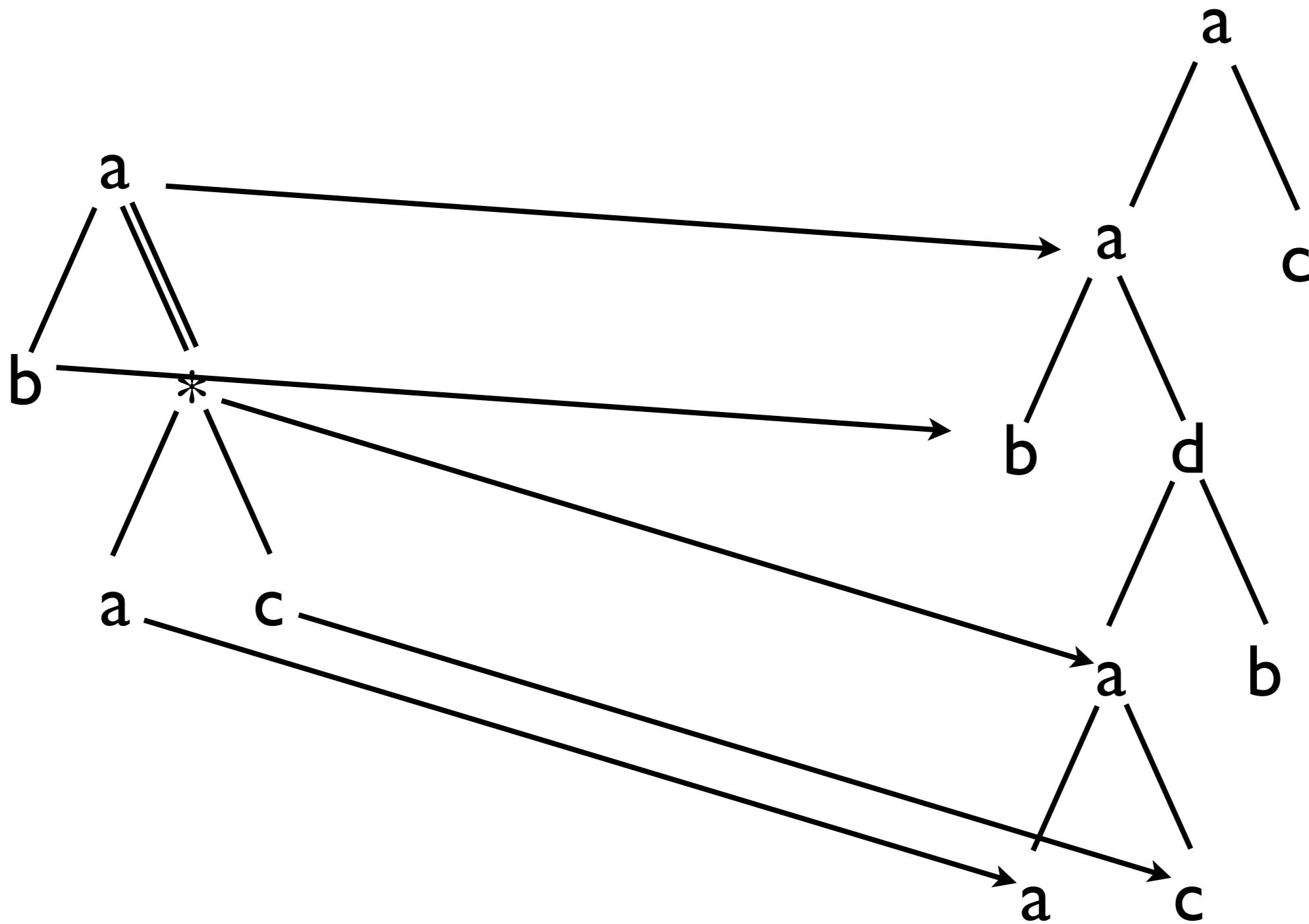
Tree Pattern



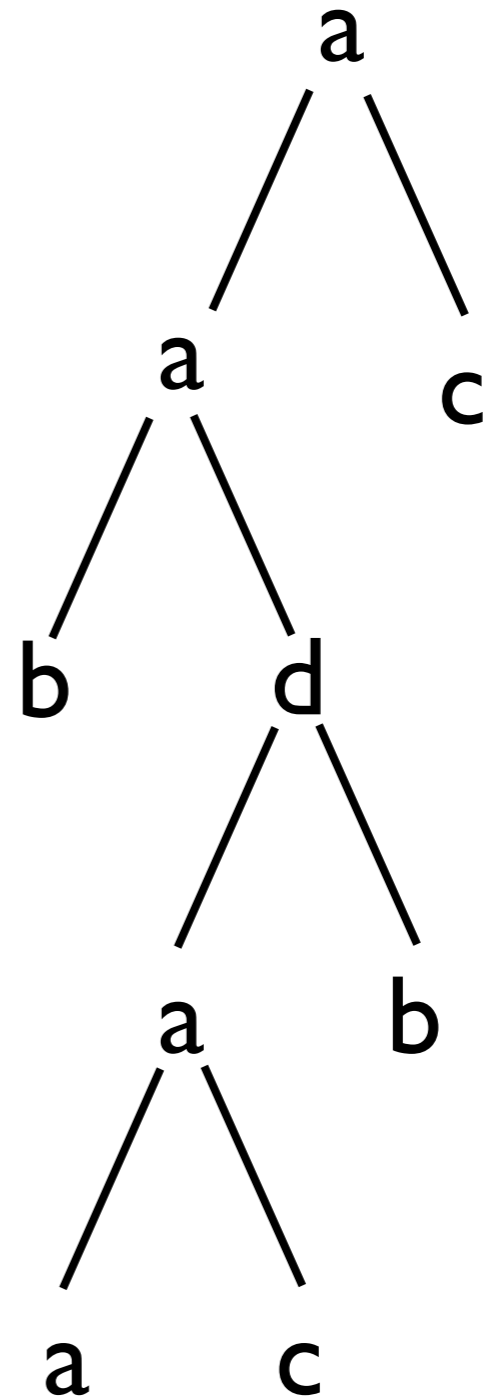
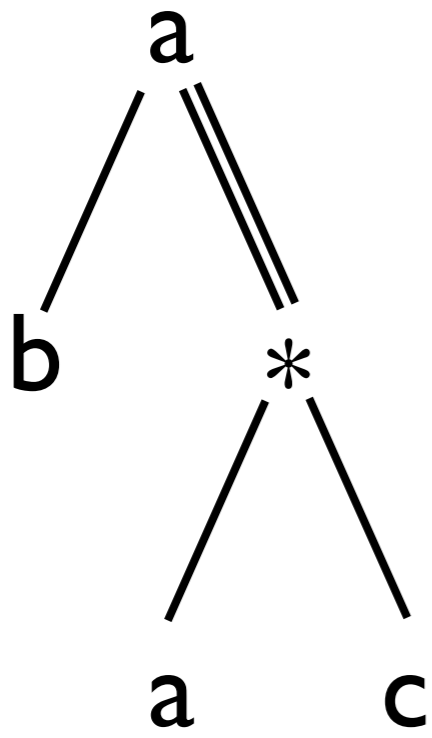
Tree Pattern



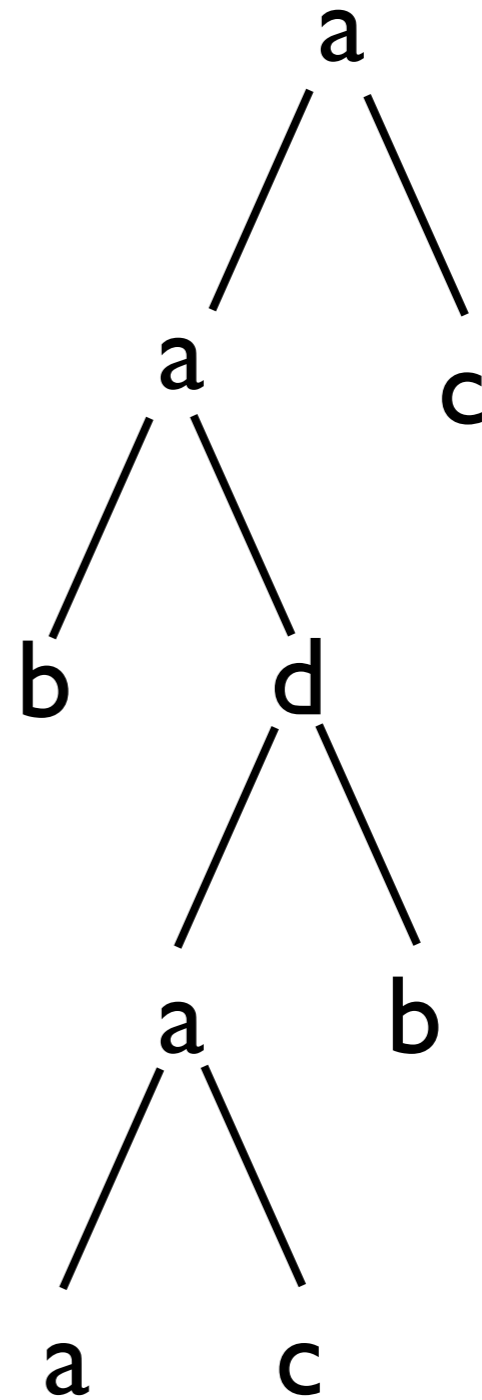
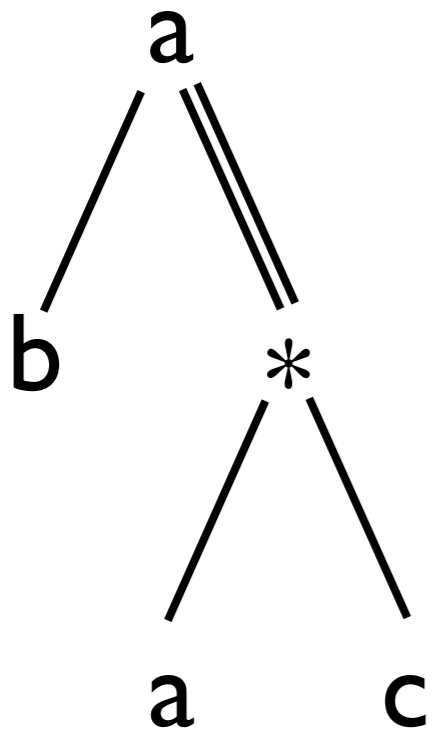
Tree Pattern



Tree Pattern

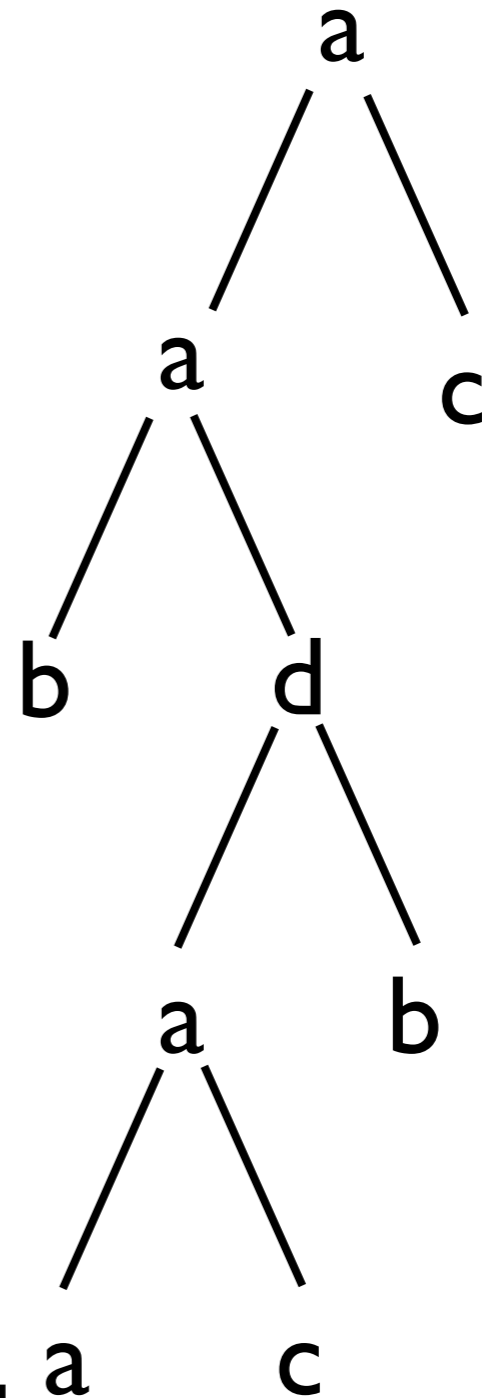
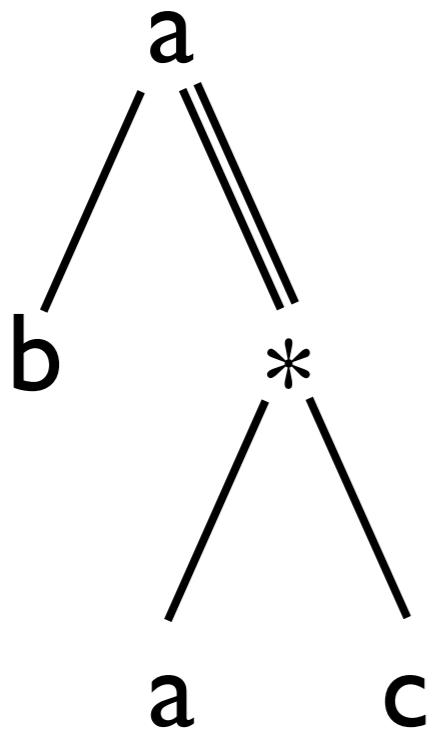


Tree Pattern



L(p)

Tree Pattern



$L(p)$ = set of trees into which p embeds

Tree Pattern Containment

Tree Pattern Containment

p includes in q :

$$L(p) \subseteq L(q)$$

Tree Pattern Containment

p includes in q:

$$L(p) \subseteq L(q)$$

does p enforce q?

Tree Pattern Containment

p includes in q:

$$L(p) \subseteq L(q)$$

does p enforce q?

p includes in q wrt DTD d:

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

Tree Pattern Containment

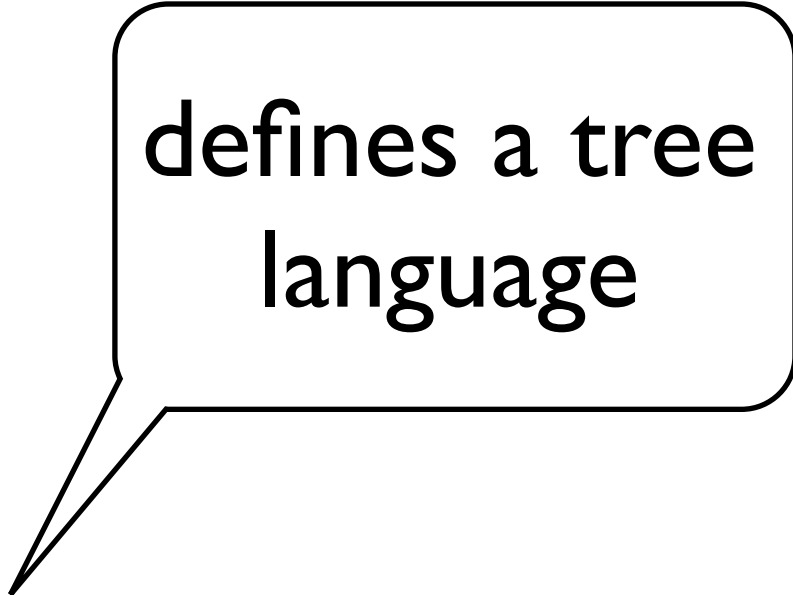
p includes in q:

$$L(p) \subseteq L(q)$$

does p enforce q?

p includes in q wrt DTD d:

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



defines a tree language

Tree Pattern Containment

p includes in q:

$$L(p) \subseteq L(q)$$

does p enforce q?

p includes in q wrt DTD d:

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

Tree Pattern Containment

p includes in q:

$$L(p) \subseteq L(q)$$

does p enforce q?

p includes in q wrt DTD d:

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

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 a schema d

Problem

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

Question: is p included in q wrt a 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 a 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 a schema d

Complexity may differ depending on:

- fragments F_1, F_2
- presence of DTD

Problem

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

Question: is p included in q wrt a schema d

Complexity may differ depending on:

- fragments F_1, F_2
- presence of DTD

Motivation: tree structured data - XML, JSON, etc.

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
branching / no branching

Many cases

Fragment: wildcard / no wildcard
child / descendant / both
branching / no branching

Two sides: $12^2 = 144$ options

Many cases

Fragment: wildcard / no wildcard
child / descendant / both
branching / no branching

Two sides: $12^2 = 144$ options

No DTD / DTD / fixed DTD

Many cases

Fragment: wildcard / no wildcard
child / descendant / both
branching / no branching

Two sides: $12^2 = 144$ options

No DTD / DTD / fixed DTD

Together: $144 \times 3 = 432$

Many cases

Fragment: wildcard / no wildcard
child / descendant / both
branching / no branching

Two sides: $12^2 = 144$ options

No DTD / DTD / fixed DTD

Together: $144 \times 3 = 432$

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

What was known?

What was known?

- a lot!

What was known?

- a lot!
- inclusion with DTD: EXPTIME-comp.

What was known?

- a lot!
- inclusion with DTD: EXPTIME-comp.
- inclusion without DTD: coNP-comp.

What was known?

- a lot!
- 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
- **uniform** overview

Our contribution

- solved most of the cases
- in **not so many** theorems
- **uniform** overview
- some **open** problems solved

This talk

This talk

- three interesting results

This talk

- three interesting results
- two observations after looking at the table

Notation

Notation

- $\text{TPQ}(/, //, *) = \text{TPQ}$ - all tree pattern queries

Notation

- $\text{TPQ}(/, //, *) = \text{TPQ}$ - all tree pattern queries
- $\text{TPQ}(//, *)$ - 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**

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**

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**

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

Interesting result (positive)

Interesting result (positive)

Inclusion of PQ in TPQ without DTD is in
PTIME

Interesting result (positive)

Inclusion of PQ in TPQ without DTD is in
PTIME

Message: **without DTD words on the left are easy**

Interesting result (positive)

Inclusion of PQ in TPQ without DTD is in
PTIME

Message: **without DTD words on the left are easy**

open for a decade

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

Thank you!