

Data Trees and Patterns

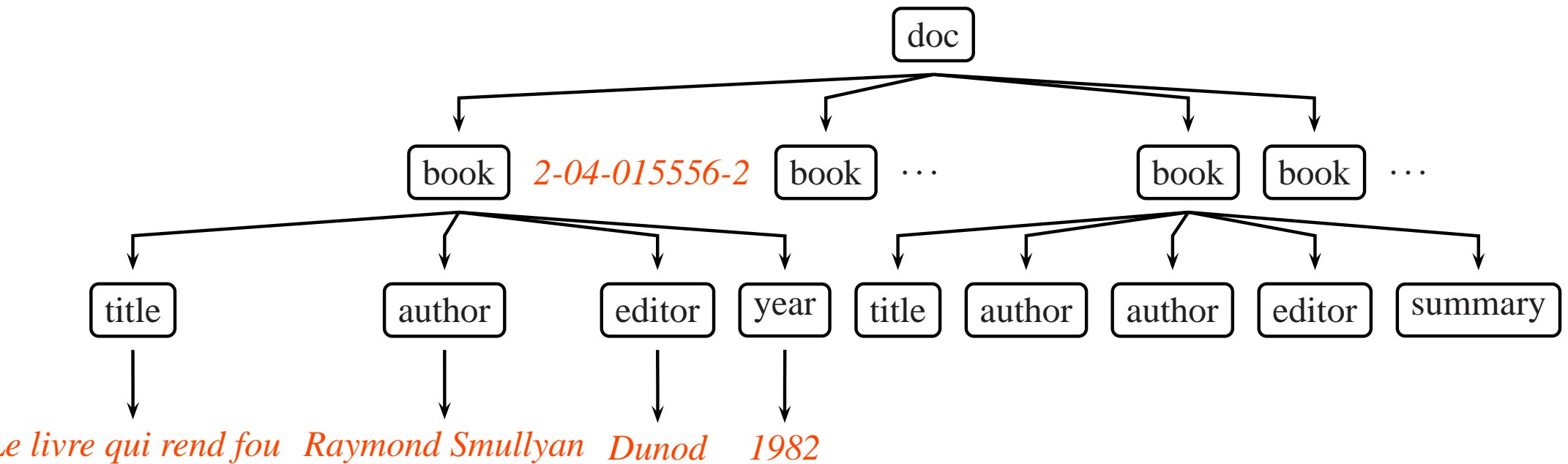
Claire David

LIAFA, Univ. Paris 7 & CNRS

INRIA-Futurs, Orsay

What is a XML Document ?

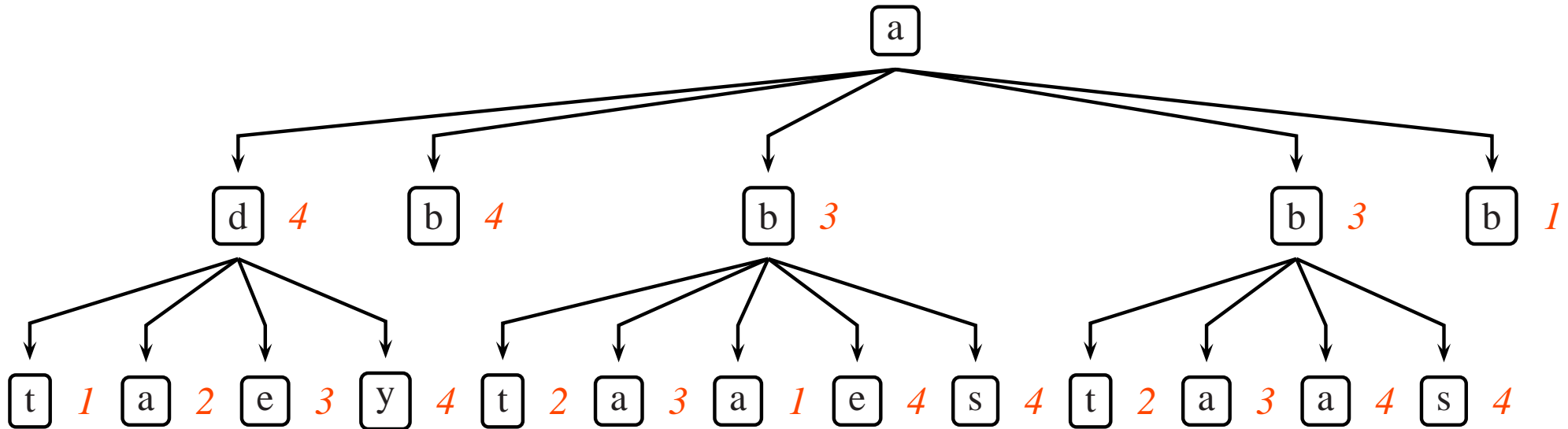
Unranked, unordered tree, whom nodes are labeled with a letter from a **finite alphabet** (tag)...



... and a **data value** from a potentially **infinite** domain (text, attribute value , reference...).

The Data Tree model.

- Unordered, unranked tree.
- Finite alphabet Σ : letters.
- Infinite domain D : data values.



Data values as an **equivalence relation** on the nodes.

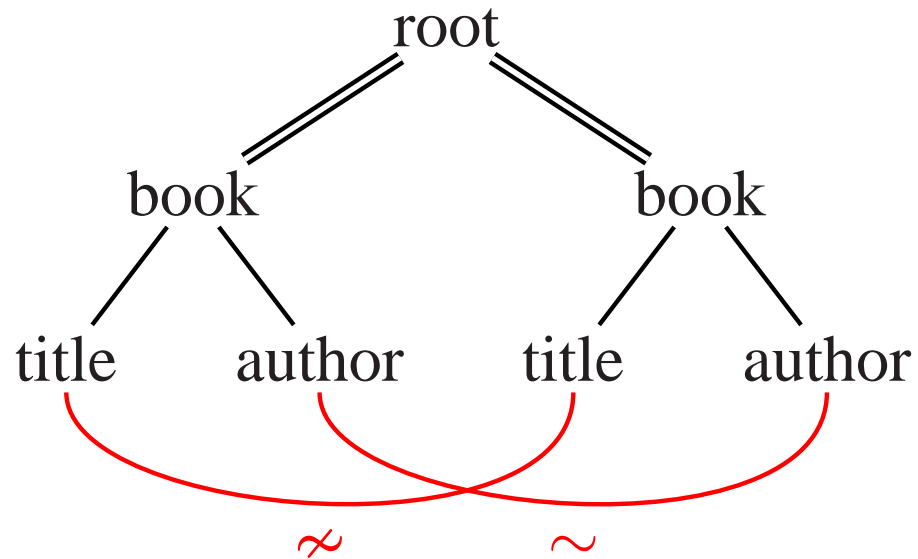
DTDs.

Constrains the finite shape of an XML document.

```
<?xml version="1.0" encoding="ISO-8859-1"?>
<!ELEMENT doc (book*)>
<!ELEMENT book (title,author*,year,editor,summary?)>
<!ATTLIST book isbn (ID)>
<!ELEMENT title (#PCDATA)>
<!ELEMENT author (#PCDATA)>
<!ELEMENT year (#PCDATA)>
<!ELEMENT editor (#PCDATA)>
<!ELEMENT summary (#PCDATA)>
```

Essentially regular tree language

Patterns : why ? how ?

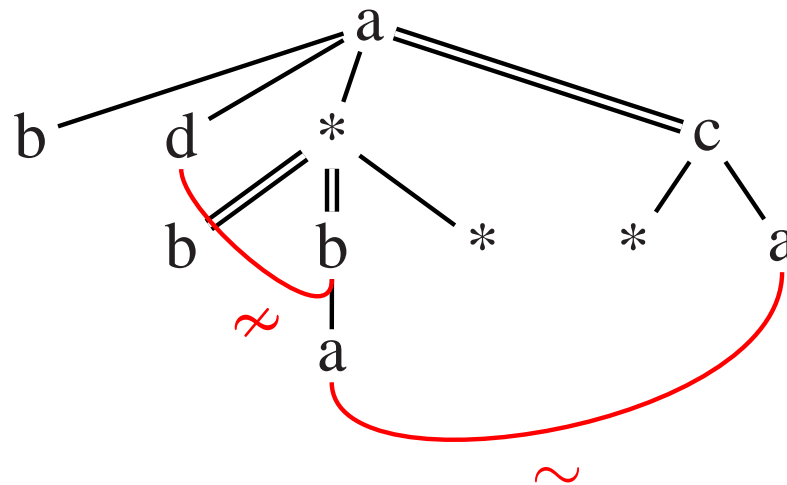


Does there exist two books from the same author ?

Goal : specify some properties on the tree about structure and data value.

Patterns : More formally

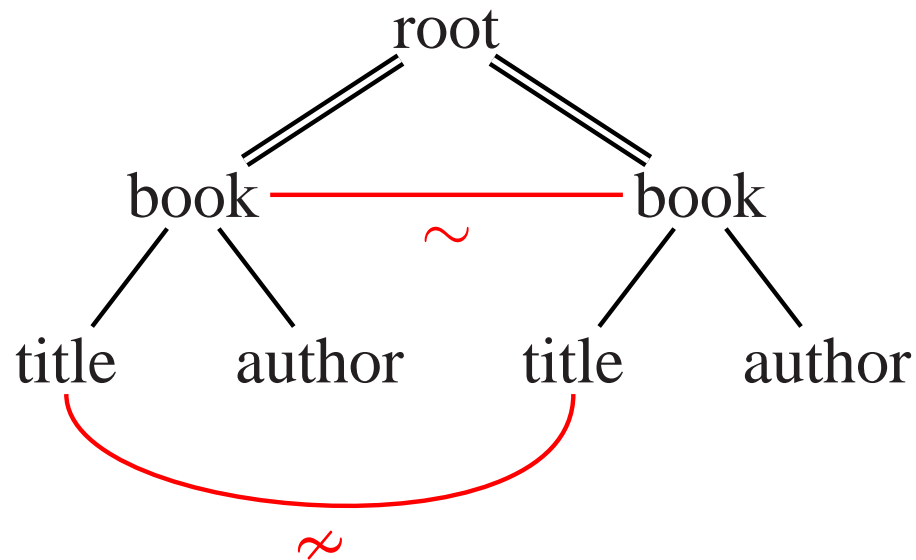
- Unordered, unranked tree
- Nodes labeled with letter or wildcard
- Child or descendant relation between nodes.
- Data constraint relations \sim or \approx between nodes



- Injective mapping.
- Preserves descendent, non-descendent relations.

Patterns : Other example

Forbidden pattern.



We want to build more complicated properties.

Consider Boolean combinaison of patterns (\vee, \wedge, \neg) .

Problems

- **Model-checking/Evaluation** : Given a data tree and a boolean combinaison of patterns, does the tree satisfies the pattern formula ?
- **Satisfiability** : Given a DTD and a boolean combinaison of patterns, does there exist a data a tree that satisfies the DTD and the pattern formula ?

To sum up...

We have :

- Data Tree models.
- DTDs.
- Data Tree Patterns.

We want to solve :

- Model checking problem.
- Satisfiability problem.

General Results : $CB(\sim, <, +1)$

– Model Checking is DP-complete.

DP is a complexity class between NP and EXPTIME

– Satisfiability is undecidable.

Fragments

We can restrict patterns :

- $CB(\sim, <)$ order fragment.
- $CB(\sim, +1)$ successor fragment.

We can restrict boolean combinaison :

- $CB^+(\sim, <, +1)$ positive fragment.
- $CB^-(\sim, <, +1)$ negative fragment.

Results for fragments

Fragments	Model-Checking	Satisfiability
$CB(\sim, <, +1)$	DP-c	Undecidable
$CB(\sim, +1)$	DP-c	2EXPTIME-c
$CB(\sim, <)$	DP-c	NEXPTIME-c
$CB^-(\sim, <, +1)$	coNP-c	Undecidable
$CB^+(\sim, <, +1)$	NP-c	NP-hard
one pattern	NP	P TIME

$$NP \cup \text{co-NP} \subseteq DP \subseteq \Sigma_2 \subseteq PSPACE$$

Satisfiability with bounded depth

Given an **integer** k , a DTD and a boolean combination of patterns, does there exist a **tree of depth less than k** that satisfies the DTD and the formula ?

k is either fixed or part of the input.

– **k fixed :**

Every fragment is decidable, Σ_2 -complete.

– **k in the input :**

Every fragment is decidable, NEXPTIME-complete.

Data word case

Fragments	MC	Sat
$CB(\sim, <, +1)$	DP-c	Undecidable
$wCB(\sim, +1)$	PSPACE	PSPACE-c
$wCB(\sim, <)$	DP-c	Σ_2 -c
$CB^-(\sim, <, +1)$	coNP-c	Undecidable
$wCB^+(\sim, <, +1)$	NP-c	NP-c
one pattern	NP-c	PSPACE

Summary

Fragments	MC	Sat	Bounded Depth Sat
$CB(\sim, <, +1)$	DP-c	Undecidable	NEXPTIME-c
$CB(\sim, +1)$	DP-c	2EXPTIME-c	NEXPTIME-c
Data Word	P TIME	P SPACE-c	
$CB(\sim, <)$	DP-c	NEXPTIME-c	NEXPTIME-c
Data Word	DP -c	Σ_2 -c	
$CB^-(\sim, <, +1)$	coNP-c	Undecidable	NEXPTIME-c
$CB^+(\sim, <, +1)$	NP-c	NP-hard	NP-hard
Data Word	NP -c	NP -c	

$$NP \cup \text{co-NP} \subseteq DP \subseteq \Sigma_2 \subseteq PSPACE$$

Discussion

A natural model of properties with data value constraints.

Powerfull formalism (vs XPath) :

- data value comparaison
- negation

Model-Checking :

- NP-complete for one pattern - XPath is polynomial.
- DP-complete in general.

Satisfiability :

- undecidable in general - XPath too.
- comparable for fragments.
- decidability for bounded depth restriction.

Perspectives

Query language : We have result for boolean formula. How can we extend this as a real query language ?

Expressivity/Complexity...

Active document : Can we use patterns to specify evolution of active document.

Thanks for your attention!