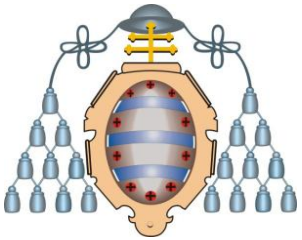
XPath



Departamento de Informática
Universidad de Oviedo
http://www.di.uniovi.es/~labra

XPath

Desarrollado en 1999 como parte de XSL

La especificación de XSL se dividía en:

- XPath: Acceso a partes del documento XML
- XSLT: Transformación
- XSL-FO: Objetos de formateo

Luego se independizó para usarse otros contextos:

- XQuery, XML Schema, XPointer, etc.

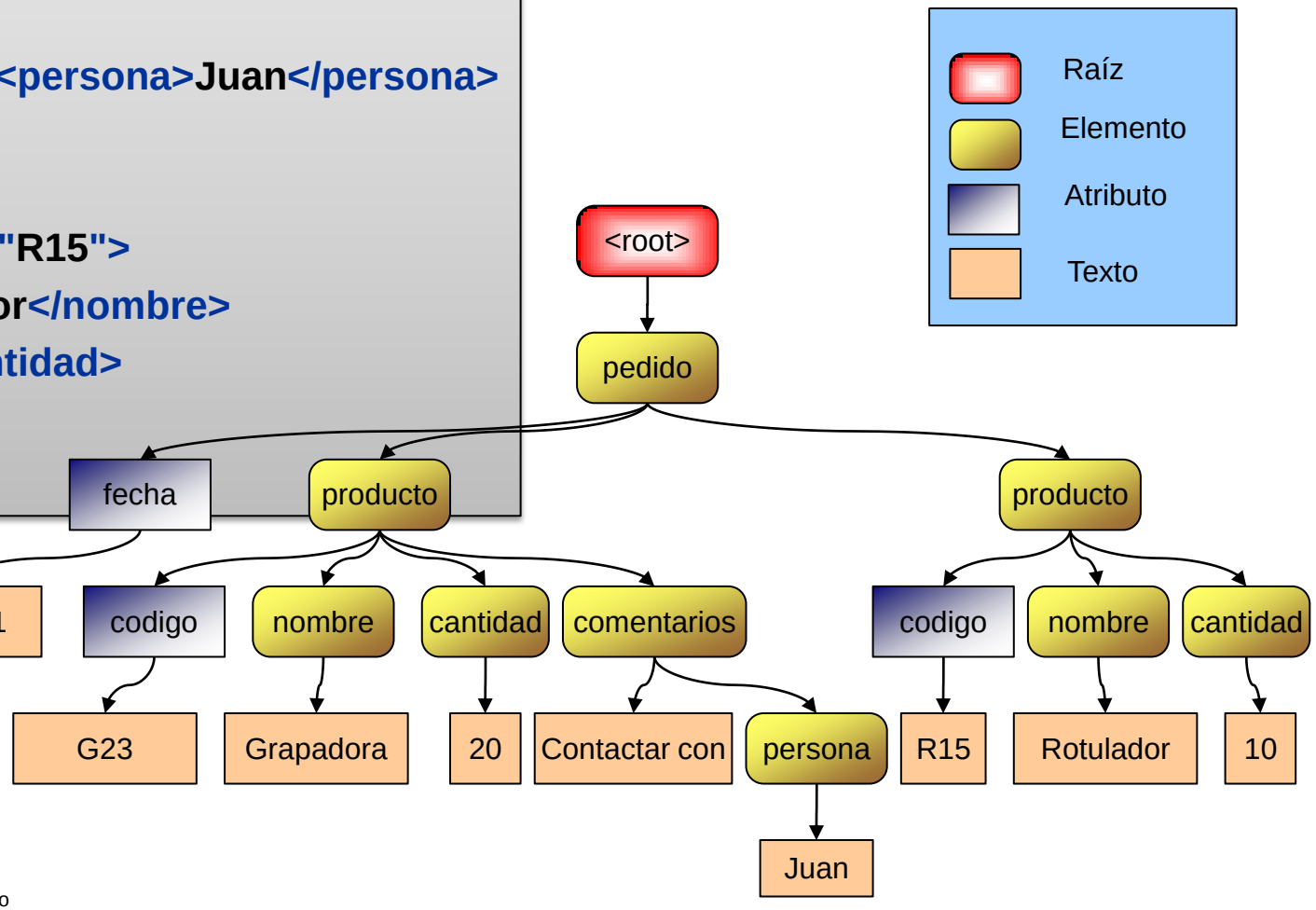
En 2005 se propuso XPath 2.0

Características

- Lenguaje sencillo con sintaxis no XML
- Permite acceder/consultar partes de un documento XML
  - Inspirado en las rutas de acceso a ficheros
- Trabaja sobre el árbol del documento XML

Árbol del documento

```
<pedido fecha="3/2/2011">
 <producto codigo="G23">
  <nombre>Grapadora</nombre>
  <cantidad>20</cantidad>
  <comentarios>
      Contactar con <persona>Juan</persona>
  </comentarios>
 </producto>
 <producto codigo="R15">
  <nombre>Rotulador</nombre>
  <cantidad>10</cantidad>
</producto>
</pedido>
```



Tipos de nodos

Los nodos del árbol pueden ser de los siguientes tipos:

- Nodo raíz
- Elemento
- Atributo
- Comentario
- Instrucciones de procesamiento
- Texto
- Espacios de nombres

Expresiones XPath

Lenguaje inspirado en acceso a directorios

Expresión Xpath: **/pedido/producto/cantidad**

Fichero XML:

```
<pedido fecha="3/2/2011">
 <producto codigo="G23">
  <nombre>Grapadora</nombre>
  <cantidad>20</cantidad>
  <comentarios>
     Contactar con <persona>Juan</persona>
  </comentarios>
 </producto>
 <producto codigo="R15">
  <nombre>Rotulador</nombre>
  <cantidad>10</cantidad>
</producto>
</pedido>
```

Resultado

```
20
10
```

Tipos de Expresiones XPath

Ruta de localización: `/pedido/producto/cantidad`

Llamada a una variable: `$cantidad`

Valor literal: `'Juan', 23`

Llamada a una función:

```
sum(/pedido/producto/cantidad)
```

Unión de conjuntos de nodos mediante | :

```
//nombre | //persona
```

XPath suele utilizarse en valores de atributos en XML

- Normas para codificar cadenas
  - Ejemplo: "`cantidad < 10`"
- Alternar comillas simples/dobles
  - Ejemplo: "`nombre = 'Juan'`"

Tipos de datos en XPath

El resultado de una expresión XPath puede ser:

- Valor booleano (true/false)
- Cadena de caracteres
- Número (entero/flotante)
- Conjunto de nodos XML
  - Los elementos del conjunto son hermanos

Contexto de evaluación

Las expresiones se evalúan en función de un contexto

El contexto contiene:

- Nodo actual del documento
- Posición (entero)
- Tamaño (entero)
- Otra información:
  - Valores de variables
  - Biblioteca de funciones
  - Declaraciones de espacios de nombres

Ruta de localización

Se compone de varios pasos separados por ***/***

```
/pedido/producto/cantidad
```

La ruta puede ser:

- Relativa: se evalúa desde el nodo de contexto
- Absoluta: se evalúa desde el nodo raíz (comienza por ***/***)

Pasos de localización

Cada paso se compone de:

- Un eje (por defecto es `child`)
- Una prueba de nodo
- Opcionalmente, varios predicados entre **[ ]**

Sintaxis:

```
eje::pruebaNodo[predicado1][predicado2]…
```

Ejemplo:

```
/descendant::producto[cantidad > 20]
```

Ejes

child:: Hijos directos del nodo actual (eje por defecto)

parent:: Padre del nodo actual

descendant:: Descendientes

ancestor:: Antecesores

descendant-or-self:: Descendientes incluido el nodo actual

ascestor-or-self:: Antecesores incluido el nodo actual

following:: Los nodos siguientes (incluidos los descencientes)

preceding:: Los precedentes (excluyendo los antecesores)

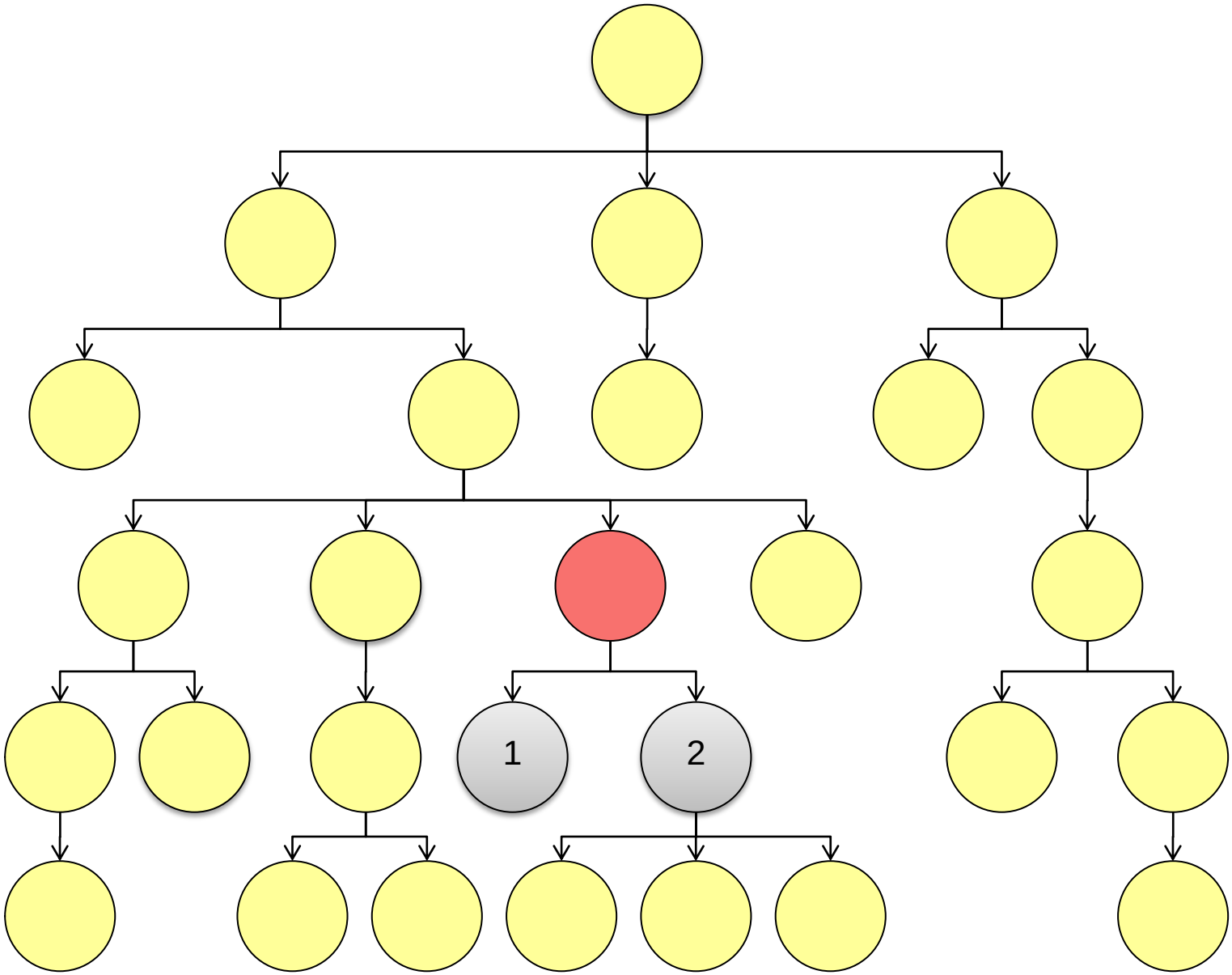
following-sibling:: Hermanos siguientes

preceding-sibling:: Hermanos precedentes

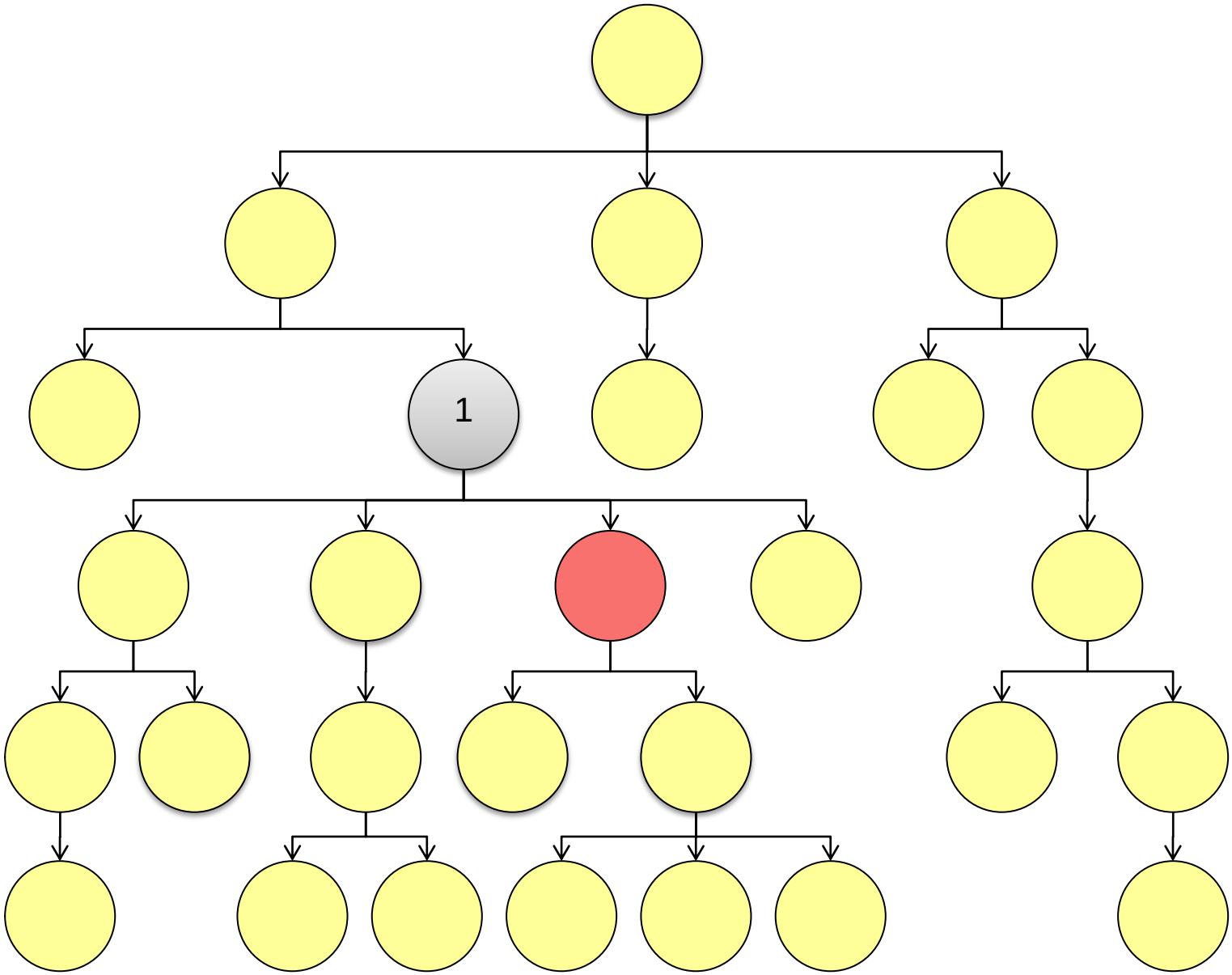
attribute:: Atributo

namespace:: Espacio de nombres

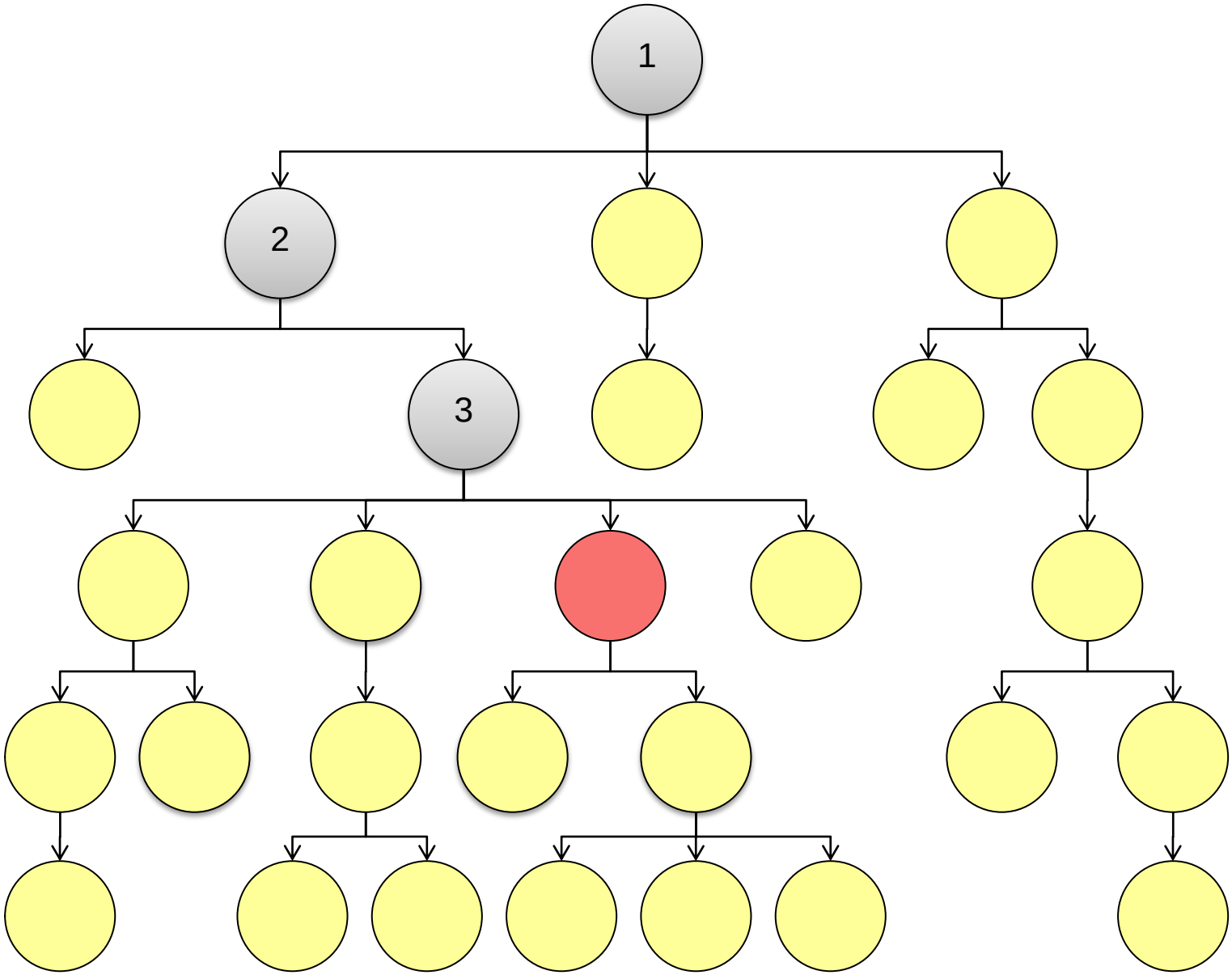
Ejes de localización: child



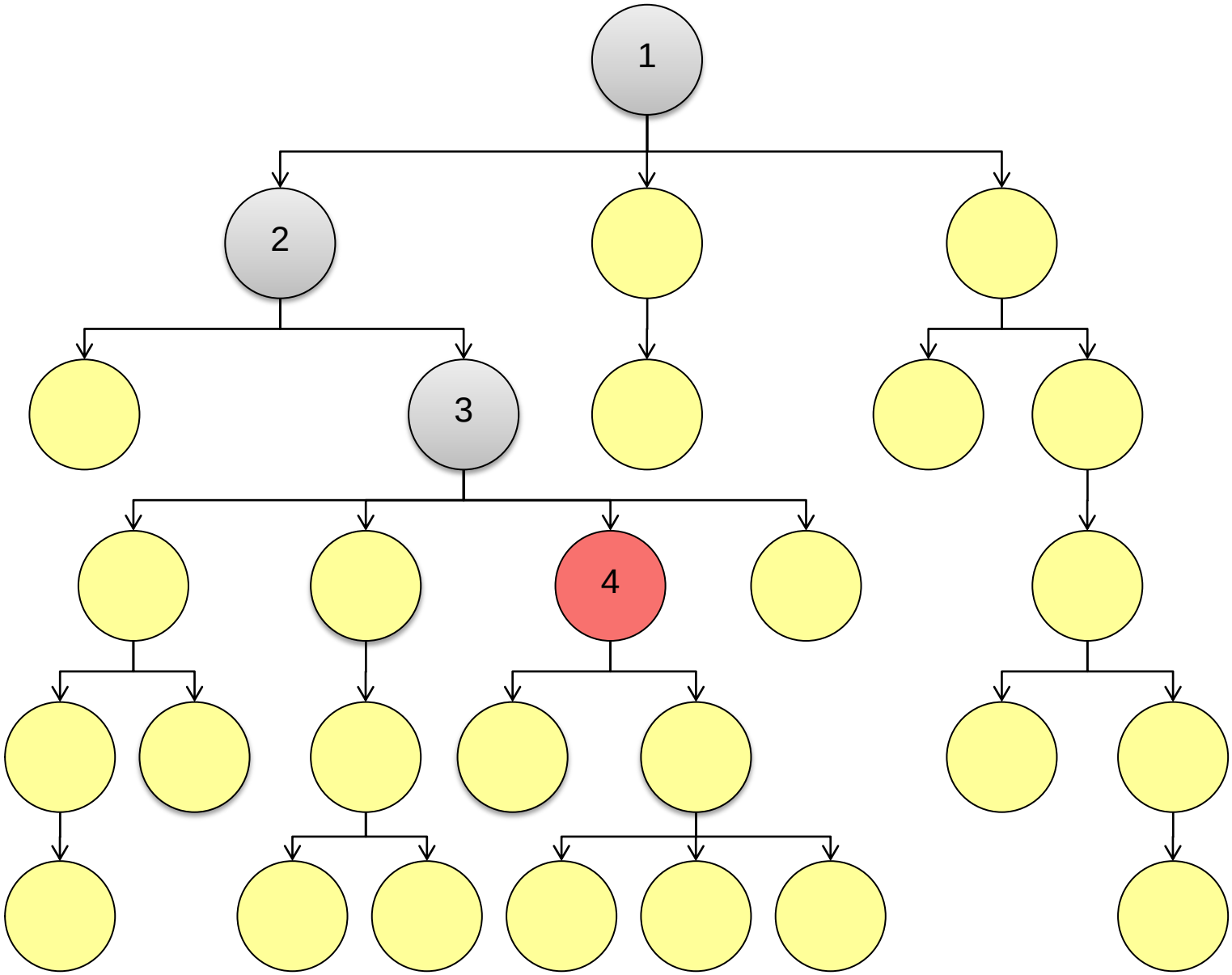
Ejes de localización: parent



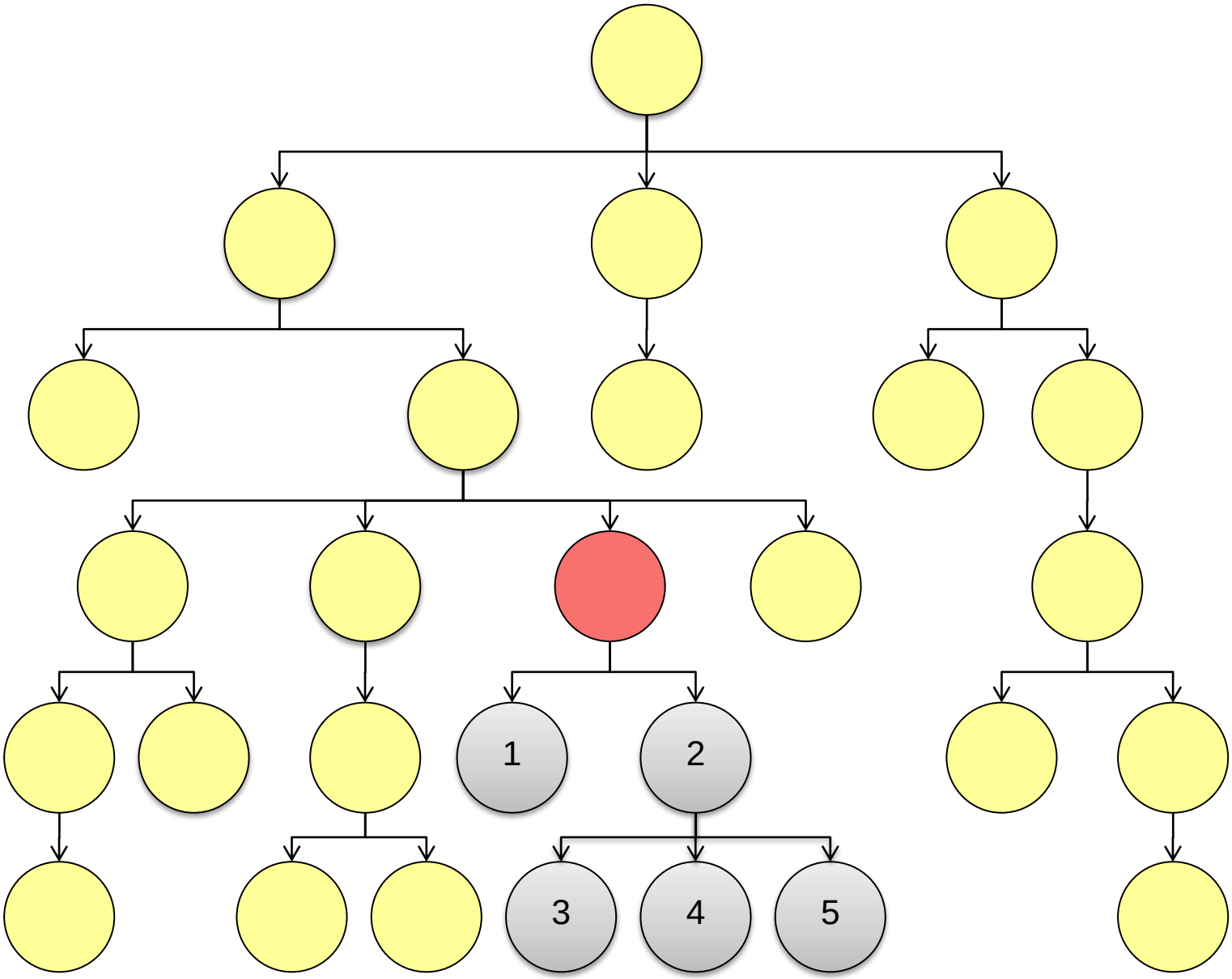
Ejes de localización: ancestor



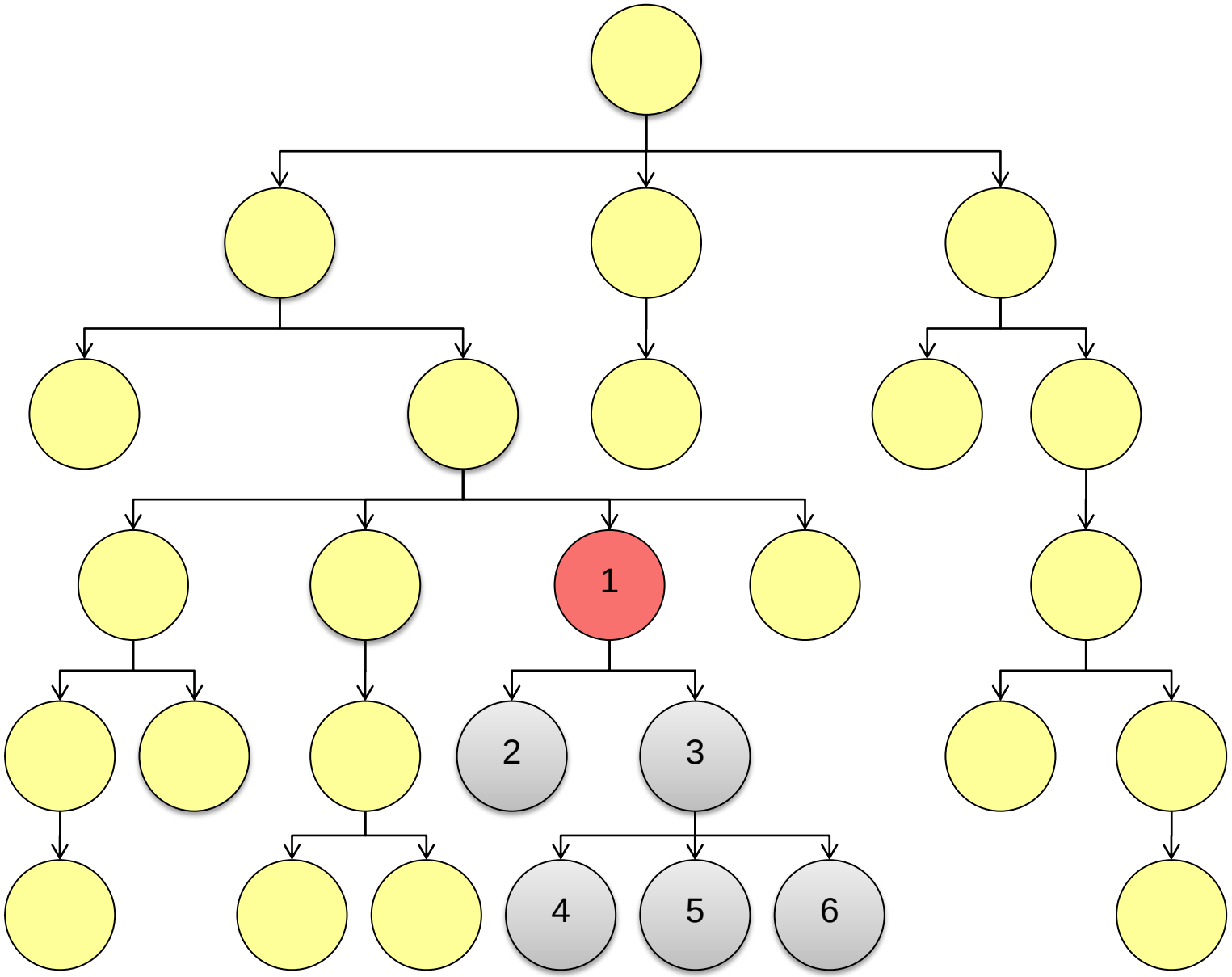
Ejes de localización: ancestor-or-self



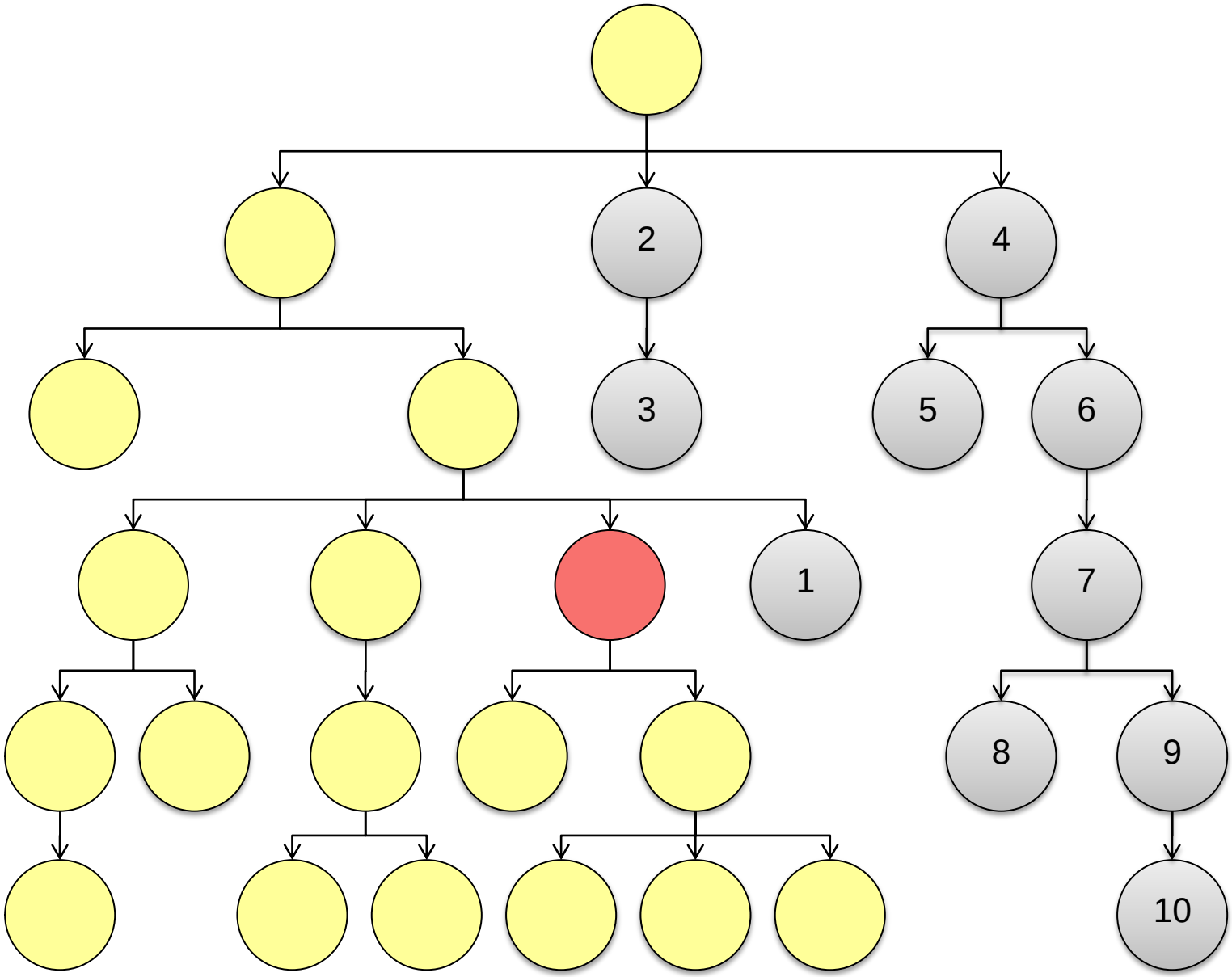
Ejes de localización: descendant



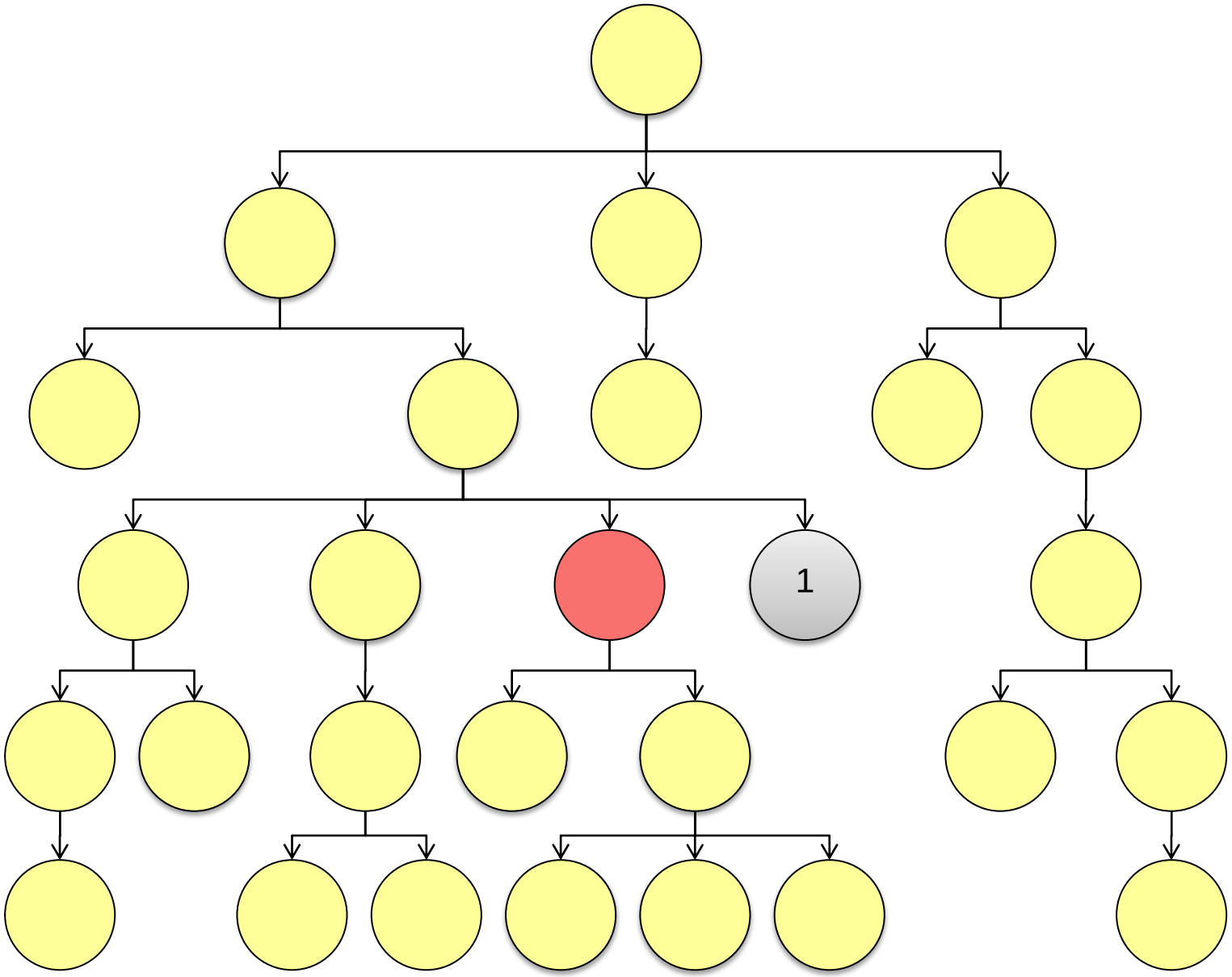
Ejes de localización: descendant-or-self



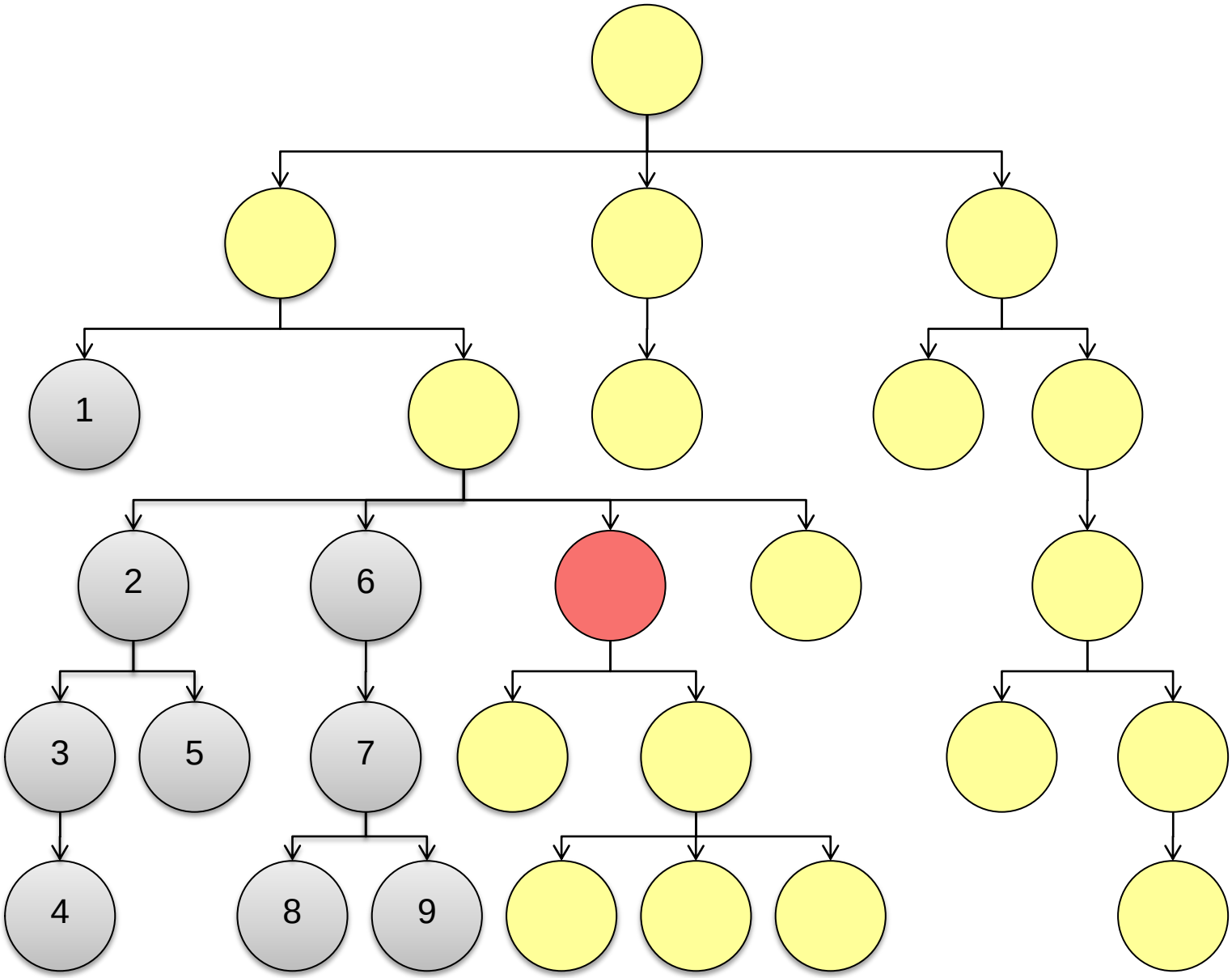
Ejes de localización: following



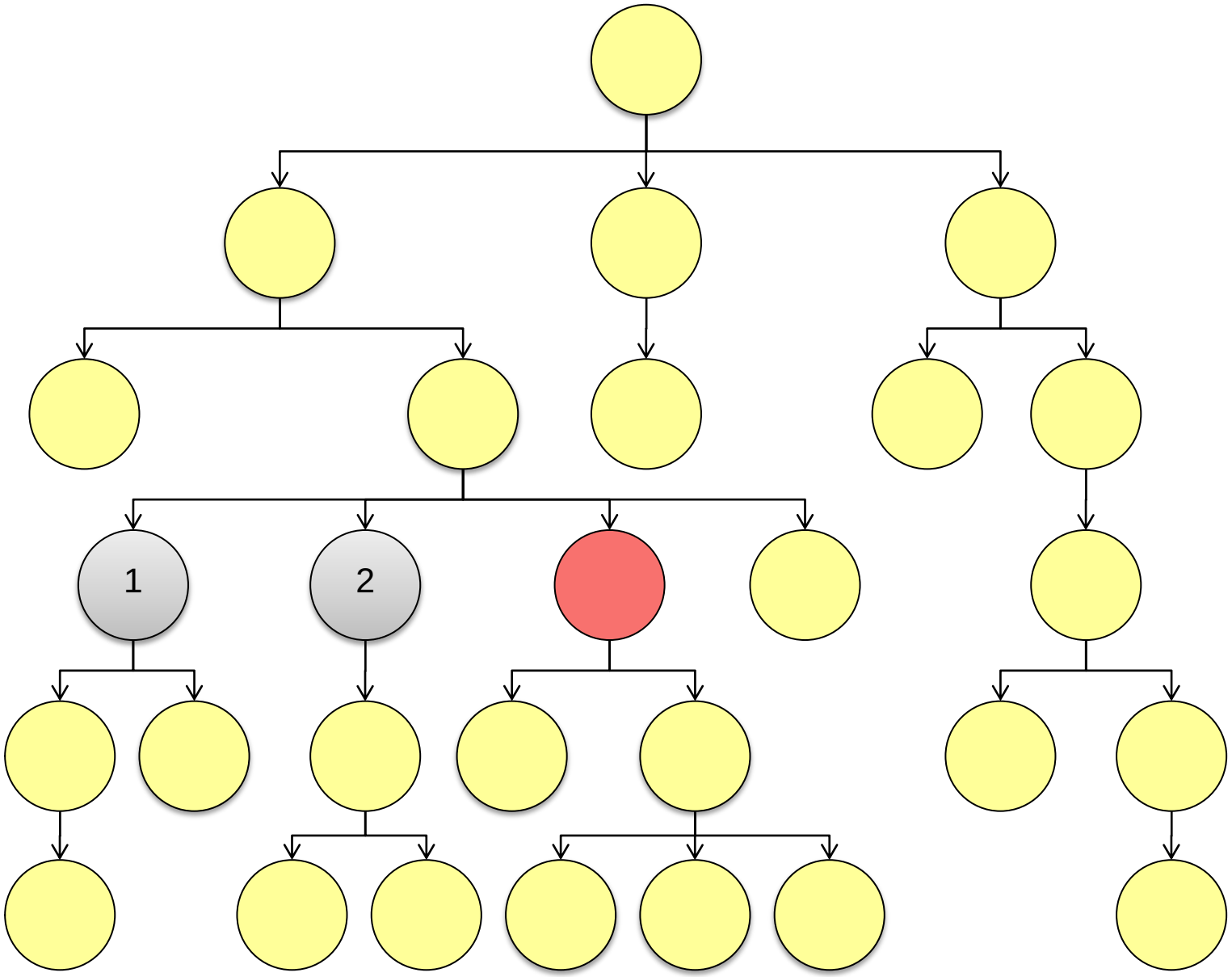
Ejes de localización: following-sibling



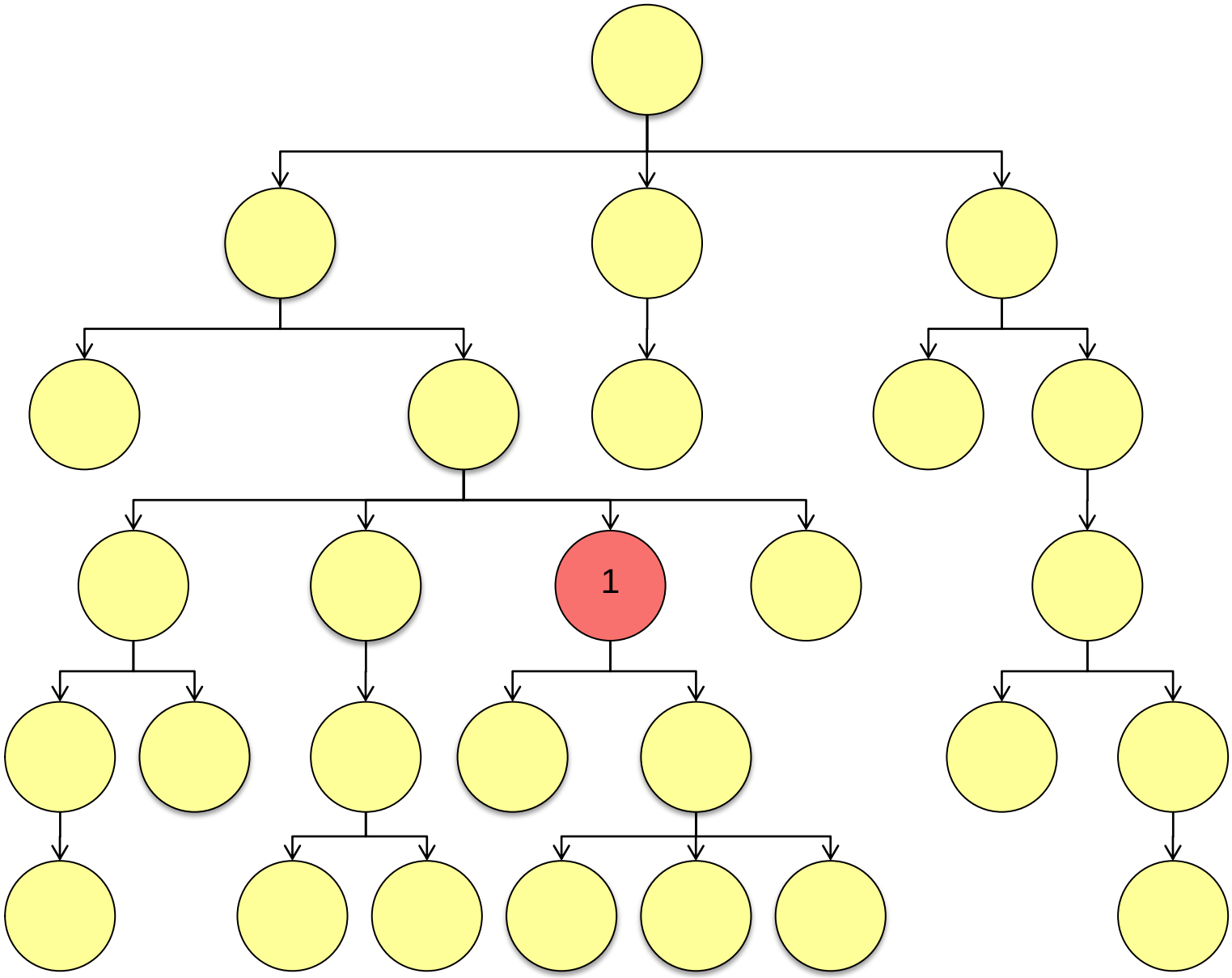
Ejes de localización: preceding



Ejes de localización: preceding-sibling



Ejes de localización: self



Predicados

Añade un nivel de verificación al paso de localización

Expresión booleana

Ejemplos:

```
//producto[cantidad > 15]
//producto[position()=2]
//producto[contains(nombre, ‘pa’)]
```

Puede haber más de un predicado

El orden es significativo

```
//producto[position()=2][contains(nombre,'pa')]
```

≠

```
//producto[contains(nombre, 'pa')] [position()=2]
```

Abreviaturas

child:: es el eje por defecto

/child::pedido/child::producto = /pedido/producto

attribute:: equivale a @

/pedido/attribute::fecha = /pedido/@fecha

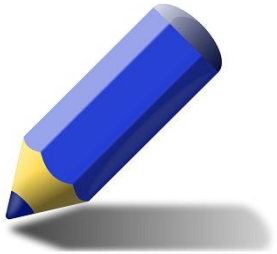
. equivale a self::node()

.. equivale a parent::node()

// equivale a descendant-or-self::node()/

[position() = nº] equivale a [nº]

//producto[position()=2] = //producto[2]



Ejercicio

El fichero Europa.xml* contiene información sobre países europeos, construir expresiones XPath para:

- Obtener el nombre del continente
- Obtener información sobre España
- Obtener el PIB de España
- Obtener países con más de 40 millones de habitantes
- Obtener nombres de países con PIB mayor que España

* http://www.di.uniovi.es/~labra/cursos/XML/files/europa.xml

Funciones predefinidas de XPath

Conversión de tipos

string() convierte a cadena de texto

- Si es un número lo representa
- Si es un booleano devuelve true/false
- Si es un conjunto de nodos, lo aplica al primer nodo

number() convierte a nº

- Si es una cadena de texto, trata de analizarla
- Si es booleano, devuelve 1/true ó 0/false
- Si es un conjunto de nodos, aplica primero String

boolean() convierte a booleano

- Si es un nº es true si es distinto de cero
- Si es una cadena de texto, es true si no está vacía
- Si es un conjunto de nodos, es true si contiene algún nodo

Funciones booleanas

Operadores and y or

Función not()

Funciones matemáticas

Operadores matemáticos habituales:

- +
- - (debe estar precedido de espacio)
- *
- div (la división no es /)
- mod

Operadores de comparación

- =, !=, >, <, <=, >= (si está en XSLT, se pone < ó >)

Otras funciones matemáticas

- sum(), floor(), ceiling(), round(), …

Funciones con cadenas

concat(c1,c2,c*) = resultado de concatenar sus argumentos

concat('uno','dos‘) = ‘unodos‘

Funciones con cadenas

substring(cad,pos,long?) = subcadena de cad a partir de la posición pos de longitud long

substring(‘camina‘,3) = ‘mina‘

substring(‘camina‘,3,2) = ‘mi‘

Funciones con cadenas

contains(cad1, cad2) = true si cad1 contiene cad2

contains(‘camina’,’ca’) = true

contains(‘camina’,’mi’) = true

Funciones con cadenas

starts-with(cad1,cad2) = true si cad1 comienza con cad1

starts-with(‘camina’,’ca’) = true

NOTA: en XPath 1.0 no existe ends-with

Funciones con cadenas

substring-before(cad1,cad2) devuelve el trozo de cadena de cad1 anterior a cad2

substring-after(cad1,cad2) devuelve el trozo de cadena de cad1 posterior a cad2

substring-before(‘camina’,’mi’) = ‘ca’

substring-after(‘camina’,’mi’) = ‘na’

Funciones con cadenas

string-length(cad) devuelve la longitud de cad

string-length(‘camina’) = 6

Funciones con cadenas

normalize-space(cad1) = devuelve el resultado de normalizar cad1 quitando los espacios en blanco redundantes

normalize-space(‘un espacio largo ’)= ‘un espacio largo’

Funciones con cadenas

translate(cad1,cad2,cad3)= devuelve el resultado de traducir todos los caracteres de cad1 que aparecen en cad2 por sus correspondientes en cad3

translate(‘camina’,’aeiou’,’AEIOU’)=‘cAmInA’

Funciones de conjuntos de nodos

position() posición del nodo

last() tamaño del contexto actual

count(nodos) cuenta el número de nodos de un conjunto

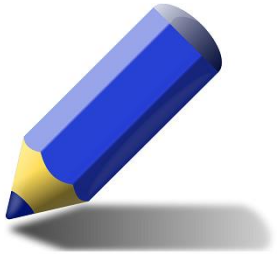
id(objeto) selecciona objetos por su id único

name(nodo) devuelve el nombre del nodo

local-name(nodo) = devuelve el nombre local

namespace-uri(nodo) = devuelve la URI del espacio de nombres

lang(cad)= devuelve true si el idioma del nodo es cad



Ejercicio

A partir del fichero de países europeos, consultar:

- Países empiezan por E
- Países que contienen la letra e
- Países que contienen la letra e (mayúscula o minúscula)
- Países que acaben por e
- PIB medio de los países europeos
- Países cuyo PIB esté por encima de la media

Documentos

document(fichero) permite acceder a un documento XML determinado

Sirve para extraer y combinar información de varios ficheros

XPath 2.0

Propuesto en 2005

Novedades:

- Concepto de secuencias
- Sentencias if y for
- Cuantificadores
- Soporte para tipos primitivos de XML Schema
- Encaje de cadenas mediante expresiones regulares

FIN