

Lógicas modales para especificación y consulta de datos semi-estructurados

Introducción. Los lenguajes de especificación de documentos XML se usan para definir un conjunto de documentos aceptados (por ej: XHTML Transitional, SVG, Jabber, etc.)



En los lenguajes más populares como DTD y XML Schema, las posibles secuencias de hijos de un nodo son descritas, básicamente, por expresiones regulares. Esto es suficiente en muchos casos pero incluso condiciones simples como

$$\#autor \leq 2 \times \#titulo$$

no pueden ser expresadas.

```
<biblioteca>
  <categoria nombre="Francia">
    <libro>
      <titulo idioma="Español">El Principito</titulo>
    </libro>
  </categoria>
</biblioteca>
```

Fig. 1: Documento XML

Los lenguajes de consulta XML se utilizan para seleccionar y obtener datos de dichos documentos



XPath, por ejemplo, resulta *indecidable* si se utiliza en toda su potencia.

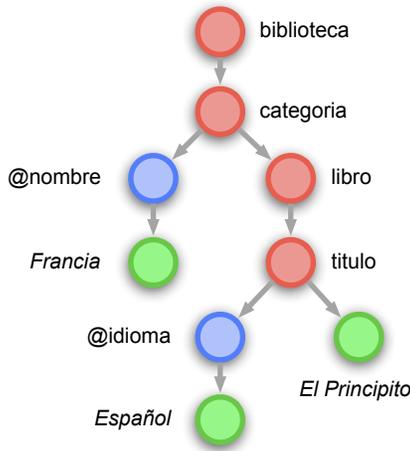


Fig. 2: Posible representación arbórea del documento de la Fig. 1

Antecedentes. En [1] Demri y Lugiez presentan la lógica EML que extiende a la lógica modal básica con restricciones aritméticas (aritmética de Presburger) y restricciones de lenguajes regulares sobre los hijos de un nodo. Aplicada sobre modelos con forma de árboles, esta lógica formaliza una manera de especificar y consultar datos semi-estructurados, en particular, datos XML representados como árboles.

Teorema [1]. El problema de satisfacción de EML es PSPACE-completo.

¡Buenas noticias! ya que tiene la misma complejidad que el problema de satisfacción de la lógica modal básica (sin ninguna extensión).

Se ha extendido otras lógicas modales (e.g. PDL) con lenguajes más poderosos que los regulares: visibly pushdown languages (VPL), lenguajes libres de contexto (CFL) con resultados variados.

Teorema [2]. El problema de satisfacción de PDL con VPLs es decidable.

Teorema [3]. El problema de satisfacción de PDL con CFLs es indecible.

Motivación y objetivos. Los lenguajes regulares son poco expresivos. Dados los antecedentes positivos en otras lógicas, extendamos EML con VPLs y CFLs y veamos que complejidades obtenemos.

Resultados preliminares. A pesar de las buenas propiedades de los VPLs, la mayoría de las extensiones resultaron *indecibles*.

Teorema. Es suficiente extender a EML con 2 VPLs ó 2 CFLs ó 1 CFL con polaridad negativa para que el problema de satisfacción resulte indecible. Los VPLs y CFLs son igual de expresivos extendiendo EML.

Teorema. El problema de satisfacción de EML extendida con 1 CFL con polaridad positiva (en cada nivel modal) es NEXPTIME.

Problema abierto. ¿Es el caso de 1 CFL positivo NEXPTIME-completo?

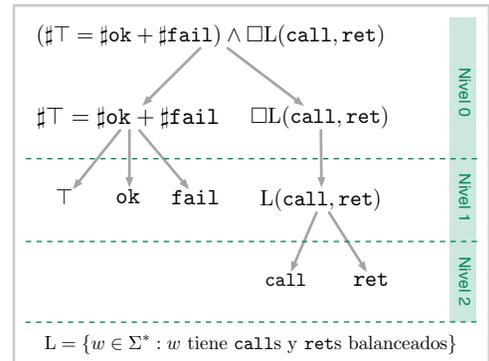


Fig. 3: Formula y sus niveles modales: Especificación parcial de un archivo XML con *stack traces* de programas. La suma de las corridas exitosas más las fallidas deben sumar el total de las corridas y el manejo de la pila debe estar bien balanceado.

[1] Stéphane Demri, Denis Lugiez: Complexity of modal logics with Presburger constraints. J. Applied Logic 8(3): 233-252
 [2] Christof Löding, Carsten Lutz, Olivier Serre: Propositional dynamic logic with recursive programs. J. Log. Algebr. Program. 73(1-2): 51-69
 [3] David Harel, Amir Pnueli, Jonathan Stavi: Propositional Dynamic Logic of Nonregular Programs. J. Comput. Syst. Sci. 26(2): 222-243