



UNIVERSIDAD DE BUENOS AIRES
Facultad de Ciencias Exactas y Naturales
Departamento de Matemática

Tesis de Licenciatura

Caracterizaciones para fragmentos de XPath
con ejes horizontales reflexo-transitivos

Nicolás Ezequiel González

Director: Sergio Alejandro Abriola

Fecha de Presentación: 29/12/2020

Resumen

XPath es un lenguaje de consulta para documentos XML, el cual se abstrae matemáticamente como una lógica que permite expresar propiedades sobre un universo semántico de árboles cuyos nodos están etiquetados y almacenan datos.

En esta tesis trabajamos principalmente con la caracterización al estilo Hennessy-Milner, relacionando la noción de bisimulación con la de equivalencia lógica de modelos puntuados. Hacemos esto para dos fragmentos de XPath poco estudiados, los cuales consideran ejes navegacionales reflexo-transitivos además de la capacidad de descender de a un paso por el árbol.

Índice

1. Introducción	3
2. Preliminares: XPath₌(↓)	5
2.1. Sintaxis y Semántica	5
2.2. Bisimilaridad	7
2.2.1. Juego de Ehrenfeucht-Fraïssé	9
2.2.2. Bisimilaridad Acotada	10
2.3. Fragmento de la Lógica de Primer Orden	11
3. XPath₌(↓, $\overset{*}{\rightarrow}$)	14
3.1. Sintaxis y Semántica	14
3.2. Expresiones Normales	15
3.3. Bisimilaridad	19
3.4. Bisimilaridad Acotada	21
3.5. Caracterización al estilo Hennessy-Milner	28
4. XPath₌(↓, $\overset{*}{\rightarrow}$, $\overset{*}{\leftarrow}$)	31
4.1. Sintaxis y Semántica	31
4.2. Expresiones Normales	31
4.3. Bisimilaridad	34
4.4. Bisimilaridad Acotada	35
4.5. Caracterización al estilo Hennessy-Milner	40
5. Conclusiones	43

1. Introducción

Sintácticamente, la lógica modal básica BML (por sus siglas en inglés) es una extensión de la lógica proposicional agregando a los operadores modales \diamond y su dual \square . Semánticamente, a fines de los '50, Saul Kripke desarrolló los modelos \mathcal{M} formados por elementos x llamados *mundos* que pueden ser *accesibles* entre sí de acuerdo a una relación binaria.

(\mathcal{M}, x) satisface $\diamond\varphi$, si y sólo si, existe un mundo y accesible desde x tal que (\mathcal{M}, y) satisface φ

(\mathcal{M}, x) satisface $\square\varphi$, si y sólo si, para todo mundo y accesible desde x se tiene que (\mathcal{M}, y) satisface φ

A partir de esto, surgieron dos nociones importantes. Por un lado, la de *equivalencia* que nos dice que dos modelos son equivalentes, si y sólo si, satisfacen las mismas propiedades expresables con la sintaxis de BML. Y por el otro, la de *bisimilaridad*, es decir, la existencia de una relación llamada *bisimulación*, que intuitivamente nos asegura que para cualquier recorrido que hagamos en alguno de los modelos, existirá una copia en el otro.

En los '70, Johan van Benthem caracterizó a BML como el fragmento de la lógica de primer orden (con cierta signatura) cuyas fórmulas son invariantes bajo bisimulaciones [3]. En [4], Martin Otto dió una demostración de este teorema desde el ángulo de la teoría de juegos, pensando a las bisimulaciones como juegos de Ehrenfeucht–Fraïssé. Otro tipo de caracterización fue dada por Hennessy y Milner, estableciendo que sobre modelos *localmente finitos* las nociones de equivalencia y bisimulación coinciden.

Con el agregado de distintas modalidades y operadores, fueron apareciendo otras lógicas con distintas capacidades expresivas. De particular interés son aquellas que permiten comparar por igualdad o desigualdad a los datos de una estructura. Por ejemplo, XPath es el lenguaje de consulta más usado para los documentos XML (Extensible Markup Language). Es un estándar abierto y constituye una World Wide Web Consortium (W3C) Recommendation [5]. Así como cada documento XML puede abstraerse matemáticamente como un árbol etiquetado con datos (ver Figura 1), el lenguaje XPath se abstrae como una lógica cuya semántica se evalúa sobre dichos árboles, pudiendo expresar propiedades sobre un nodo o un camino entre dos nodos. Además de la consulta de etiquetas y relaciones de accesibilidad (hijo, padre, descendiente, ancestro, siguiente hermano, anterior hermano, etc.), lo que principalmente distingue a la lógica XPath es la comparación de datos almacenados en pares de nodos.

Objetivos y Organización El objetivo de esta tesis es dar con las nociones adecuadas de bisimulación para fragmentos de XPath con ejes navegacionales reflexo-transitivos horizontales y así poder enunciar resultados conocidos como caracterizaciones al estilo Hennessy-Milner.

La tesis estará organizada en tres secciones y en cada una estudiaremos un fragmento distinto de XPath.

En la Sección 2 daremos una presentación general del fragmento $\text{XPath}_=(\downarrow)$. Definiremos su sintaxis y semántica para luego pasar a las nociones de equivalencia y bisimilaridad entre modelos punteados. Enunciaremos las caracterizaciones al estilo Hennessy-Milner y van Benthem para finalmente dar un aplicación interesante sobre la no expresividad de propiedades. Todos los resultados de esta sección se encuentran en el trabajo [7] de D. Figueira, S. Figueira y Areces.

En la Sección 3 y 4 estudiaremos los fragmentos $\text{XPath}_=(\downarrow, \xrightarrow{*})$ y $\text{XPath}_=(\downarrow, \xrightarrow{*}, \xleftarrow{*})$, respectivamente. Empezaremos por las expresiones directas y normales que nos serán útiles para tener una idea clara de cómo definir bisimulación, y al mismo tiempo, hacerlo de una manera sencilla y adecuada. Luego, pasaremos a las variantes acotadas de equivalencia y bisimilaridad con el objetivo de probar que estas coinciden sobre pares de modelos punteados. Finalizaremos cada sección con la caracterización deseada. La principal diferencia entre ambos fragmentos estará en las expresiones normales, basándonos para las de $\text{XPath}_=(\downarrow, \xrightarrow{*}, \xleftarrow{*})$ en lo hecho por Diego Figueira en [9] donde estudió los problemas de decidibilidad y satisfacibilidad para las expresiones de nodo.

```
<film name="Walden">
  <director name="Jonas Mekas"/>
</film>
<film name="Satantango">
  <director name="Bela Tarr"/>
  <runtime minutes="450"/>
</film>
```

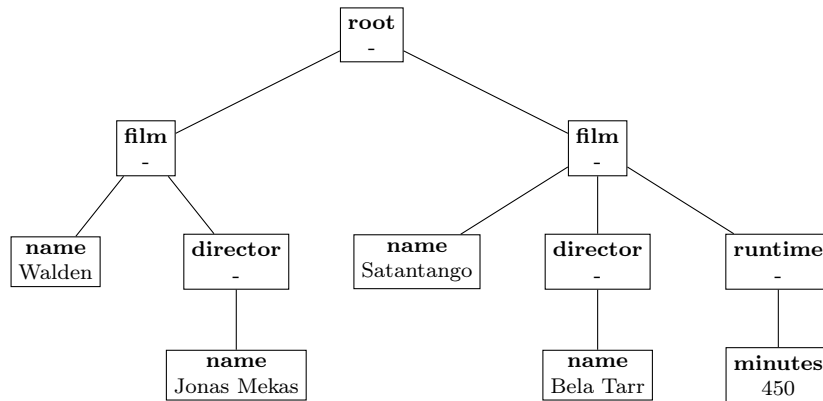


Figura 1: Un documento XML (arriba) y su representación abstracta como un árbol **etiquetado** con datos (abajo).

2. Preliminares: XPath₌(↓)

2.1. Sintaxis y Semántica

Definición 2.1. Un *árbol dirigido* $\mathcal{T} = \langle T, \downarrow^{\mathcal{T}} \rangle$ es un grafo dirigido acíclico con un nodo distinguido $r \in T$ llamado *raíz* tal que para todo $x \in T$ hay un único camino desde r hacia x .

De ser claro por contexto, eliminaremos el superíndice \mathcal{T} de la relación $\downarrow^{\mathcal{T}}$. Supongamos que $x, y \in T$ y $n \in \mathbb{N}_0$ entonces diremos que

y es un hijo de x / x es el padre de y ($\uparrow y := x$) si $x \downarrow y$.

y es un descendiente de x a distancia n ($x \downarrow_n y$) / *x es el ancestro de y a distancia n* ($\uparrow^n y := x$) si existe un (necesariamente único) camino de longitud n desde x hacia y .

el par (\mathcal{T}, x) es un árbol *punteado* y la terna (\mathcal{T}, x, y) es un árbol *bipunteado*.

También, diremos que \mathcal{T} es *finitamente ramificado* si todos sus nodos tienen una cantidad finita de hijos (no necesariamente la misma).

Definición 2.2. Dado un conjunto finito \mathbb{L} de *etiquetas* y otro conjunto $\mathcal{O} = \{\varepsilon, \downarrow\}$ de *navegadores*, la sintaxis de XPath₌(↓) estará formada por dos tipos de fórmulas, las *expresiones de nodo* y las *expresiones de camino* que se definen como sigue:

Todo $a \in \mathbb{L}$ es una una expresión de nodo.

Si φ y ψ son expresiones de nodo entonces también los son las *combinaciones booleanas* $\neg\varphi$ y $\varphi \wedge \psi$.

Si α y β son expresiones de camino entonces $\langle \alpha = \beta \rangle$ y $\langle \alpha \neq \beta \rangle$ son expresiones de nodo, a las cuales llamaremos *tests de datos*.

Si α es una expresión de camino entonces $\langle \alpha \rangle$ es una expresión de nodo.

Nada más es una expresión de nodo.

Todo $o \in \mathcal{O}$ es una expresión de camino.

Si α y β son expresiones de camino entonces la *concatenación* $\alpha\beta$ y la *unión* $\alpha \cup \beta$ también lo son.

Si φ es una expresión de nodo entonces el *test de nodo* $[\varphi]$ es una expresión de camino.

Nada más es una expresión de camino.

Definición 2.3. Un *árbol con datos* $\mathcal{T} = \langle T, \downarrow, label, data \rangle$ es un árbol dirigido junto con funciones $label : T \rightarrow \mathbb{L}$ (etiquetas) y $data : T \rightarrow \mathbb{D}_{\mathcal{T}}$ (datos), donde \mathbb{L} es el conjunto finito de la Definición 2.2 mientras que $\mathbb{D}_{\mathcal{T}}$ es otro conjunto que puede depender de \mathcal{T} y el cual puede ser infinito.

Ejemplo 2.4. En la Figura 2 se puede observar la representación de un árbol con datos \mathcal{T} . Cada vértice es de la forma *nodo* (*etiqueta*, *dato*). Por ejemplo, para la raíz $x \in T$ se tiene $label(x) = a$ y $data(x) = 1$.

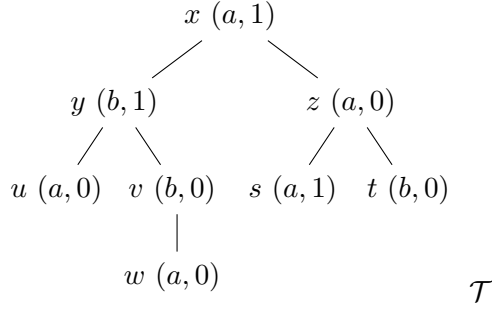


Figura 2: Representación de un árbol con datos.

Definición 2.5. Dado un árbol con datos $\mathcal{T} = \langle T, \downarrow, label, data \rangle$ definimos la semántica de cada expresión de $XPath_{=}(\downarrow)$.

$$\begin{aligned}
\llbracket \varepsilon \rrbracket^{\mathcal{T}} &:= \{(x, y) \in T^2 \mid x = y\} \\
\llbracket \downarrow \rrbracket^{\mathcal{T}} &:= \{(x, y) \in T^2 \mid x \downarrow y\} \\
\llbracket [\varphi] \rrbracket^{\mathcal{T}} &:= \{(x, x) \in T^2 \mid x \in \llbracket \varphi \rrbracket^{\mathcal{T}}\} \\
\llbracket \langle \alpha \beta \rangle \rrbracket^{\mathcal{T}} &:= \{(x, y) \in T^2 \mid \exists z \in T : (x, z) \in \llbracket \langle \alpha \rangle \rrbracket^{\mathcal{T}} \wedge (z, y) \in \llbracket \langle \beta \rangle \rrbracket^{\mathcal{T}}\} \\
\llbracket \langle \alpha \cup \beta \rangle \rrbracket^{\mathcal{T}} &:= \llbracket \langle \alpha \rangle \rrbracket^{\mathcal{T}} \cup \llbracket \langle \beta \rangle \rrbracket^{\mathcal{T}} \\
\llbracket \langle a \rangle \rrbracket^{\mathcal{T}} &:= \{x \in T \mid label(x) = a\} && a \in \mathbb{L} \\
\llbracket \neg \varphi \rrbracket^{\mathcal{T}} &:= T - \llbracket \varphi \rrbracket^{\mathcal{T}} \\
\llbracket \varphi \wedge \psi \rrbracket^{\mathcal{T}} &:= \llbracket \varphi \rrbracket^{\mathcal{T}} \cap \llbracket \psi \rrbracket^{\mathcal{T}} \\
\llbracket \langle \alpha = \beta \rangle \rrbracket^{\mathcal{T}} &:= \{x \in T \mid \exists y, z \in T : (x, y) \in \llbracket \langle \alpha \rangle \rrbracket^{\mathcal{T}} \wedge (x, z) \in \llbracket \langle \beta \rangle \rrbracket^{\mathcal{T}} \wedge data(y) = data(z)\} \\
\llbracket \langle \alpha \neq \beta \rangle \rrbracket^{\mathcal{T}} &:= \{x \in T \mid \exists y, z \in T : (x, y) \in \llbracket \langle \alpha \rangle \rrbracket^{\mathcal{T}} \wedge (x, z) \in \llbracket \langle \beta \rangle \rrbracket^{\mathcal{T}} \wedge data(y) \neq data(z)\} \\
\llbracket \langle \alpha \rangle \rrbracket^{\mathcal{T}} &:= \{x \in T \mid \exists y \in T : (x, y) \in \llbracket \langle \alpha \rangle \rrbracket^{\mathcal{T}}\}
\end{aligned}$$

Como es costumbre, definimos $\varphi \vee \psi := \neg(\neg\varphi \wedge \neg\psi)$. Por otra parte, notar que $\langle \alpha \rangle$ tiene la misma semántica que $\langle \alpha = \alpha \rangle$, por lo que tiene sentido redefinir $\langle \alpha \rangle := \langle \alpha = \alpha \rangle$.

Diremos que el árbol punteado (\mathcal{T}, x) *satisface* a la expresión de nodo φ , si y sólo si, $x \in \llbracket \varphi \rrbracket^{\mathcal{T}}$. Lo indicaremos con la notación $(\mathcal{T}, x) \models \varphi$. De manera análoga, un árbol bipunteado (\mathcal{T}, x, y) *satisface* a la expresión de camino α , si y sólo si, $(x, y) \in \llbracket \langle \alpha \rangle \rrbracket^{\mathcal{T}}$. Escribiremos $(\mathcal{T}, x, y) \models \alpha$.

Ejemplo 2.6. ¿Cómo es la sintaxis de cada una de las siguientes propiedades?

“Este nodo tiene etiqueta a .” $\varphi = a$

“Este nodo tiene hijos con datos distintos.” $\psi = \langle \downarrow \neq \downarrow \rangle$

“El nodo final de este camino tiene etiqueta b y es un nieto o un hijo del nodo inicial.” $\alpha = \downarrow [b] \cup \downarrow \downarrow [b]$

Veamos ahora la semántica sobre el árbol \mathcal{T} de la Figura 2.

$$\llbracket \varphi \rrbracket^{\mathcal{T}} = \{x, z, u, s, w\}$$

$$\llbracket \psi \rrbracket^{\mathcal{T}} = \{x, z\}$$

$$\llbracket \alpha \rrbracket^{\mathcal{T}} = \{(x, y), (x, v), (x, t), (y, v), (z, t)\}$$

2.2. Bisimilaridad

Definición 2.7. Dados árboles punteados (\mathcal{T}, x) y (\mathcal{T}', x') diremos que son \downarrow -*equivalentes* y lo notaremos $(\mathcal{T}, x) \equiv^{\downarrow} (\mathcal{T}', x')$ si valen las mismas propiedades sobre los nodos distinguidos. Es decir, si vale

$$(\mathcal{T}, x) \models \varphi \Leftrightarrow (\mathcal{T}', x') \models \varphi \quad \forall \varphi \in \text{XPath}_{=}(\downarrow) \text{ expresión de nodo}$$

Definición 2.8. Si \mathcal{T} y \mathcal{T}' son dos árboles con datos, diremos que una relación $Z \subseteq T \times T'$ es una \downarrow -*bisimulación* si para todo $(x, x') \in Z$ se cumplen las tres siguientes propiedades:

Harmony. $label(x) = label(x')$

Zig. Si $x \downarrow_n u$ y $x \downarrow_m v$ entonces existen $u', v' \in T'$ tales que:

- $\uparrow^n u' = x' = \uparrow^m v'$
- $(\uparrow^i u, \uparrow^i u') \in Z \quad \forall i \in \{0, 1, \dots, n-1\}$
- $(\uparrow^i v, \uparrow^i v') \in Z \quad \forall i \in \{0, 1, \dots, m-1\}$
- $data(u) = data(v) \Leftrightarrow data(u') = data(v')$

Zag. Si $x' \downarrow_n u'$ y $x' \downarrow_m v'$ entonces existen $u, v \in T$ tales que:

- $\uparrow^n u = x = \uparrow^m v$
- $(\uparrow^i u, \uparrow^i u') \in Z \quad \forall i \in \{0, 1, \dots, n-1\}$
- $(\uparrow^i v, \uparrow^i v') \in Z \quad \forall i \in \{0, 1, \dots, m-1\}$
- $data(u) = data(v) \Leftrightarrow data(u') = data(v')$

Definición 2.9. Dos modelos punteados (\mathcal{T}, x) y (\mathcal{T}', x') son \downarrow -*bisimilares* y lo notaremos $(\mathcal{T}, x) \simeq^{\downarrow} (\mathcal{T}', x')$ si existe una \downarrow -bisimulación Z tal que $(x, x') \in Z$.

Ejemplo 2.10. Si \mathcal{T} y \mathcal{T}' son los árboles de la Figura 3, la relación

$$Z = \{(x, x'), (y_1, y'_1), (y_2, y'_2), (y_3, y'_1), (z_1, z'_1), (z_2, z'_2), (z_3, z'_1)\}$$

es una \downarrow -bisimulación entre (\mathcal{T}, x) y (\mathcal{T}', x') . Por lo tanto, $(\mathcal{T}, x) \simeq^{\downarrow} (\mathcal{T}', x')$.

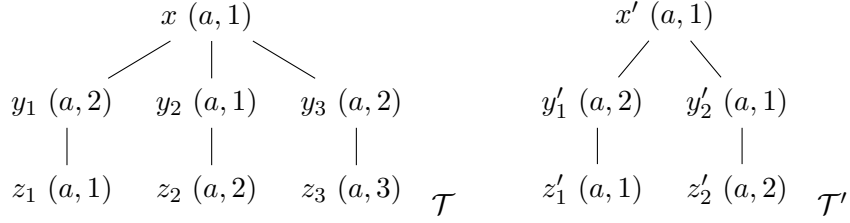


Figura 3: $(\mathcal{T}, x) \simeq^\downarrow (\mathcal{T}', x')$

Observación 2.11. La unión de una familia de \downarrow -bisimulaciones, también es una \downarrow -bisimulación. Por lo tanto, toda \downarrow -bisimulación está incluida dentro de una \downarrow -bisimulación maximal.

La noción de \downarrow -bisimulación es más fuerte que la de \downarrow -equivalencia, aunque sobre árboles finitamente ramificados coinciden. El siguiente enunciado expresa esta propiedad y es conocido como una *caracterización al estilo Hennessy-Milner*.

Teorema 2.12 (Hennessy-Milner). Dados árboles punteados (\mathcal{T}, x) y (\mathcal{T}', x') se tiene que

$$(\mathcal{T}, x) \simeq^\downarrow (\mathcal{T}', x') \implies (\mathcal{T}, x) \equiv^\downarrow (\mathcal{T}', x')$$

Y si ambos son finitamente ramificados, vale la vuelta.

Observación 2.13. La hipótesis de finitamente ramificados no es menor. De lo contrario, podríamos considerar a los árboles de la Figura 4, los cuales cumplen que:

- $data(x) = data(y) \forall (x, y) \in T^2 \cup T'^2$.
- $label(z) = a \forall z \in T \cup T'$.
- $\mathcal{T}, u \equiv^\downarrow \mathcal{T}', u'$. Esto se debe a que \mathcal{T} es una subestructura de \mathcal{T}' .
- $\mathcal{T}, u \not\equiv^\downarrow \mathcal{T}', u'$. Si lo fueran, tendríamos una \downarrow -bisimulación Z . Entonces, para el hijo v' de u' que tiene infinitos descendientes, existiría un v tal que $v Z v'$. Necesariamente, por **Zag**, como v' tiene descendientes de cualquier profundidad, también debería pasar lo mismo con v . Pero esto no ocurre, porque todos los hijos de u tienen descendientes hasta cierta profundidad.

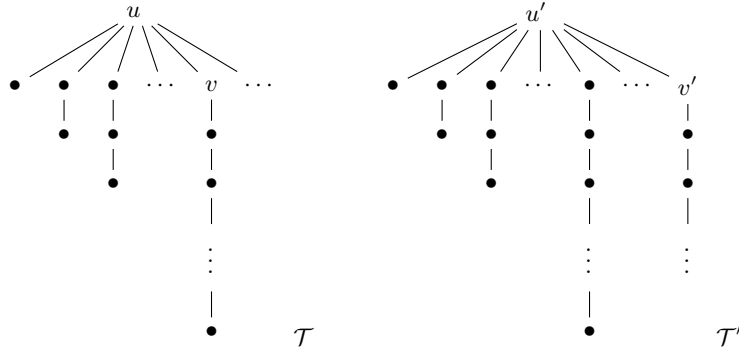


Figura 4: $(\mathcal{T}, u) \equiv^\downarrow (\mathcal{T}', u')$ pero $(\mathcal{T}, u) \not\equiv^\downarrow (\mathcal{T}', u')$

2.2.1. Juego de Ehrenfeucht-Fraïssé

Una \downarrow -bisimulación puede pensarse como un juego entre dos jugadores a los que llamaremos *Spoiler* y *Duplicator*. El objetivo de Spoiler es encontrar una expresión de nodo φ tal que se satisfaga en un modelo pero no en el otro, mientras que Duplicator busca evitar esto.

Supongamos que tenemos dos árboles \mathcal{T} y \mathcal{T}' junto con fichas p y p' que inicialmente se encuentran sobre los nodos $u \in T$ y $u' \in T'$, respectivamente. Si $label(u) \neq label(u')$ entonces el juego termina y Spoiler es el ganador ya que $(\mathcal{T}, u) \models label(u)$ pero $(\mathcal{T}', u') \not\models label(u)$.

Supongamos que en cierto turno las fichas p y p' se encuentran sobre $x \in T$ y $x' \in T'$ con la misma etiqueta.

- Spoiler elige una ficha y dos caminos que empiecen desde el nodo donde se encuentra la ficha seleccionada. Si por ejemplo elige p entonces escoge dos caminos $x \downarrow_n v$ y $x \downarrow_m w$ (de no poder hacer esto por ser x una hoja, Duplicator es el ganador).
- Siguiendo con el ejemplo, Duplicator responde eligiendo dos caminos $x' \downarrow_n v'$ y $x' \downarrow_m w'$ de tal manera que $data(v) = data(w) \Leftrightarrow data(v') = data(w')$ (de no poder hacer esto, Spoiler es el ganador).
- Finalmente, Spoiler elige:
 - O bien, un $i \in \{0, 1, \dots, n-1\}$ y mueve las fichas p y p' a los lugares $\uparrow^i v$ y $\uparrow^i v'$, respectivamente.
 - O bien, un $j \in \{0, 1, \dots, m-1\}$ y mueve las fichas p y p' a los lugares $\uparrow^j w$ y $\uparrow^j w'$, respectivamente.

Si ahora los nodos donde se encuentran las fichas tienen distintas etiquetas, Spoiler es el ganador. De lo contrario, se procede como antes. En caso de que el juego no termine nunca debido a que Spoiler no pueda ganarlo, entonces Duplicator es el ganador.

Chequear que los lugares donde se encuentran las fichas tengan la misma etiqueta es lo mismo que verificar la condición **Harmony**. Por otra parte, la selección y respuesta de caminos que hace cada jugador se vincula con las condiciones **Zig** y **Zag**.

Teorema 2.14. Dados árboles punteados (\mathcal{T}, u) y (\mathcal{T}', u') entonces $(\mathcal{T}, u) \simeq^\downarrow (\mathcal{T}', u')$ si y sólo si Duplicator tiene una estrategia ganadora con las fichas inicialmente sobre u y u' .

2.2.2. Bisimilaridad Acotada

A continuación, veremos las nociones acotadas de \downarrow -equivalencia y \downarrow -bisimilaridad. Informalmente, quiere decir que dados modelos punteados estudiaremos las propiedades que comparten los árboles que surgen al restringirnos a los descendientes que se encuentran a cierta profundidad prefijada de los nodos distinguidos.

Definición 2.15. Dada una fórmula η de $\text{XPath}_=(\downarrow)$ definimos la *profundidad hacia abajo* $\text{dd}(\eta)$ de la siguiente manera:

$$\begin{aligned}
\text{dd}(a) &:= 0 & a \in \mathbb{L} \\
\text{dd}(\neg\varphi) &:= \text{dd}(\varphi) \\
\text{dd}(\varphi \wedge \psi) &:= \text{máx}\{\text{dd}(\varphi), \text{dd}(\psi)\} \\
\text{dd}(\langle \alpha \odot \beta \rangle) &:= \text{máx}\{\text{dd}(\alpha), \text{dd}(\beta)\} & \odot \in \{=, \neq\} \\
\text{dd}(\alpha \cup \beta) &:= \text{máx}\{\text{dd}(\alpha), \text{dd}(\beta)\} \\
\text{dd}(\lambda) &:= 0 & \lambda \text{ es el string vacío} \\
\text{dd}(\varepsilon\alpha) &:= \text{dd}(\alpha) \\
\text{dd}([\varphi]\alpha) &:= \text{máx}\{\text{dd}(\varphi), \text{dd}(\alpha)\} \\
\text{dd}(\downarrow\alpha) &:= \text{dd}(\alpha) + 1
\end{aligned}$$

Denotaremos por $\ell\text{-XPath}_=(\downarrow)$ al subconjunto de fórmulas $\eta \in \text{XPath}_=(\downarrow)$ tales que $\text{dd}(\eta) \leq \ell$.

Definición 2.16. Diremos que dos árboles punteados (\mathcal{T}, x) y (\mathcal{T}', x') son \downarrow_ℓ -*equivalentes* y escribiremos $(\mathcal{T}, x) \equiv_\ell^\downarrow (\mathcal{T}', x')$ si satisfacen las mismas expresiones de nodo con profundidad hacia abajo menor o igual a ℓ . Es decir, $(\mathcal{T}, x) \equiv_\ell^\downarrow (\mathcal{T}', x')$, si y sólo si,

$$(\mathcal{T}, x) \models \varphi \Leftrightarrow (\mathcal{T}', x') \models \varphi \quad \forall \varphi \in \ell\text{-XPath}_=(\downarrow)$$

Definición 2.17. Sean árboles con datos \mathcal{T} y \mathcal{T}' . Supongamos que para cada $j \in \{0, \dots, \ell\}$ tenemos una relación $Z_j \subseteq T \times T'$. Diremos que la familia $\mathcal{Z} = \{Z_j\}_{0 \leq j \leq \ell}$ es una \downarrow_ℓ -*bisimulación* si para todo Z_j y $(x, x') \in Z_j$ se cumplen las tres siguientes:

Harmony. $label(x) = label(x')$

Zig. Dados $n, m \leq j$, para todo $x \downarrow_n v$ y $x \downarrow_m w$ existen $x' \downarrow_n v'$ y $x' \downarrow_m w'$ tales que:

- $(\uparrow^i v, \uparrow^i v') \in Z_{j-n+i}$ para todo $i \in \{0, \dots, n-1\}$
- $(\uparrow^i w, \uparrow^i w') \in Z_{j-m+i}$ para todo $i \in \{0, \dots, m-1\}$

Zag. Dados $n, m \leq j$, para todo $x' \downarrow_n v'$ y $x' \downarrow_m w'$ existen $x \downarrow_n v$ y $x \downarrow_m w$ tales que:

- $(\uparrow^i v, \uparrow^i v') \in Z_{j-n+i}$ para todo $i \in \{0, \dots, n-1\}$
- $(\uparrow^i w, \uparrow^i w') \in Z_{j-m+i}$ para todo $i \in \{0, \dots, m-1\}$

Definición 2.18. Dos árboles punteados (\mathcal{T}, x) y (\mathcal{T}', x') son ℓ -bisimilares si existe una \downarrow_ℓ -bisimulación $\mathcal{Z} = \{Z_j\}_{0 \leq j \leq \ell}$ tal que $(x, x') \in Z_\ell$. En tal caso, escribiremos $(\mathcal{T}, x) \simeq_\ell^\downarrow (\mathcal{T}', x')$.

Observación 2.19. Si $(\mathcal{T}, x) \simeq^\downarrow (\mathcal{T}', x')$ vía una \downarrow -bisimulación Z entonces $(\mathcal{T}, x) \simeq_\ell^\downarrow (\mathcal{T}', x')$ vía $\mathcal{Z} = \{Z_j\}_{0 \leq j \leq \ell}$ con $Z_j := Z$ para todo $j \in \{0, \dots, \ell\}$.

Teorema 2.20. Dos modelos punteados son \downarrow_ℓ -equivalentes, si y sólo si, son \downarrow_ℓ -bisimilares.

$$(\mathcal{T}, x) \equiv_\ell^\downarrow (\mathcal{T}', x') \iff (\mathcal{T}, x) \simeq_\ell^\downarrow (\mathcal{T}', x')$$

2.3. Fragmento de la Lógica de Primer Orden

Definición 2.21. Todo árbol con datos \mathcal{T} es un modelo para la lógica de primer orden con signatura $\sigma := \{\sim, R_\downarrow, (P_a)_{a \in \mathbb{L}}\}$ formada por dos relaciones binarias R_\downarrow y \sim , y una relación unaria P_a por cada etiqueta $a \in \mathbb{L}$. Al conjunto de fórmulas con esta signatura lo denotaremos por $\text{FO}(\sigma)$.

La semántica de cada símbolo de σ sobre \mathcal{T} se define a continuación:

$$\begin{aligned} R_\downarrow^\mathcal{T} &:= \{(x, y) \in T^2 \mid x \downarrow y\} \\ \sim^\mathcal{T} &:= \{(x, y) \in T^2 \mid data(x) = data(y)\} \\ P_a^\mathcal{T} &:= \{x \in T \mid label(x) = a\} \end{aligned}$$

Definición 2.22. Toda fórmula de $\text{XPath}_=(\downarrow)$ puede traducirse a una de

FO(σ) por medio de las funciones ST_x y $ST_{x,y}$ definidas a continuación:

$$\begin{aligned}
ST_x(a) &:= P_a(x) \\
ST_x(\varphi \wedge \psi) &:= ST_x(\varphi) \wedge ST_x(\psi) \\
ST_x(\neg\varphi) &:= \neg ST_x(\varphi) \\
ST_x(\langle\alpha = \beta\rangle) &:= (\exists y)(\exists z)[(y \sim z) \wedge ST_{x,y}(\alpha) \wedge ST_{x,y}(\beta)] \\
ST_x(\langle\alpha \neq \beta\rangle) &:= (\exists y)(\exists z)[(y \not\sim z) \wedge ST_{x,y}(\alpha) \wedge ST_{x,y}(\beta)] \\
ST_x(\langle\alpha\rangle) &:= (\exists y) ST_{x,y}(\alpha) \\
ST_{x,y}(\varepsilon) &:= (x = y) \\
ST_{x,y}(\downarrow) &:= R_{\downarrow}(x, y) \\
ST_{x,y}([\varphi]) &:= (x = y) \wedge ST_x(\varphi) \\
ST_{x,y}(\alpha\beta) &:= (\exists z)[ST_{x,z}(\alpha) \wedge ST_{z,y}(\beta)] \\
ST_{x,y}(\alpha \cup \beta) &:= ST_{x,y}(\alpha) \vee ST_{x,y}(\beta)
\end{aligned}$$

Notar que ST_x transforma expresiones de nodo en fórmulas con una variable libre x , mientras que $ST_{x,y}$ manda expresiones de camino en fórmulas con dos variables libres x e y .

Teorema 2.23. Las funciones ST_x y $ST_{x,y}$ preservan verdades. Es decir, una propiedad es válida en XPath Hacia Abajo, si y sólo si, lo es en la lógica de primer orden con signatura σ .

$$\begin{aligned}
(\mathcal{T}, u) \models \varphi &\iff \mathcal{T} \models ST_x(\varphi)[u] \\
(\mathcal{T}, u, v) \models \alpha &\iff \mathcal{T} \models ST_{x,y}(\alpha)[u, v]
\end{aligned}$$

El Teorema 2.12 nos dice que las expresiones de nodo de $\text{XPath}_=(\downarrow)$ son invariantes por \downarrow -bisimulaciones entre modelos punteados. Si lo llevamos al contexto de la lógica de primer orden, el siguiente enunciado afirma que las fórmulas de FO(σ) con una variable libre e invariantes por \downarrow -bisimulaciones son justamente (salvo equivalencia lógica) aquellas que resultan ser la traducción de una expresión de nodo de $\text{XPath}_=(\downarrow)$. Esto se conoce como *caracterización al estilo van Benthem*. La demostración del siguiente enunciado es similar a la que se encuentra en [4] dada por Otto para lógicas modales.

Teorema 2.24 (van Benthem). Dada $\varphi(x) \in \text{FO}(\sigma)$, son equivalentes:

1. φ es invariante por \downarrow -bisimulaciones entre árboles punteados, o sea:

$$(\mathcal{T}, u) \simeq^{\downarrow} (\mathcal{T}', u') \implies (\mathcal{T} \models \varphi[u] \iff \mathcal{T}' \models \varphi[u'])$$

2. Existe $\psi \in \text{XPath}_=(\downarrow)$ expresión de nodo tal que φ es lógicamente equivalente a $ST_x(\psi)$. Es decir,

$$\mathcal{T} \models \varphi[u] \iff \mathcal{T} \models ST_x(\psi)[u]$$

Una aplicación del Teorema 2.24 (aunque, con el Teorema 2.12 ya nos alcance) tiene que ver con mostrar que una propiedad no puede expresarse con la sintaxis de XPath Hacia Abajo. La idea es dar dos modelos punteados \downarrow -bisimilares tal que solo uno satisfaga dicha propiedad.

Ejemplo 2.25. Sea $\eta(x)$ la propiedad que expresa que para un nodo x existe y hijo de x y z hijo de y tales que todos tienen dato distinto dos a dos. Notar que $\eta(x) \in \text{FO}(\sigma)$ pues

$$\eta(x) = (\exists y)(\exists z)[R_{\downarrow}(x, y) \wedge R_{\downarrow}(y, z) \wedge (x \not\sim y) \wedge (y \not\sim z) \wedge (z \not\sim x)]$$

Si consideramos los árboles de la Figura 3 entonces son \downarrow -bisimilares por el Ejemplo 2.10 y está claro que η se satisface en (\mathcal{T}, x) pero no en (\mathcal{T}', x') . Por lo tanto, por el Teorema 2.12 y 2.23, η no puede ser equivalente a la traducción de una fórmula de XPath.

Esto nos dice que (como era de esperar) el nivel de expresividad de XPath Hacia Abajo es estrictamente menor al de la lógica de primer orden con signatura σ .

3. XPath= $(\downarrow, \overset{*}{\rightarrow})$

3.1. Sintaxis y Semántica

Para poder trabajar con ejes horizontales necesitaremos que los hijos de todo nodo estén ordenados de cierta manera. Es decir, habrá que restringirnos a modelos donde tenga sentido hablar del siguiente o anterior hermano de un nodo.

Definición 3.1. Un subconjunto $S \subseteq \mathbb{Z}$ es un *segmento* si para todo $n, m \in S$ se cumple que si $n \leq k \leq m$ entonces $k \in S$.

Definición 3.2. Un *árbol ordenado* $\mathcal{T} = \langle T, \downarrow, \delta \rangle$ es un árbol dirigido junto con una función $\delta : T \rightarrow \mathbb{Z}$ tal que si $H_{\mathcal{T}}(x) := \{y \in T \mid x \downarrow y\}$ es el conjunto de hijos de un nodo x entonces la restricción $\delta|_{H_{\mathcal{T}}(x)}$ es una biyección cuya imagen es un segmento de \mathbb{Z} para todo $x \in T$

En este contexto, si $y, z \in H_{\mathcal{T}}(x)$ entonces

diremos que z es el siguiente hermano de y / y es el anterior hermano de z si $\delta(z) = \delta(y) + 1$ y lo indicaremos con la notación $y \rightarrow z$.

diremos que z es un hermano a derecha de y / y es un hermano a izquierda de z si $\delta(y) < \delta(z)$ y lo indicaremos con la notación $y \overset{+}{\rightarrow} z$.

escribiremos $y \overset{*}{\rightarrow} z$ si $\delta(y) \leq \delta(z)$.

Notar que las relaciones \rightarrow , $\overset{+}{\rightarrow}$ y $\overset{*}{\rightarrow}$ sobre T cumplen que

$\overset{*}{\rightarrow}$ es un orden parcial

$\overset{+}{\rightarrow}$ es la *clausura transitiva* de \rightarrow

$\overset{*}{\rightarrow}$ es la *clausura reflexo-transitiva* de \rightarrow

Definición 3.3. Un *árbol ordenado con datos* $\mathcal{T} = \langle T, \downarrow, \delta, label, data \rangle$ es un árbol ordenado junto con funciones $label : T \rightarrow \mathbb{L}$ (etiquetas) y $data : T \rightarrow \mathbb{D}_{\mathcal{T}}$ (datos). Donde, nuevamente, \mathbb{L} es el conjunto finito de la Definición 2.2 y $\mathbb{D}_{\mathcal{T}}$ es otro conjunto que depende de \mathcal{T} , pudiendo ser infinito.

Observación 3.4. En la representación de un árbol ordenado con datos, los valores específicos de la función δ no son necesarios. Simplemente, ubicaremos los nodos de izquierda a derecha de manera creciente.

Ejemplo 3.5. Si consideramos el árbol de la Figura 2, entonces podemos pensar que está ordenado. Por ejemplo, tendremos que $s \overset{*}{\rightarrow} t$ ya que son hijos de z y s se encuentra primero de izquierda a derecha. Más aún, $s \rightarrow t$.

Definición 3.6. La sintaxis de $\text{XPath}_=(\downarrow, \overset{*}{\rightarrow})$ es similar a la dada en la Definición 2.2 para $\text{XPath}_=(\downarrow)$ salvo que ahora $\mathcal{O} = \{\varepsilon, \downarrow, \overset{*}{\rightarrow}\}$.

La semántica de $\overset{*}{\rightarrow}$ evaluada sobre un árbol ordenado con datos \mathcal{T} es $\llbracket \overset{*}{\rightarrow} \rrbracket^{\mathcal{T}} := \{(x, y) \in T^2 \mid x \overset{*}{\rightarrow} y\}$.

Para el resto de los símbolos es análoga a la dada en la Definición 2.5, aplicada ahora sobre un árbol ordenado con datos.

3.2. Expresiones Normales

Definición 3.7. Dos fórmulas φ y ψ de $\text{XPath}_=(\downarrow, \overset{*}{\rightarrow})$ son *lógicamente equivalentes* (\equiv) si para toda estructura la semántica es la misma. Es decir,

$$\varphi \equiv \psi \stackrel{\text{def}}{\iff} \llbracket \varphi \rrbracket^{\mathcal{T}} = \llbracket \psi \rrbracket^{\mathcal{T}} \text{ para todo } \mathcal{T} \text{ árbol ordenado con datos}$$

Observación 3.8. Supongamos que

$$\begin{array}{ll} \varphi \equiv \varphi' & \alpha \equiv \alpha' \\ \psi \equiv \psi' & \beta \equiv \beta' \end{array}$$

entonces la equivalencia lógica se preserva por cualquier construcción

$$\begin{array}{ll} \neg\varphi \equiv \neg\varphi' & \alpha\beta \equiv \alpha'\beta' \\ \varphi \wedge \psi \equiv \varphi' \wedge \psi' & \alpha \cup \beta \equiv \alpha' \cup \beta' \\ \langle \alpha = \beta \rangle \equiv \langle \alpha' = \beta' \rangle & [\varphi] \equiv [\varphi'] \\ \langle \alpha \neq \beta \rangle \equiv \langle \alpha' \neq \beta' \rangle & \end{array}$$

Definición 3.9. Una expresión de camino α es *directa* si es de la forma

$$\alpha = [\psi_0]o_1[\psi_1] \dots o_n[\psi_n] \quad o_i \in \{\downarrow, \overset{*}{\rightarrow}\} \forall i \in \{1, \dots, n\}$$

A continuación, definiremos para cada fórmula el conjunto de las subexpresiones de camino que aparecen en sus tests de datos. Esto nos permitirá definir de manera más sencilla aquellas expresiones de nodo cuyo conjunto de subexpresiones está formado únicamente por expresiones directas.

Definición 3.10. Para cada fórmula η de $\text{XPath}_=(\downarrow, \overset{*}{\rightarrow})$ definimos recursivamente al conjunto de subexpresiones de camino $\text{sub}(\eta)$ que aparecen en los tests de datos de su sintaxis:

$$\begin{array}{ll} \text{sub}(a) := \emptyset & a \in \mathbb{L} \\ \text{sub}(o) := \emptyset & o \in \{\varepsilon, \downarrow, \overset{*}{\rightarrow}\} \\ \text{sub}(\neg\varphi) := \text{sub}(\varphi) & \\ \text{sub}(\varphi \wedge \psi) := \text{sub}(\varphi) \cup \text{sub}(\psi) & \\ \text{sub}(\langle \alpha \odot \beta \rangle) := \text{sub}(\alpha) \cup \text{sub}(\beta) \cup \{\alpha, \beta\} & \odot \in \{=, \neq\} \\ \text{sub}(\alpha\beta) := \text{sub}(\alpha) \cup \text{sub}(\beta) & \\ \text{sub}(\alpha \cup \beta) := \text{sub}(\alpha) \cup \text{sub}(\beta) & \\ \text{sub}([\varphi]) := \text{sub}(\varphi) & \end{array}$$

Notar que si α es una expresión de camino entonces $\alpha \notin \text{sub}(\alpha)$.

Definición 3.11. Una expresión de camino α es *normal* si toda $\gamma \in \text{sub}(\alpha) \cup \{\alpha\}$ es directa.

Una expresión de nodo φ es *normal* si para toda $\gamma \in \text{sub}(\varphi)$ se tiene que γ es directa.

Definición 3.12. Definimos la *tautología de nodo* $\top := \bigvee_{a \in \mathbb{L}} a$.

Notar que \top está bien definido pues \mathbb{L} es finito. Además, efectivamente se cumple que $(\mathcal{T}, x) \models \top$ para todo modelo punteado (\mathcal{T}, x) .

Observación 3.13. Algunas propiedades que nos serán útiles.

1. La unión es asociativa.

$$(\alpha \cup \beta) \cup \gamma \equiv \alpha \cup (\beta \cup \gamma)$$

2. La concatenación es distributiva respecto a la unión.

$$\gamma(\alpha \cup \beta) \equiv \gamma\alpha \cup \gamma\beta \quad (\alpha \cup \beta)\gamma \equiv \alpha\gamma \cup \beta\gamma$$

3. La concatenación de tests de nodo es un test de nodo.

$$[\varphi][\psi] \equiv [\varphi \wedge \psi]$$

4. Podemos sacar uniones en los tests de datos como disyunciones.

$$\langle \alpha \cup \beta \odot \gamma \rangle \equiv \langle \alpha \odot \gamma \rangle \vee \langle \beta \odot \gamma \rangle \quad \odot \in \{=, \neq\}$$

Definición 3.14. La *profundidad de accesibilidad* $\text{ad}(\eta)$ de una fórmula η mide la distancia a la cual basta explorar un árbol desde un nodo inicial para poder determinar si es verdadera o no.

Formalmente, se define recursivamente como sigue:

$$\begin{aligned} \text{ad}(a) &:= 0 & a \in \mathbb{L} \\ \text{ad}(\neg\varphi) &:= \text{ad}(\varphi) \\ \text{ad}(\varphi \wedge \psi) &:= \text{máx}\{\text{ad}(\varphi), \text{ad}(\psi)\} \\ \text{ad}(\langle \alpha \odot \beta \rangle) &:= \text{máx}\{\text{ad}(\alpha), \text{ad}(\beta)\} & \odot \in \{=, \neq\} \\ \text{ad}(\alpha \cup \beta) &:= \text{máx}\{\text{ad}(\alpha), \text{ad}(\beta)\} \\ \text{ad}(\lambda) &:= 0 & \lambda \text{ es el string vacío} \\ \text{ad}(\varepsilon\alpha) &:= \text{ad}(\alpha) \\ \text{ad}([\varphi]\alpha) &:= \text{máx}\{\text{ad}(\varphi), \text{ad}(\alpha)\} \\ \text{ad}(o\alpha) &:= \text{ad}(\alpha) + 1 & o \in \{\downarrow, \overset{*}{\rightarrow}\} \end{aligned}$$

Al conjunto de expresiones $\eta \in \text{XPath}_{=(\downarrow, \overset{*}{\rightarrow})}$ tales que $\text{ad}(\eta) \leq \ell$ lo denotaremos $\ell\text{-XPath}_{=(\downarrow, \overset{*}{\rightarrow})}$.

Observación 3.15. Como consecuencia de las definiciones anteriores, para la disyunción se tiene que $\text{ad}(\varphi \vee \psi) = \text{máx}\{\text{ad}(\varphi), \text{ad}(\psi)\}$.

Ejemplo 3.16. Veamos un ejemplo de expresión de camino normal y otra que no lo sea.

$$\begin{aligned}\alpha &= [a \wedge \langle [b] \xrightarrow{*} [a] \rangle] \downarrow [b] && \text{Normal} \\ \beta &= [a] \xrightarrow{*} [b] \xrightarrow{*} [\langle \alpha \neq \varepsilon \cup \downarrow \rangle] \downarrow [a] \xrightarrow{*} && \text{No Normal}\end{aligned}$$

El problema principal con β es que $(\varepsilon \cup \downarrow) \in \text{sub}(\beta)$ y no es directa. Sin embargo, es fácil ver que β resulta lógicamente equivalente a una expresión normal β' debido a 4 de la Observación 3.13. Además, podemos utilizar a la tautología \top para agregar tests de nodo superfluos. Así llegamos a que si $\beta' := [a] \xrightarrow{*} [b] \xrightarrow{*} [\langle \alpha \neq [\top] \rangle \vee \langle \alpha \neq [\top] \downarrow [\top] \rangle] \downarrow [a] \xrightarrow{*} [\top]$ entonces $\beta \equiv \beta'$. Además, $\text{ad}(\beta) = 4 = \text{ad}(\beta')$.

Proposición 3.17. Para toda φ expresión de nodo y α expresión de camino, existen φ' expresión de nodo normal y α' unión de expresiones de camino normales tales que

$$\begin{aligned}\varphi &\equiv \varphi' && \text{ad}(\varphi) = \text{ad}(\varphi') \\ \alpha &\equiv \alpha' && \text{ad}(\alpha) = \text{ad}(\alpha')\end{aligned}$$

Demostración. Lo probaremos por inducción estructural mutua sobre φ y α .

Expresiones de nodo.

Etiqueta. Si $\varphi = a$ con $a \in \mathbb{L}$ entonces $\varphi' = \varphi$ es normal y $\text{ad}(\varphi') = \text{ad}(\varphi)$.

Negación. Si $\varphi = \neg\psi$ por hipótesis inductiva existe ψ' normal tal que $\psi \equiv \psi'$ y $\text{ad}(\psi) = \text{ad}(\psi')$. Si $\varphi' = \neg\psi'$ entonces $\varphi \equiv \varphi'$ con φ' normal y $\text{ad}(\varphi) = \text{ad}(\varphi')$.

Conjunción. Si $\varphi = \psi_1 \wedge \psi_2$ por hipótesis inductiva existen ψ'_1 y ψ'_2 normales tales que $\psi_i \equiv \psi'_i$ y $\text{ad}(\psi_i) = \text{ad}(\psi'_i)$ para $i = 1, 2$. Luego, $\varphi' = \psi'_1 \wedge \psi'_2$ es normal, equivalente a φ y además, $\text{ad}(\varphi) = \text{ad}(\varphi')$.

Test de datos. Si $\varphi = \langle \beta \odot \gamma \rangle$ con $\odot \in \{=, \neq\}$, por hipótesis inductiva existen $\beta' = \bigcup_{i=1}^M \beta'_i$ y $\gamma' = \bigcup_{j=1}^N \gamma'_j$ tales que

$$\begin{aligned}\beta'_i \text{ normal } \forall i \in \{1, \dots, M\} &&& \gamma'_j \text{ normal } \forall j \in \{1, \dots, N\} \\ \text{ad}(\beta) = \text{ad}(\beta') &&& \text{ad}(\gamma) = \text{ad}(\gamma') \\ = \text{máx}_{1 \leq i \leq M} \{\text{ad}(\beta'_i)\} &&& = \text{máx}_{1 \leq j \leq N} \{\text{ad}(\gamma'_j)\}\end{aligned}$$

Definimos $\varphi' = \bigvee_{i=1}^M \bigvee_{j=1}^N \langle \beta'_i \odot \gamma'_j \rangle$, la cual es normal.

Por un lado, $\varphi \equiv \langle \beta' \odot \gamma' \rangle$ y

$$\begin{aligned} \text{ad}(\varphi) &= \text{máx}\{\text{ad}(\beta), \text{ad}(\gamma)\} \\ &= \text{máx}\{\text{ad}(\beta'), \text{ad}(\gamma')\} = \text{ad}(\langle \beta' \odot \gamma' \rangle) \end{aligned}$$

Por el otro, por 4 de la Observación 3.13, $\langle \beta' \odot \gamma' \rangle \equiv \varphi'$ y

$$\begin{aligned} \text{ad}(\langle \beta' \odot \gamma' \rangle) &= \text{máx}\{\text{ad}(\beta'), \text{ad}(\gamma')\} \\ &= \text{máx}_{\substack{1 \leq i \leq M \\ 1 \leq j \leq N}} \{\text{ad}(\beta'_i), \text{ad}(\gamma'_j)\} = \text{ad}(\varphi') \end{aligned}$$

Juntando todo, $\varphi \equiv \varphi'$ y $\text{ad}(\varphi) = \text{ad}(\varphi')$.

Expresiones de camino.

Caso Base. Si $\alpha \in \{\varepsilon, \downarrow, \overset{*}{\rightarrow}\}$ entonces

$$\varepsilon \equiv [\top] \quad \downarrow \equiv [\top] \downarrow [\top] \quad \overset{*}{\rightarrow} \equiv [\top] \overset{*}{\rightarrow} [\top]$$

Test de nodo. Si $\alpha = [\psi]$ por hipótesis inductiva existe ψ' normal tal que $\psi \equiv \psi'$ con $\text{ad}(\psi) = \text{ad}(\psi')$. Luego, si $\alpha' = [\psi']$ entonces α' es normal tal que $\alpha \equiv \alpha'$ y $\text{ad}(\alpha) = \text{ad}(\alpha')$.

Unión. Si $\alpha = \beta \cup \gamma$ por hipótesis inductiva existen β' y γ' uniones de expresiones normales tales que $\beta \equiv \beta'$, $\gamma \equiv \gamma'$, $\text{ad}(\beta) = \text{ad}(\beta')$ y $\text{ad}(\gamma) = \text{ad}(\gamma')$. Por 1 de la Observación 3.13, si $\alpha' = \beta' \cup \gamma'$ entonces α' es una unión de expresiones normales y $\alpha' \equiv \alpha$. Además, escribiendo explícitamente la unión, es fácil ver que $\text{ad}(\alpha) = \text{ad}(\alpha')$.

Concatenación. Supongamos $\alpha = \beta\gamma$ y $\beta \neq \lambda \neq \gamma$. Por hipótesis inductiva existe $\gamma' = \bigcup_{i=1}^M \gamma'_i$ con γ'_i normal para todo $i \in \{1, \dots, M\}$ tal que $\gamma \equiv \gamma'$ y $\text{ad}(\gamma) = \text{ad}(\gamma')$.

- Si $\beta = \varepsilon$ aplicando 2 de la Observación 3.13 y teniendo en cuenta que $\varepsilon\eta \equiv \eta$ para toda expresión de camino $\eta \neq \lambda$ llegamos a que

$$\begin{aligned} \alpha \equiv \varepsilon \left(\bigcup_{i=1}^M \gamma'_i \right) &\equiv \bigcup_{i=1}^M \varepsilon\gamma'_i \equiv \bigcup_{i=1}^M \gamma'_i = \gamma' \\ \text{ad}(\alpha) &= \text{ad}(\gamma) = \text{ad}(\gamma') \end{aligned}$$

Por lo tanto, para este caso nos sirve $\alpha' = \gamma'$.

- Si $\beta \in \{\downarrow, \overset{*}{\rightarrow}\}$ entonces $\beta \equiv [\top]\beta$. Para cada $i \in \{1, \dots, M\}$ tenemos que $[\top]\beta\gamma'_i$ es normal. Luego,

$$\begin{aligned}\alpha &\equiv [\top]\beta \left(\bigcup_{i=1}^M \gamma'_i \right) \equiv \bigcup_{i=1}^M [\top]\beta\gamma'_i \\ \text{ad}(\alpha) &= 1 + \text{ad}(\gamma) = 1 + \text{ad}(\gamma') \\ &= 1 + \max_{1 \leq i \leq M} \{\text{ad}(\gamma'_i)\} = \max_{1 \leq i \leq M} \{1 + \text{ad}(\gamma'_i)\} \\ &= \text{ad}(\alpha')\end{aligned}$$

- Si $\beta = [\psi]$ por hipótesis inductiva existe ψ' normal tal que $\psi \equiv \psi'$ y $\text{ad}(\psi) = \text{ad}(\psi')$. Cada γ'_i empieza con un test de nodo $[\eta_i]$. Por 3 de la Observación 3.13 tenemos que $[\psi'][\eta_i] \equiv [\psi' \wedge \eta_i]$. Por lo tanto, $[\psi']\gamma'_i \equiv \gamma''_i$ con γ''_i normal y $\text{ad}(\gamma''_i) = \max\{\text{ad}(\psi'), \text{ad}(\gamma'_i)\}$ para todo $i \in \{1, \dots, M\}$. Luego,

$$\begin{aligned}\alpha &\equiv [\psi'] \left(\bigcup_{i=1}^M \gamma'_i \right) \equiv \bigcup_{i=1}^M [\psi']\gamma'_i \equiv \bigcup_{i=1}^M \gamma''_i = \alpha' \\ \text{ad}(\alpha) &= \max\{\text{ad}(\psi), \text{ad}(\gamma)\} = \max\{\text{ad}(\psi'), \text{ad}(\gamma')\} \\ &= \max \left\{ \text{ad}(\psi'), \max_{1 \leq i \leq M} \{\text{ad}(\gamma'_i)\} \right\} \\ &= \max_{1 \leq i \leq M} \left\{ \max\{\text{ad}(\psi'), \text{ad}(\gamma'_i)\} \right\} \\ &= \max_{1 \leq i \leq M} \{\text{ad}(\gamma''_i)\} = \text{ad}(\alpha')\end{aligned}$$

□

3.3. Bisimilaridad

La afirmación de la Proposición 3.17 nos da la pista de cómo adaptar el concepto de bisimulación para el fragmento $\text{XPath}_{=}(\downarrow, \overset{*}{\rightarrow})$.

Definición 3.18. Dados modelos punteados (\mathcal{T}, x) y (\mathcal{T}', x') diremos que son $\{\downarrow, \overset{*}{\rightarrow}\}$ -*equivalentes* y lo notaremos $(\mathcal{T}, x) \equiv^{\{\downarrow, \overset{*}{\rightarrow}\}} (\mathcal{T}', x')$ si satisfacen las mismas fórmulas de $\text{XPath}_{=}(\downarrow, \overset{*}{\rightarrow})$. Es decir, si:

$$(\mathcal{T}, x) \models \varphi \Leftrightarrow (\mathcal{T}', x') \models \varphi \quad \forall \varphi \in \text{XPath}_{=}(\downarrow, \overset{*}{\rightarrow}) \text{ expresión de nodo}$$

Definición 3.19. Dado un árbol ordenado con datos \mathcal{T} , diremos que una secuencia $\tilde{v} = v_0 e_1 v_1 \dots e_n v_n$ es una $\{\downarrow, \overset{*}{\rightarrow}\}$ -*trayectoria* en \mathcal{T} si

$$\begin{aligned}v_i &\in T & \forall i \in \{0, \dots, n\} \\ e_i &\in \{\downarrow, \overset{*}{\rightarrow}\} & \forall i \in \{1, \dots, n\} \\ (\mathcal{T}, v_i, v_{i+1}) &\models e_{i+1} & \forall i \in \{0, \dots, n-1\}\end{aligned}$$

Al n lo llamaremos la *longitud* de \tilde{v} .

Denotaremos por $\{\downarrow, \overset{*}{\rightarrow}\}$ -Tray $_n(\mathcal{T}, x)$ al conjunto de $\{\downarrow, \overset{*}{\rightarrow}\}$ -trayectorias \tilde{v} de longitud n en \mathcal{T} tales que $v_0 = x$.

Definición 3.20. Si $\tilde{v} = v_0 e_1 v_1 \dots e_n v_n$ es una $\{\downarrow, \overset{*}{\rightarrow}\}$ -trayectoria en \mathcal{T} y $\tilde{v}' = v'_0 e'_1 v'_1 \dots e'_n v'_n$ es una $\{\downarrow, \overset{*}{\rightarrow}\}$ -trayectoria en \mathcal{T}' entonces diremos que son *semejantes* y lo notaremos $\tilde{v} \sim \tilde{v}'$ si $e_i = e'_i$ para todo $i \in \{1, \dots, n\}$.

Observación 3.21. Debido a que en un árbol existe a lo sumo un camino (con la relación \downarrow) entre cualquier par de nodos, esto facilita a dar con la noción de \downarrow -bisimulación (Definición 2.8). En cambio, si trabajamos con árboles ordenados, como ahora hay dos relaciones \downarrow y $\overset{*}{\rightarrow}$, pueden haber muchos caminos entre dos nodos. Esto obliga a definir $\{\downarrow, \overset{*}{\rightarrow}\}$ -bisimulación utilizando las $\{\downarrow, \overset{*}{\rightarrow}\}$ -trayectorias.

Definición 3.22. Dados \mathcal{T} y \mathcal{T}' árboles ordenados con datos, diremos que una relación $Z \subseteq T \times T'$ es una $\{\downarrow, \overset{*}{\rightarrow}\}$ -bisimulación si para todo $(x, x') \in Z$ se cumplen las tres siguientes propiedades:

Harmony. $label(x) = label(x')$

Zig. Para todo $\tilde{v} \in \{\downarrow, \overset{*}{\rightarrow}\}$ -Tray $_n(\mathcal{T}, x)$ y $\tilde{w} \in \{\downarrow, \overset{*}{\rightarrow}\}$ -Tray $_m(\mathcal{T}, x)$ existen $\tilde{v}' \in \{\downarrow, \overset{*}{\rightarrow}\}$ -Tray $_n(\mathcal{T}', x')$ y $\tilde{w}' \in \{\downarrow, \overset{*}{\rightarrow}\}$ -Tray $_m(\mathcal{T}', x')$ tales que

- $\tilde{v} \sim \tilde{v}'$ y $\tilde{w} \sim \tilde{w}'$
- $(v_i, v'_i) \in Z \forall i \in \{1, \dots, n\}$
- $(w_i, w'_i) \in Z \forall i \in \{1, \dots, m\}$
- $data(v_n) = data(w_m) \Leftrightarrow data(v'_n) = data(w'_m)$

Zag. Para todo $\tilde{v}' \in \{\downarrow, \overset{*}{\rightarrow}\}$ -Tray $_n(\mathcal{T}', x')$ y $\tilde{w}' \in \{\downarrow, \overset{*}{\rightarrow}\}$ -Tray $_m(\mathcal{T}', x')$ existen $\tilde{v} \in \{\downarrow, \overset{*}{\rightarrow}\}$ -Tray $_n(\mathcal{T}, x)$ y $\tilde{w} \in \{\downarrow, \overset{*}{\rightarrow}\}$ -Tray $_m(\mathcal{T}, x)$ tales que

- $\tilde{v} \sim \tilde{v}'$ y $\tilde{w} \sim \tilde{w}'$
- $(v_i, v'_i) \in Z \forall i \in \{1, \dots, n\}$
- $(w_i, w'_i) \in Z \forall i \in \{1, \dots, m\}$
- $data(v_n) = data(w_m) \Leftrightarrow data(v'_n) = data(w'_m)$

Definición 3.23. Dados dos modelos punteados (\mathcal{T}, x) y (\mathcal{T}', x') diremos que son $\{\downarrow, \overset{*}{\rightarrow}\}$ -bisimilares y lo notaremos $(\mathcal{T}, x) \simeq^{\{\downarrow, \overset{*}{\rightarrow}\}} (\mathcal{T}', x')$ si existe una $\{\downarrow, \overset{*}{\rightarrow}\}$ -bisimulación $Z \subseteq T \times T'$ tal que $(x, x') \in Z$.

Ejemplo 3.24. En la Figura 5 se pueden observar dos modelos punteados (\mathcal{T}, x) y (\mathcal{T}', x') que son $\{\downarrow, \overset{*}{\rightarrow}\}$ -bisimilares por medio de la siguiente $\{\downarrow, \overset{*}{\rightarrow}\}$ -bisimulación $Z := \{(x, x'), (y, y'), (z_1, z'), (z_2, z')\}$.

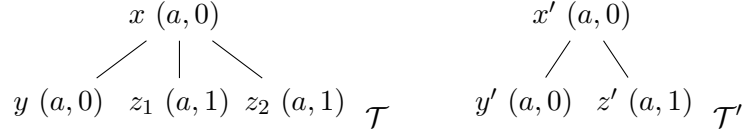


Figura 5: $(\mathcal{T}, x) \simeq^{\{\downarrow, \overset{*}{\rightarrow}\}} (\mathcal{T}', x)$

Observación 3.25. Notar que la noción de $\{\downarrow, \overset{*}{\rightarrow}\}$ -bisimulación es más fuerte que la de \downarrow -bisimulación sobre árboles ordenados. Es decir,

$$(\mathcal{T}, x) \simeq^{\{\downarrow, \overset{*}{\rightarrow}\}} (\mathcal{T}', x') \Rightarrow (\mathcal{T}, x) \simeq^{\downarrow} (\mathcal{T}', x')$$

y de hecho, lo es estrictamente.

En efecto, si consideramos los árboles de la Figura 6 entonces es claro que $Z^{\downarrow} := \{(x, x'), (y, z'), (z, y')\}$ es una \downarrow -bisimulación entre (\mathcal{T}, x) y (\mathcal{T}', x') . Pero no es posible definir una $\{\downarrow, \overset{*}{\rightarrow}\}$ -bisimulación Z tal que $(x, x') \in Z$, pues para la $\{\downarrow, \overset{*}{\rightarrow}\}$ -trayectoria $x \downarrow y \overset{*}{\rightarrow} z$ en $\{\downarrow, \overset{*}{\rightarrow}\}$ -Tray₂(\mathcal{T}, x) hay tres opciones de $\{\downarrow, \overset{*}{\rightarrow}\}$ -trayectorias semejantes en $\{\downarrow, \overset{*}{\rightarrow}\}$ -Tray₂($\mathcal{T}', x')$

$$x' \downarrow y' \overset{*}{\rightarrow} z' \quad x' \downarrow y' \overset{*}{\rightarrow} y' \quad x' \downarrow z' \overset{*}{\rightarrow} z'$$

y por lo tanto debería ocurrir que $(y, y') \in Z$ o $(z, z') \in Z$ pero ambos pares difieren en sus etiquetas, lo cual contradice a Harmony.

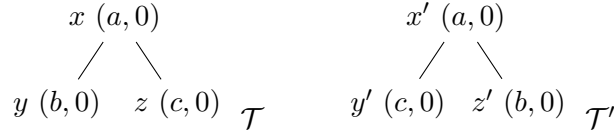


Figura 6: $(\mathcal{T}, x) \simeq^{\downarrow} (\mathcal{T}', x')$ pero $(\mathcal{T}, x) \not\simeq^{\{\downarrow, \overset{*}{\rightarrow}\}} (\mathcal{T}', x')$

3.4. Bisimilaridad Acotada

Definición 3.26. Dos modelos punteados (\mathcal{T}, x) y (\mathcal{T}', x') son $\{\downarrow, \overset{*}{\rightarrow}\}_{\ell}$ -equivalentes y lo notaremos $(\mathcal{T}, x) \equiv_{\ell}^{\{\downarrow, \overset{*}{\rightarrow}\}} (\mathcal{T}', x')$ si satisfacen las mismas expresiones de nodo φ tales que $\text{ad}(\varphi) \leq \ell$. Es decir, si:

$$(\mathcal{T}, x) \models \varphi \Leftrightarrow (\mathcal{T}', x') \models \varphi \quad \forall \varphi \in \ell\text{-XPath}_{=}(\downarrow, \overset{*}{\rightarrow}) \text{ expresión de nodo}$$

Definición 3.27. Dados árboles ordenados \mathcal{T} y \mathcal{T}' , una familia $\mathcal{Z} = \{Z_j\}_{0 \leq j \leq \ell}$ con $Z_k \subseteq T \times T'$ para todo $j \in \{0, \dots, \ell\}$ es una $\{\downarrow, \overset{*}{\rightarrow}\}_{\ell}$ -bisimulación si para todo $j \leq \ell$ y $(x, x') \in Z_j$ se cumplen:

Harmony. $label(x) = label(x')$

Zig. Para todo $\tilde{v} \in \{\downarrow, \rightarrow^*\}$ -Tray $_n(\mathcal{T}, x)$ y $\tilde{w} \in \{\downarrow, \rightarrow^*\}$ -Tray $_m(\mathcal{T}, x)$ con $n, m \leq j$ existen $\tilde{v}' \in \{\downarrow, \rightarrow^*\}$ -Tray $_n(\mathcal{T}', x')$ y $\tilde{w}' \in \{\downarrow, \rightarrow^*\}$ -Tray $_m(\mathcal{T}', x')$ tales que

- $\tilde{v} \sim \tilde{v}'$ y $\tilde{w} \sim \tilde{w}'$
- $(v_i, v'_i) \in Z_{j-i} \forall i \in \{1, \dots, n\}$
- $(w_i, w'_i) \in Z_{j-i} \forall i \in \{1, \dots, m\}$
- $data(v_n) = data(w_m) \Leftrightarrow data(v'_n) = data(w'_m)$

Zag. Para todo $\tilde{v}' \in \{\downarrow, \rightarrow^*\}$ -Tray $_n(\mathcal{T}', x')$ y $\tilde{w}' \in \{\downarrow, \rightarrow^*\}$ -Tray $_m(\mathcal{T}', x')$ con $n, m \leq j$ existen $\tilde{v} \in \{\downarrow, \rightarrow^*\}$ -Tray $_n(\mathcal{T}, x)$ y $\tilde{w} \in \{\downarrow, \rightarrow^*\}$ -Tray $_m(\mathcal{T}, x)$ tales que

- $\tilde{v} \sim \tilde{v}'$ y $\tilde{w} \sim \tilde{w}'$
- $(v_i, v'_i) \in Z_{j-i} \forall i \in \{1, \dots, n\}$
- $(w_i, w'_i) \in Z_{j-i} \forall i \in \{1, \dots, m\}$
- $data(v_n) = data(w_m) \Leftrightarrow data(v'_n) = data(w'_m)$

Definición 3.28. Dos modelos punteados (\mathcal{T}, x) y (\mathcal{T}', x') son $\{\downarrow, \rightarrow^*\}_\ell$ -bisimilares si existe una $\{\downarrow, \rightarrow^*\}_\ell$ -bisimulación $\mathcal{Z} = \{Z_j\}_{0 \leq j \leq \ell}$ tal que $(x, x') \in Z_\ell$. En tal caso, escribiremos $(\mathcal{T}, x) \simeq_{\{\downarrow, \rightarrow^*\}_\ell} (\mathcal{T}', x')$.

Definición 3.29. Sean $\tilde{v} \in \{\downarrow, \rightarrow^*\}$ -Tray $_n(\mathcal{T}, x)$ y α expresión de camino tales que

$$\begin{aligned} \alpha &= [\psi_0]o_1[\psi_1] \dots o_n[\psi_n] & o_i &\in \{\downarrow, \rightarrow^*\} \\ \tilde{v} &= v_0e_1v_1 \dots e_nv_n \end{aligned}$$

entonces diremos que el par (\mathcal{T}, \tilde{v}) *satisface* α y escribiremos $(\mathcal{T}, \tilde{v}) \models \alpha$ si se cumple que

$$\begin{aligned} o_i &= e_i & \forall i &\in \{1, \dots, n\} \\ (\mathcal{T}, v_i) &\models \psi_i & \forall i &\in \{0, \dots, n\} \end{aligned}$$

Lema 3.30. Sean \mathcal{T} y \mathcal{T}' árboles ordenados con datos. Supongamos $x \in \mathcal{T}$ y $x' \in \mathcal{T}'$ tales que $label(x) = label(x')$. Entonces, para toda φ expresión de nodo normal tal que $ad(\varphi) = 0$ se cumple que

$$(\mathcal{T}, x) \models \varphi \iff (\mathcal{T}', x') \models \varphi \quad (1)$$

Demostración. Lo probaremos por inducción estructural sobre φ .

Etiqueta. Como $label(x) = label(x')$ entonces vale **1** para todo $\varphi = a$ con $a \in \mathbb{L}$.

Negación. Si ψ es normal con $ad(\psi) = 0$ y cumple **1** entonces $\varphi = \neg\psi$ es normal con $ad(\varphi) = 0$ y también cumple **1**.

Conjunción. Si φ_1 y φ_2 son normales con $ad(\varphi_1) = ad(\varphi_2) = 0$ y cumplen **1** entonces $\varphi = \varphi_1 \wedge \varphi_2$ es normal con $ad(\varphi) = 0$ y también cumple **1**.

Test de datos. Supongamos $\varphi = \langle \alpha \odot \beta \rangle$ con $\odot \in \{=, \neq\}$. Si φ es normal, entonces α y β son normales. Además, si $ad(\varphi) = 0$ tenemos que $ad(\alpha) = ad(\beta) = 0$. Necesariamente, $\alpha = [\varphi_1]$ y $\beta = [\varphi_2]$. Luego, φ_1 y φ_2 son normales y $ad(\varphi_1) = ad(\varphi_2) = 0$. Por hipótesis inductiva,

$$(\mathcal{T}, x) \models \varphi_i \Leftrightarrow (\mathcal{T}', x') \models \varphi_i \quad i = 1, 2$$

Esto implica que $(\mathcal{T}, x) \models \varphi \Leftrightarrow (\mathcal{T}', x') \models \varphi$

□

Lema 3.31. Sea $\mathcal{Z} = \{Z_j\}_{0 \leq j \leq \ell}$ una $\{\downarrow, \overset{*}{\rightarrow}\}_\ell$ -bisimulación entre (\mathcal{T}, x) y (\mathcal{T}', x') .

Para todo $0 \leq j \leq \ell$, α normal y φ normal tales que $ad(\alpha) \leq j$ y $ad(\varphi) \leq j$ se tiene que

1. Si $(x, x') \in Z_j$ entonces $(\mathcal{T}, x) \models \varphi \Leftrightarrow (\mathcal{T}', x') \models \varphi$.
2. Si $\tilde{v} \in \{\downarrow, \overset{*}{\rightarrow}\}\text{-Tray}_n(\mathcal{T}, x)$, $\tilde{v}' \in \{\downarrow, \overset{*}{\rightarrow}\}\text{-Tray}_n(\mathcal{T}', x')$, $\tilde{v} \sim \tilde{v}'$ y $(v_i, v'_i) \in Z_{j-i}$ para todo $i \in \{0, \dots, n\}$ entonces $(\mathcal{T}, \tilde{v}) \models \alpha \Leftrightarrow (\mathcal{T}', \tilde{v}') \models \alpha$.

Demostración. Haremos inducción sobre $j \leq \ell$.

- Si $j = 0$ entonces $ad(\varphi) = ad(\alpha) = 0$. Necesariamente, $\alpha = [\psi]$ con $ad(\psi) = 0$. Además, ψ es normal pues α lo es. Como $(x, x') \in Z_0$, por **Harmony**, $label(x) = label(x')$. Aplicando el Lema 3.30 tenemos probado **1** y **2**.
- Supongamos $j > 0$. Probaremos **1** y **2** por inducción estructural mutua sobre φ y α .
 2. Supongamos primero que $\alpha = [\varphi]$ normal tal que $ad(\alpha) \leq j$. Notar que φ es normal y $ad(\varphi) \leq j$. Supongamos $\tilde{v} \in \{\downarrow, \overset{*}{\rightarrow}\}\text{-Tray}_n(\mathcal{T}, x)$, $\tilde{v}' \in \{\downarrow, \overset{*}{\rightarrow}\}\text{-Tray}_n(\mathcal{T}', x')$, $\tilde{v} \sim \tilde{v}'$ y $(v_i, v'_i) \in Z_{j-i}$ para todo $i \in \{0, \dots, n\}$. Si $(\mathcal{T}, \tilde{v}) \models \alpha$ necesariamente $n = 0$ y entonces $v = x$, $v' = x'$ y $(x, x') \in Z_j$. Como $(\mathcal{T}, x) \models \varphi$, por hipótesis inductiva, aplicando el ítem **1** sobre φ , tenemos que $(\mathcal{T}', x') \models \varphi$. Por lo tanto, $(\mathcal{T}', v') \models \alpha$.

Supongamos ahora que $\alpha = [\psi_0]o_1[\psi_1] \dots o_n[\psi_n]$ con $o_i \in \{\downarrow, \overset{*}{\rightarrow}\}$ para todo $i \in \{1, \dots, n\}$ y $\text{ad}(\alpha) \leq j$. Supongamos que existen $\tilde{v} \in \{\downarrow, \overset{*}{\rightarrow}\}\text{-Tray}_n(\mathcal{T}, x)$, $\tilde{v}' \in \{\downarrow, \overset{*}{\rightarrow}\}\text{-Tray}_n(\mathcal{T}', x')$, $\tilde{v} \sim \tilde{v}'$ y $(v_i, v'_i) \in Z_{j-i}$ para todo $i \in \{0, \dots, n\}$ tales que $(\mathcal{T}, \tilde{v}) \models \alpha$ pero $(\mathcal{T}', \tilde{v}') \not\models \alpha$. Entonces, necesariamente, existe $k \in \{1, \dots, n\}$ ($k = 0$ es el caso anterior) tal que $(\mathcal{T}, v_k) \models \psi_k$ pero $(\mathcal{T}', v'_k) \not\models \psi_k$. Notar que ψ_k es normal, $\text{ad}(\psi_k) \leq j - k$ y $(v_k, v'_k) \in Z_{j-k}$. Como $0 \leq j - k < j$, por hipótesis inductiva aplicada sobre $j - k$, tenemos que por debería pasar que $(\mathcal{T}, v_k) \models \psi_k$, lo cual es absurdo. Esto vino de suponer que $(\mathcal{T}', v'_k) \not\models \psi_k$. Por lo tanto, $(\mathcal{T}, v_k) \models \psi_k$.

1. Si $\varphi = a \in \mathbb{L}$ entonces como $(x, x') \in Z_j$ por **Harmony** tenemos que $\text{label}(x) = \text{label}(x')$.

Si ψ y φ cumplen **1** entonces también lo harán $\neg\psi$ y $\psi \wedge \varphi$.

Veamos $\varphi = \langle \beta \odot \gamma \rangle$ con $\odot \in \{=, \neq\}$. Notar que β y γ son expresiones normales con $\max\{\text{ad}(\beta), \text{ad}(\gamma)\} \leq j$. Si $(\mathcal{T}, x) \models \varphi$ entonces existen $\tilde{v} \in \{\downarrow, \overset{*}{\rightarrow}\}\text{-Tray}_n(\mathcal{T}, x)$ y $\tilde{w} \in \{\downarrow, \overset{*}{\rightarrow}\}\text{-Tray}_m(\mathcal{T}, x)$ tales que $(\mathcal{T}, \tilde{v}) \models \beta$, $(\mathcal{T}, \tilde{w}) \models \gamma$ y $\text{data}(v_n) \odot \text{data}(w_m)$ (notar que necesariamente $n, m \leq j$). Como $(x, x') \in Z_j$ por **Zig** existen $\tilde{v}' \in \{\downarrow, \overset{*}{\rightarrow}\}\text{-Tray}_n(\mathcal{T}', x')$ y $\tilde{w}' \in \{\downarrow, \overset{*}{\rightarrow}\}\text{-Tray}_m(\mathcal{T}', x')$ tales que $\tilde{v} \sim \tilde{v}'$, $\tilde{w} \sim \tilde{w}'$, $(v_i, v'_i) \in Z_{j-i}$ para todo $i \in \{0, \dots, n\}$, $(w_i, w'_i) \in Z_{j-i}$ para todo $i \in \{0, \dots, m\}$ y $\text{data}(v'_n) \odot \text{data}(w'_m)$. Por hipótesis inductiva, aplicando **1** sobre β y γ , tenemos que $(\mathcal{T}', \tilde{v}') \models \beta$ y $(\mathcal{T}', \tilde{w}') \models \gamma$. Por lo tanto, $(\mathcal{T}', x') \models \varphi$. Análogamente, si $(\mathcal{T}', x') \models \varphi$ por **Zag** llegamos a que $(\mathcal{T}, x) \models \varphi$.

□

Proposición 3.32. Dos modelos punteados $\{\downarrow, \overset{*}{\rightarrow}\}_\ell$ -bisimilares son $\{\downarrow, \overset{*}{\rightarrow}\}_\ell$ -equivalentes.

$$(\mathcal{T}, x) \simeq_\ell^{\{\downarrow, \overset{*}{\rightarrow}\}} (\mathcal{T}', x') \implies (\mathcal{T}, x) \equiv_\ell^{\{\downarrow, \overset{*}{\rightarrow}\}} (\mathcal{T}', x')$$

Demostración. Sea $\mathcal{Z} = \{Z_j\}_{0 \leq j \leq \ell}$ una $\{\downarrow, \overset{*}{\rightarrow}\}_\ell$ -bisimulación entre (\mathcal{T}, x) y (\mathcal{T}', x') . Queremos probar que para toda $\varphi \in \ell\text{-XPath}_=(\downarrow, \overset{*}{\rightarrow})$ se cumple que

$$(\mathcal{T}, x) \models \varphi \iff (\mathcal{T}', x') \models \varphi$$

Supongamos $(\mathcal{T}, x) \models \varphi$ entonces por la Proposición 3.17 existe φ' normal tal que $\varphi \equiv \varphi'$ y $\text{ad}(\varphi) = \text{ad}(\varphi')$. Luego, $(\mathcal{T}, x) \models \varphi'$. Como $(x, x') \in Z_\ell$, por **2** del Lema 3.31, tenemos que $(\mathcal{T}', x') \models \varphi'$. Por lo tanto, $(\mathcal{T}', x') \models \varphi$.

Análogamente, si $(\mathcal{T}', x') \models \varphi$ entonces $(\mathcal{T}, x) \models \varphi$. □

Observación 3.33. Dada una familia finita de expresiones de nodo $\mathcal{N} = \{\varphi_1, \dots, \varphi_n\}$, si denotamos por $\mathcal{B}(\mathcal{N})$ al conjunto de combinaciones booleanas generadas por los elementos de \mathcal{N} , entonces, de la misma manera que

se prueba en el contexto de la lógica proposicional, $\mathcal{B}(\mathcal{N})$ es finito módulo equivalencia lógica.

Lema 3.34. Existe una cantidad finita de expresiones de nodo φ tales que $\text{ad}(\varphi) \leq \ell$, salvo equivalencia lógica.

Demostración. Lo probaremos por inducción sobre ℓ .

Si $\ell = 0$ entonces cada test de datos es de la forma $\langle [\varphi] = [\psi] \rangle$ o $\langle [\varphi] \neq [\psi] \rangle$. Las primeras son equivalentes a $\varphi \wedge \psi$, y las otras, a $\perp := \neg \top$. Esto nos dice que podemos evitar el uso de tests de datos. Luego, cualquier expresión de nodo es equivalente a una combinación booleana de etiquetas. Como la cantidad de etiquetas es finita, por la Observación 3.33 existe una cantidad finita de expresiones de nodo η , salvo equivalencia lógica.

Si $\ell > 1$, cada $\varphi \in \ell\text{-XPath}_{=, \neq}$ es equivalente a una φ' normal tal que $\text{ad}(\varphi') \leq \ell$ (ver Proposición 3.17). Si φ' es un test de datos entonces $\varphi' = \langle [\psi_1]o_1\alpha_1 \odot [\psi_2]o_2\alpha_2 \rangle$ con $\odot \in \{=, \neq\}$, $o_1, o_2 \in \{\downarrow, \xrightarrow{*}\}$ y α_1, α_2 expresiones de camino normales. Notar que:

$$\begin{aligned} \varphi' &= \langle [\psi_1]o_1\alpha_1 \odot [\psi_2]o_2\alpha_2 \rangle \equiv \psi_1 \wedge \psi_2 \wedge \varphi'' \\ \varphi'' &= \langle o_1\alpha_1 \odot o_2\alpha_2 \rangle \qquad \text{máx}\{\text{ad}(\alpha_1), \text{ad}(\alpha_2)\} < \ell \end{aligned}$$

Esto implica que toda expresión de nodo es equivalente a una combinación booleana de etiquetas y tests de datos que tienen la misma forma que φ'' . Como por hipótesis inductiva hay una cantidad finita (salvo equivalencia lógica) de expresiones φ'' , llegamos a que por la Observación 3.33 existen finitas expresiones de nodo η con $\text{ad}(\eta) \leq \ell$, módulo equivalencia lógica.

□

Definición 3.35. Dado $\tilde{v} \in \{\downarrow, \xrightarrow{*}\}\text{-Tray}_n(\mathcal{T}, x)$ tal que $\tilde{v} = v_0e_1v_1 \dots e_nv_n$ definimos $\alpha_{\tilde{v}}$ la expresión (normal) de camino asociada a \tilde{v} como

$$\alpha_{\tilde{v}} := [\top]e_1[\top] \dots [\top]e_n[\top]$$

Notar que $(\mathcal{T}, \tilde{v}) \models \alpha_{\tilde{v}}$.

Proposición 3.36. Dos modelos punteados $\{\downarrow, \xrightarrow{*}\}_\ell$ -equivalentes son $\{\downarrow, \xrightarrow{*}\}_\ell$ -bisimilares. Es decir:

$$(\mathcal{T}, x) \equiv_{\ell}^{\{\downarrow, \xrightarrow{*}\}} (\mathcal{T}', x') \implies (\mathcal{T}, x) \simeq_{\ell}^{\{\downarrow, \xrightarrow{*}\}} (\mathcal{T}', x')$$

Demostración. Para todo $i \in \{0, \dots, \ell\}$ definimos las relaciones $Z_i \subseteq T \times T'$ de la siguiente manera

$$(y, y') \in Z_i \quad \stackrel{\text{def}}{\iff} \quad (\mathcal{T}, y) \equiv_i^{\{\downarrow, \overset{*}{\rightarrow}\}} (\mathcal{T}', y')$$

Probaremos que $\mathcal{Z} := \{Z_j\}_{0 \leq j \leq \ell}$ es una ℓ -bisimulación entre (\mathcal{T}, x) y (\mathcal{T}', x') . Notar que por hipótesis tenemos que $(x, x') \in Z_\ell$.

Supongamos que $(y, y') \in Z_h$.

Harmony. Supongamos que $\text{label}(y) = a$ y llamemos $\varphi = a$. Como $(\mathcal{T}, y) \models \varphi$ y $(\mathcal{T}, y) \equiv_h^{\{\downarrow, \overset{*}{\rightarrow}\}} (\mathcal{T}', y')$, debido a que $\text{ad}(\varphi) = 0 \leq h$ tenemos que $(\mathcal{T}', y') \models \varphi$.

Zig. Supongamos $n, m \leq h$. Sea $\tilde{v} \in \{\downarrow, \overset{*}{\rightarrow}\}\text{-Tray}_n(\mathcal{T}, y)$ y $\tilde{w} \in \{\downarrow, \overset{*}{\rightarrow}\}\text{-Tray}_m(\mathcal{T}, y)$ tales que $\text{data}(v_n) \odot \text{data}(w_m)$ con $\odot \in \{=, \neq\}$. Consideremos el siguiente conjunto

$$P := \{(\tilde{v}', \tilde{w}') \in \{\downarrow, \overset{*}{\rightarrow}\}\text{-Tray}_n(\mathcal{T}', y') \times \{\downarrow, \overset{*}{\rightarrow}\}\text{-Tray}_m(\mathcal{T}', y') : \\ \tilde{v} \sim \tilde{v}' \wedge \tilde{w} \sim \tilde{w}' \wedge \text{data}(v'_n) \odot \text{data}(w'_m)\}$$

Veamos que P es no vacío. Si $\alpha_{\tilde{v}}$ y $\alpha_{\tilde{w}}$ son las expresiones de camino asociadas a \tilde{v} y \tilde{w} , entonces como $(\mathcal{T}, y) \models \langle \alpha_{\tilde{v}} \odot \alpha_{\tilde{w}} \rangle$ y $\text{ad}(\langle \alpha_{\tilde{v}} \odot \alpha_{\tilde{w}} \rangle) \leq h$, dado que $(\mathcal{T}, y) \equiv_h^{\{\downarrow, \overset{*}{\rightarrow}\}} (\mathcal{T}', y')$ tenemos que $(\mathcal{T}', y') \models \langle \alpha_{\tilde{v}} \odot \alpha_{\tilde{w}} \rangle$. Por lo tanto, existen $\tilde{v}' \in \{\downarrow, \overset{*}{\rightarrow}\}\text{-Tray}_n(\mathcal{T}', y')$ y $\tilde{w}' \in \{\downarrow, \overset{*}{\rightarrow}\}\text{-Tray}_m(\mathcal{T}', y')$ tales que $\tilde{v} \sim \tilde{v}'$, $\tilde{w} \sim \tilde{w}'$, $(\mathcal{T}', \tilde{v}') \models \alpha_{\tilde{v}}$, $(\mathcal{T}', \tilde{w}') \models \alpha_{\tilde{w}}$ y $\text{data}(v'_n) \odot \text{data}(w'_m)$. O sea, $(\tilde{v}', \tilde{w}') \in P \neq \emptyset$.

Supongamos que no vale **Zig** en (y, y') . Luego, por definición de \mathcal{Z} , para todo $(\tilde{v}', \tilde{w}') \in P$ tenemos que alguna de las siguientes no se cumple

$$\exists i \in \{0, \dots, n\} : (\mathcal{T}, v_i) \not\equiv_{h-i}^{\{\downarrow, \overset{*}{\rightarrow}\}} (\mathcal{T}', v'_i) \quad (2)$$

$$\exists j \in \{0, \dots, m\} : (\mathcal{T}, w_j) \not\equiv_{h-j}^{\{\downarrow, \overset{*}{\rightarrow}\}} (\mathcal{T}', w'_j) \quad (3)$$

Ahora, para cada $(\tilde{v}', \tilde{w}') \in P$ definimos expresiones de nodo

$$\begin{array}{ll} \varphi_{(\tilde{v}', \tilde{w}')}^0, \varphi_{(\tilde{v}', \tilde{w}')}^1, \dots, \varphi_{(\tilde{v}', \tilde{w}')}^n & \psi_{(\tilde{v}', \tilde{w}')}^0, \psi_{(\tilde{v}', \tilde{w}')}^1, \dots, \psi_{(\tilde{v}', \tilde{w}')}^m \\ \text{ad}(\varphi_{(\tilde{v}', \tilde{w}')}^i) \leq h - i & \text{ad}(\psi_{(\tilde{v}', \tilde{w}')}^j) \leq h - j \\ \forall i \in \{0, \dots, n\} & \forall j \in \{0, \dots, m\} \end{array}$$

de la siguiente manera:

- Supongamos que vale **2** entonces consideremos el mínimo i tal que $(\mathcal{T}, v_i) \not\equiv_i^{\{\downarrow, \overset{*}{\rightarrow}\}} (\mathcal{T}', v'_i)$. Luego, existe $\varphi_{(\tilde{v}', \tilde{w}')}^i$ con $\text{ad}(\varphi_{(\tilde{v}', \tilde{w}')}^i) \leq h - i$ tal que $(\mathcal{T}, v_i) \models \varphi_{(\tilde{v}', \tilde{w}')}^i$ pero $(\mathcal{T}', v'_i) \not\models \varphi_{(\tilde{v}', \tilde{w}')}^i$. Para el resto de los $k \in \{0, \dots, n\} \setminus \{i\}$ y $j \in \{0, \dots, m\}$ definimos $\varphi_{(\tilde{v}', \tilde{w}')}^k := \top$ y $\psi_{(\tilde{v}', \tilde{w}')}^j := \top$.
- Si no vale **2** entonces vale **3**. Consideremos el mínimo j tal que $(\mathcal{T}, w_j) \not\equiv_j^{\{\downarrow, \overset{*}{\rightarrow}\}} (\mathcal{T}', w'_j)$. Luego, existe $\psi_{(\tilde{v}', \tilde{w}')}^j$ con $\text{ad}(\psi_{(\tilde{v}', \tilde{w}')}^j) \leq h - j$ tal que $(\mathcal{T}, w_j) \models \psi_{(\tilde{v}', \tilde{w}')}^j$ pero $(\mathcal{T}', w'_j) \not\models \psi_{(\tilde{v}', \tilde{w}')}^j$. Para el resto de los $k \in \{0, \dots, m\} \setminus \{j\}$ e $i \in \{0, \dots, n\}$ definimos $\psi_{(\tilde{v}', \tilde{w}')}^k := \top$ y $\varphi_{(\tilde{v}', \tilde{w}')}^i := \top$.

Luego, para cada $i \in \{0, \dots, n\}$ y $j \in \{0, \dots, m\}$ definimos

$$\Phi^i := \bigwedge_{(\tilde{v}', \tilde{w}') \in P} \varphi_{(\tilde{v}', \tilde{w}')}^i \quad \Psi^j := \bigwedge_{(\tilde{v}', \tilde{w}') \in P} \psi_{(\tilde{v}', \tilde{w}')}^j$$

Notar que P puede ser infinito, pero por el Lema **3.34**, tanto Φ^i como Ψ^j están bien definidas, módulo equivalencia lógica. Finalmente, si $\tilde{v} = v_0 e_1 v_1 \dots e_n v_n$ y $\tilde{w} = w_0 o_1 w_1 \dots o_m w_m$ definimos

$$\begin{aligned} \alpha &:= [\Phi^0] e_1 [\Phi^1] \dots e_n [\Phi^n] & (\mathcal{T}, \tilde{v}) &\models \alpha \\ \beta &:= [\Psi^0] o_1 [\Psi^1] \dots o_m [\Psi^m] & (\mathcal{T}, \tilde{w}) &\models \beta \end{aligned}$$

Por lo tanto, $(\mathcal{T}, y) \models \langle \alpha \odot \beta \rangle$. Como $\text{ad}(\langle \alpha \odot \beta \rangle) \leq h$ y $(\mathcal{T}, y) \equiv_h^{\{\downarrow, \overset{*}{\rightarrow}\}} (\mathcal{T}', y')$ entonces $(\mathcal{T}', y') \models \langle \alpha \odot \beta \rangle$. Luego, existe $(\tilde{v}', \tilde{w}') \in P$ tal que $(\mathcal{T}', \tilde{v}') \models \alpha$ y $(\mathcal{T}', \tilde{w}') \models \beta$. Pero, por construcción, $(\mathcal{T}', \tilde{v}') \not\models \alpha$ o $(\mathcal{T}', \tilde{w}') \not\models \beta$. En cualquiera de los casos, llegamos a un absurdo, el cual vino de suponer que no valía **Zig**.

Zag. Análogo a **Zig**. □

Teorema 3.37. Dos modelos punteados son $\{\downarrow, \overset{*}{\rightarrow}\}_\ell$ -equivalentes, si y sólo si, son $\{\downarrow, \overset{*}{\rightarrow}\}_\ell$ -bisimilares

$$(\mathcal{T}, x) \simeq_\ell^{\{\downarrow, \overset{*}{\rightarrow}\}} (\mathcal{T}', x') \iff (\mathcal{T}, x) \equiv_\ell^{\{\downarrow, \overset{*}{\rightarrow}\}} (\mathcal{T}', x')$$

Demostración. Se deduce de los resultados anteriores.

(\Rightarrow) Por la Proposición **3.32**.

(\Leftarrow) Por la Proposición **3.36**. □

3.5. Caracterización al estilo Hennessy-Milner

Proposición 3.38. Sean \mathcal{T} y \mathcal{T}' árboles ordenados finitamente ramificados. Entonces

$$(\mathcal{T}, x) \equiv^{\{\downarrow, \overset{*}{\rightarrow}\}} (\mathcal{T}', x') \implies (\mathcal{T}, x) \simeq^{\{\downarrow, \overset{*}{\rightarrow}\}} (\mathcal{T}', x')$$

Demostración. Definamos $Z := \{(x, x') \in T \times T' : (\mathcal{T}, x) \equiv^{\{\downarrow, \overset{*}{\rightarrow}\}} (\mathcal{T}', x')\}$. Notar que por hipótesis tenemos que $(x, x') \in Z$. Vamos a probar que Z es una $\{\downarrow, \overset{*}{\rightarrow}\}$ -bisimulación.

Sea $(y, y') \in Z$.

Harmony. Supongamos que $label(y) = a$ y llamemos $\varphi = a$. Como $(\mathcal{T}, y) \models \varphi$ y $(\mathcal{T}, y) \equiv^{\{\downarrow, \overset{*}{\rightarrow}\}} (\mathcal{T}', y')$ entonces $(\mathcal{T}', y') \models \varphi$. Por lo tanto, $label(y) = label(y')$.

Zig. Sea $\tilde{v} \in \{\downarrow, \overset{*}{\rightarrow}\}$ -Tray $_n(\mathcal{T}, y)$ y $\tilde{w} \in \{\downarrow, \overset{*}{\rightarrow}\}$ -Tray $_m(\mathcal{T}, y)$ tales que $data(v_n) \odot data(w_m)$ con $\odot \in \{=, \neq\}$. Consideremos el siguiente conjunto

$$P := \{(\tilde{v}', \tilde{w}') \in \{\downarrow, \overset{*}{\rightarrow}\}$$
-Tray $_n(\mathcal{T}', y') \times \{\downarrow, \overset{*}{\rightarrow}\}$ -Tray $_m(\mathcal{T}', y') : \tilde{v} \sim \tilde{v}' \wedge \tilde{w} \sim \tilde{w}' \wedge data(v'_n) \odot data(w'_m)\}$

Veamos que P es no vacío. Si $\alpha_{\tilde{v}}$ y $\alpha_{\tilde{w}}$ son las expresiones de camino asociadas a \tilde{v} y \tilde{w} , entonces como $(\mathcal{T}, y) \models \langle \alpha_{\tilde{v}} \odot \alpha_{\tilde{w}} \rangle$ y $(\mathcal{T}, y) \equiv^{\{\downarrow, \overset{*}{\rightarrow}\}} (\mathcal{T}', y')$ tenemos que $(\mathcal{T}', y') \models \langle \alpha_{\tilde{v}} \odot \alpha_{\tilde{w}} \rangle$. Por lo tanto, existen $\tilde{v}' \in \{\downarrow, \overset{*}{\rightarrow}\}$ -Tray $_n(\mathcal{T}', y')$ y $\tilde{w}' \in \{\downarrow, \overset{*}{\rightarrow}\}$ -Tray $_m(\mathcal{T}', y')$ tales que $\tilde{v} \sim \tilde{v}'$, $\tilde{w} \sim \tilde{w}'$, $(\mathcal{T}', \tilde{v}') \models \alpha_{\tilde{v}}$, $(\mathcal{T}', \tilde{w}') \models \alpha_{\tilde{w}}$ y $data(v'_n) \odot data(w'_m)$. O sea, $(\tilde{v}', \tilde{w}') \in P \neq \emptyset$.

Supongamos que no vale **Zig** en (y, y') . Luego, por definición de Z , para todo $(\tilde{v}', \tilde{w}') \in P$ tenemos que alguna de las siguientes no se cumple

$$\exists i \in \{0, \dots, n\} : (\mathcal{T}, v_i) \not\equiv^{\{\downarrow, \overset{*}{\rightarrow}\}} (\mathcal{T}', v'_i) \quad (4)$$

$$\exists j \in \{0, \dots, m\} : (\mathcal{T}, w_j) \not\equiv^{\{\downarrow, \overset{*}{\rightarrow}\}} (\mathcal{T}', w'_j) \quad (5)$$

Ahora, para cada $(\tilde{v}', \tilde{w}') \in P$ definimos expresiones de nodo

$$\varphi_{(\tilde{v}', \tilde{w}')}^0, \varphi_{(\tilde{v}', \tilde{w}')}^1, \dots, \varphi_{(\tilde{v}', \tilde{w}')}^n \quad \psi_{(\tilde{v}', \tilde{w}')}^0, \psi_{(\tilde{v}', \tilde{w}')}^1, \dots, \psi_{(\tilde{v}', \tilde{w}')}^m$$

de la siguiente manera:

- Supongamos que vale 4 entonces consideremos el mínimo i tal que $(\mathcal{T}, v_i) \not\equiv^{\{\downarrow, \overset{*}{\rightarrow}\}} (\mathcal{T}', v'_i)$. Luego, existe $\varphi_{(\tilde{v}', \tilde{w}')}^i$ tal que $(\mathcal{T}, v_i) \models \varphi_{(\tilde{v}', \tilde{w}')}^i$ pero $(\mathcal{T}', v'_i) \not\models \varphi_{(\tilde{v}', \tilde{w}')}^i$. Para el resto de los $k \in \{0, \dots, n\} \setminus \{i\}$ y $j \in \{0, \dots, m\}$ definimos $\varphi_{(\tilde{v}', \tilde{w}')}^k := \top$ y $\psi_{(\tilde{v}', \tilde{w}')}^j := \top$.

- Si no vale 4 entonces vale 5. Consideremos el mínimo j tal que $(\mathcal{T}, w_j) \not\equiv^{\{\downarrow, \overset{*}{\rightarrow}\}} (\mathcal{T}', w'_j)$. Luego, existe $\psi_{(\tilde{v}', \tilde{w}')}^j$ tal que $(\mathcal{T}, w_j) \models \psi_{(\tilde{v}', \tilde{w}')}^j$ pero $(\mathcal{T}', w'_j) \not\models \psi_{(\tilde{v}', \tilde{w}')}^j$. Para el resto de los $k \in \{0, \dots, m\} \setminus \{j\}$ e $i \in \{0, \dots, n\}$ definimos $\psi_{(\tilde{v}', \tilde{w}')}^k := \top$ y $\varphi_{(\tilde{v}', \tilde{w}')}^i := \top$.

Luego, para cada $i \in \{0, \dots, n\}$ y $j \in \{0, \dots, m\}$ definimos

$$\Phi^i := \bigwedge_{(\tilde{v}', \tilde{w}') \in P} \varphi_{(\tilde{v}', \tilde{w}')}^i \quad \Psi^j := \bigwedge_{(\tilde{v}', \tilde{w}') \in P} \psi_{(\tilde{v}', \tilde{w}')}^j$$

Como P es finito debido a que \mathcal{T}' es un árbol finitamente ramificado, tanto Φ^i como Ψ^j están bien definidas (notar que esta es la diferencia principal con la demostración de la Proposición 3.36). Finalmente, si $\tilde{v} = v_0 e_1 v_1 \dots e_n v_n$ y $\tilde{w} = w_0 o_1 w_1 \dots o_m w_m$ definimos

$$\begin{aligned} \alpha &:= [\Phi^0] e_1 [\Phi^1] \dots e_n [\Phi^n] & (\mathcal{T}, \tilde{v}) &\models \alpha \\ \beta &:= [\Psi^0] o_1 [\Psi^1] \dots o_m [\Psi^m] & (\mathcal{T}, \tilde{w}) &\models \beta \end{aligned}$$

Por lo tanto, $(\mathcal{T}, y) \models \langle \alpha \odot \beta \rangle$. Como $(\mathcal{T}, y) \equiv^{\{\downarrow, \overset{*}{\rightarrow}\}} (\mathcal{T}', y')$ entonces $(\mathcal{T}', y') \models \langle \alpha \odot \beta \rangle$. Luego, existe $(\tilde{v}', \tilde{w}') \in P$ tal que $(\mathcal{T}', \tilde{v}') \models \alpha$ y $(\mathcal{T}', \tilde{w}') \models \beta$. Pero, por construcción, $(\mathcal{T}', \tilde{v}') \not\models \alpha$ o $(\mathcal{T}', \tilde{w}') \not\models \beta$. En cualquiera de los casos, llegamos a un absurdo, el cual vino de suponer que no valía **Zig**.

Zag. Análogo a **Zig**.

□

Observación 3.39. En general, $\{\downarrow, \overset{*}{\rightarrow}\}$ -equivalencia no implica $\{\downarrow, \overset{*}{\rightarrow}\}$ -bisimilaridad. Si consideramos los árboles de la Figura 4 entonces los modelos (\mathcal{T}, u) y (\mathcal{T}', u') resultan ser $\{\downarrow, \overset{*}{\rightarrow}\}$ -equivalentes. Si fuesen $\{\downarrow, \overset{*}{\rightarrow}\}$ -bisimilares necesariamente serían \downarrow -bisimilares, lo cual es falso.

Observación 3.40. Notar que si Z es una $\{\downarrow, \overset{*}{\rightarrow}\}$ -bisimulación entonces la familia $\mathcal{Z} = (Z_j)_{j \leq \ell}$ con $Z_j := Z$ es una $\{\downarrow, \overset{*}{\rightarrow}\}_\ell$ -bisimulación. Por lo tanto, tenemos que para todo ℓ se cumple que

$$(\mathcal{T}, x) \simeq^{\{\downarrow, \overset{*}{\rightarrow}\}} (\mathcal{T}', x') \implies (\mathcal{T}, x) \simeq_\ell^{\{\downarrow, \overset{*}{\rightarrow}\}} (\mathcal{T}', x')$$

Teorema 3.41. Sean (\mathcal{T}, x) y (\mathcal{T}', x') modelos punteados, entonces

$$(\mathcal{T}, x) \simeq^{\{\downarrow, \overset{*}{\rightarrow}\}} (\mathcal{T}', x') \implies (\mathcal{T}, x) \equiv^{\{\downarrow, \overset{*}{\rightarrow}\}} (\mathcal{T}', x')$$

Demostración. Si $(\mathcal{T}, x) \simeq^{\{\downarrow, \overset{*}{\rightarrow}\}} (\mathcal{T}', x')$ entonces por la Observación 3.40, para todo $\ell \in \mathbb{N}_0$ tenemos que $(\mathcal{T}, x) \simeq_\ell^{\{\downarrow, \overset{*}{\rightarrow}\}} (\mathcal{T}', x')$. Y esto es lo mismo que para todo $\ell \in \mathbb{N}_0$ tengamos que $(\mathcal{T}, x) \equiv_\ell^{\{\downarrow, \overset{*}{\rightarrow}\}} (\mathcal{T}', x')$, debido al Teorema 3.37. Por lo tanto, $(\mathcal{T}, x) \equiv^{\{\downarrow, \overset{*}{\rightarrow}\}} (\mathcal{T}', x')$. \square

Como aplicación, podemos utilizar el Teorema 3.41 para probar que no existe una $\{\downarrow, \overset{*}{\rightarrow}\}$ -bisimulación entre dos modelos dados.

Ejemplo 3.42. Si consideramos los árboles \mathcal{T} y \mathcal{T}' de la Figura 7 entonces los modelos (\mathcal{T}, x) y (\mathcal{T}', x') no son $\{\downarrow, \overset{*}{\rightarrow}\}$ -bisimilares ya que si

$$\varphi := \langle \downarrow [\langle \downarrow \neq \overset{*}{\rightarrow} [\langle \downarrow \neq \overset{*}{\rightarrow} \downarrow \rangle] \downarrow] \rangle$$

entonces $(\mathcal{T}, x) \models \varphi$ pero $(\mathcal{T}', x') \not\models \varphi$. Por lo tanto, debido al Teorema 3.41 necesariamente $(\mathcal{T}, x) \not\equiv^{\{\downarrow, \overset{*}{\rightarrow}\}} (\mathcal{T}', x')$.

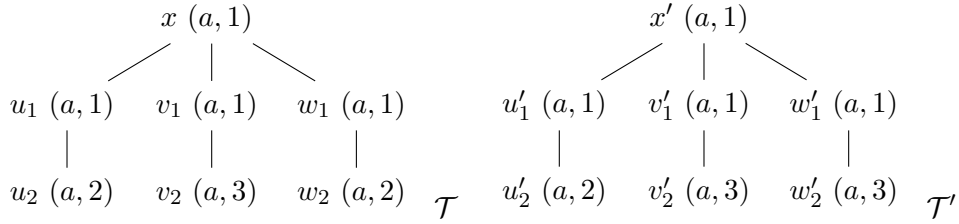


Figura 7: $(\mathcal{T}, x) \not\equiv^{\{\downarrow, \overset{*}{\rightarrow}\}} (\mathcal{T}', x')$

4. XPath₌(↓, $\overset{*}{\rightarrow}$, $\overset{*}{\leftarrow}$)

4.1. Sintaxis y Semántica

Definición 4.1. La sintaxis de XPath₌(↓, $\overset{*}{\rightarrow}$, $\overset{*}{\leftarrow}$) se define de igual manera que en la Definición 2.2 para XPath₌(↓) salvo que ahora $O = \{\varepsilon, \downarrow, \overset{*}{\rightarrow}, \overset{*}{\leftarrow}\}$.

Definición 4.2. La semántica de $\overset{*}{\leftarrow}$ sobre un árbol ordenado con datos \mathcal{T} es $\llbracket \overset{*}{\leftarrow} \rrbracket^{\mathcal{T}} := \{(x, y) \in T^2 \mid y \overset{*}{\rightarrow} x\}$.

Para el resto de los símbolos es igual a la Definición 2.5 y Definición 3.6.

4.2. Expresiones Normales

Observación 4.3. La Definición 3.7 se puede extender sin problemas para las expresiones del fragmento XPath₌(↓, $\overset{*}{\rightarrow}$, $\overset{*}{\leftarrow}$). La Observación 3.8 y la Observación 3.13 siguen siendo válidas en XPath₌(↓, $\overset{*}{\rightarrow}$, $\overset{*}{\leftarrow}$).

Definición 4.4. Una expresión de camino α en XPath₌(↓, $\overset{*}{\rightarrow}$, $\overset{*}{\leftarrow}$) es *horizontal* si $\alpha = [\psi_0]o_1[\psi_1] \dots o_n[\psi_n]$ con $o_i \in \{\overset{*}{\rightarrow}, \overset{*}{\leftarrow}\}$ para todo $i \in \{1, \dots, n\}$. Y si todos los o_i son iguales, diremos que α es *horizontal directa*.

Una expresión de camino β es *hacia abajo* si $\beta = \lambda$ o $\beta = \downarrow[\psi_1] \dots \downarrow[\psi_n]$.

Por último, diremos que una expresión γ es *directa* si $\gamma = \alpha\beta$ con α horizontal directa y β hacia abajo.

Definición 4.5. Extendemos la Definición 3.10 a XPath₌(↓, $\overset{*}{\rightarrow}$, $\overset{*}{\leftarrow}$) definiendo $\text{sub}(\overset{*}{\leftarrow}) := \emptyset$.

Diremos que una expresión de camino α es *normal* si para todo $\gamma \in \text{sub}(\alpha) \cup \{\alpha\}$ se tiene que γ es directa. Por otra parte, una expresión de nodo φ es *normal* si todo $\gamma \in \text{sub}(\varphi)$ es directa.

Para abreviar, si α es horizontal directa y normal, diremos que es *horizontal normal*.

Lema 4.6. Toda expresión horizontal α cuyos tests de nodo son expresiones normales es equivalente a una unión de expresiones horizontales normales.

Demostración. Supongamos $\alpha = [\psi_0]o_1[\psi_1] \dots o_n[\psi_n]$ con $o_i \in \{\overset{*}{\rightarrow}, \overset{*}{\leftarrow}\}$ para todo $i \in \{1, \dots, n\}$ y ψ_j es normal para todo $j \in \{0, \dots, n\}$. Llamemos al n la *longitud* de α .

Probaremos algo un poco más fuerte, y es que $\alpha \equiv \bigcup_{k=1}^N \alpha_k$ donde α_k es horizontal normal y con longitud menor o igual a la de α para todo $k \in \{1, \dots, N\}$. Esto lo haremos por inducción sobre n . Los casos $n = 0$ y $n = 1$ son obvios.

Si $n = 2$ entonces hay dos casos: $o_1 = o_2$ y $o_1 \neq o_2$. Para el primero tenemos que α ya es horizontal normal. Veamos el segundo caso, y supongamos que $\alpha = [\psi_0] \overset{*}{\rightarrow} [\psi_1] \overset{*}{\leftarrow} [\psi_2]$. Si $(x, y) \in \llbracket \alpha \rrbracket^{\mathcal{T}}$ entonces existe $z \in T$ tal que $x \overset{*}{\rightarrow} z$, $y \overset{*}{\leftarrow} z$ y $(\mathcal{T}, z) \models \psi_1$. Pueden pasar dos situaciones:

Si $x \xrightarrow{*} y \xrightarrow{*} z$ entonces $(x, y) \in \llbracket \alpha_1 \rrbracket^{\mathcal{T}}$ con $\alpha_1 := [\psi_0] \xrightarrow{*} \langle [\psi_2] \xrightarrow{*} [\psi_1] \rangle$.

Si $y \xrightarrow{*} x \xrightarrow{*} z$ entonces $(x, y) \in \llbracket \alpha_2 \rrbracket^{\mathcal{T}}$ con $\alpha_2 := \langle [\psi_0] \xrightarrow{*} [\psi_1] \rangle \xleftarrow{*} [\psi_2]$.

Como $\llbracket \alpha_1 \rrbracket^{\mathcal{T}} \cup \llbracket \alpha_2 \rrbracket^{\mathcal{T}} \subseteq \llbracket \alpha \rrbracket^{\mathcal{T}}$ tenemos que $\alpha \equiv \alpha_1 \cup \alpha_2$. Notar que α_1 y α_2 son horizontales normales con longitud igual a 1. El caso $\alpha = [\psi_0] \xleftarrow{*} [\psi_1] \xrightarrow{*} [\psi_2]$ es análogo.

Supongamos $n > 2$. Por hipótesis inductiva $\beta = [\psi_1] o_2 [\psi_2] \dots o_n [\psi_n]$ es equivalente a $\bigcup_{k=1}^M \beta_k$ con β_k horizontal normal y con longitud menor o igual a β para todo $k \in \{1, \dots, M\}$. Por lo tanto, $\alpha \equiv \bigcup_{k=1}^M [\psi_0] o_1 \beta_k$. Supongamos que $o_1 = \xrightarrow{*}$ (si $o_1 = \xleftarrow{*}$ es análogo). Entonces, si todos los β_k son de la forma $[-] \xrightarrow{*} [-] \xrightarrow{*} \dots \xrightarrow{*} [-]$, listo. Si no, existe $k_0 \in \{1, \dots, M\}$ tal que $\beta_{k_0} = [\varphi_1] \xleftarrow{*} [\varphi_2] \xleftarrow{*} \dots \xleftarrow{*} [-]$. Por el caso $n = 2$, como ψ_0, φ_1 y φ_2 son normales, tenemos que $[\psi_0] \xrightarrow{*} [\varphi_1] \xleftarrow{*} [\varphi_2]$ es equivalente a $[\eta_0] \xrightarrow{*} [\eta_1] \cup [\eta_2] \xleftarrow{*} [\eta_3]$ con η_i normal para $i \in \{1, 2, 3, 4\}$. Luego,

$$\begin{aligned} [\psi_0] \xrightarrow{*} \beta_{k_0} &\equiv \gamma_1 \cup \gamma_2 \\ \gamma_1 &:= [\eta_0] \xrightarrow{*} [\eta_1] \xleftarrow{*} \dots \xleftarrow{*} [-] \\ \gamma_2 &:= [\eta_2] \xleftarrow{*} [\eta_3] \xleftarrow{*} \dots \xleftarrow{*} [-] \end{aligned}$$

Tenemos que γ_2 es horizontal normal mientras que γ_1 no. Pero la longitud de γ_1 es estrictamente menor que la de α . Aplicando hipótesis inductiva, γ_1 es equivalente a una unión de expresiones normales con longitudes a lo sumo igual a la de γ_1 .

□

Lema 4.7. Para toda α expresión horizontal con tests de nodo normales, existe ψ normal tal que $\downarrow \alpha \equiv \downarrow [\psi]$.

Demostración. Supongamos $\alpha = [\psi_0] o_1 [\psi_1] \dots o_n [\psi_n]$ con $o_i \in \{\xrightarrow{*}, \xleftarrow{*}\}$ para todo $i \in \{1, \dots, n\}$ y ψ_i es normal para todo $i \in \{0, \dots, n\}$. Haremos inducción sobre la longitud n . El caso base $n = 0$ es trivial.

Si $n = 1$ entonces $\downarrow \alpha = \downarrow [\psi_0] o_1 [\psi_1]$. Supongamos $o_1 = \xrightarrow{*}$ (el caso $o_1 = \xleftarrow{*}$ es análogo), entonces $\downarrow [\psi_0] \xrightarrow{*} [\psi_1] \equiv \downarrow \langle [\psi_1] \xleftarrow{*} [\psi_0] \rangle$. Como $\psi := \langle [\psi_1] \xleftarrow{*} [\psi_0] \rangle$ es normal, llegamos a lo que queríamos.

Supongamos $n > 1$. Si $\beta = [\psi_0] o_1 [\psi_1] \dots o_{n-1} [\psi_{n-1}]$ entonces β es normal con longitud $n - 1$. Por lo tanto, existe $[\varphi]$ normal tal que $\downarrow \beta \equiv \downarrow [\varphi]$. Luego, $\downarrow \alpha \equiv \downarrow [\varphi] o_n [\psi_n]$. Por el caso $n = 1$, existe ψ normal tal que $\downarrow [\varphi] o_n [\psi_n] \equiv \downarrow [\psi]$.

□

Observación 4.8. Observar que la demostración del Lema 4.7 no podría reproducirse en el fragmento XPath= $(\downarrow, \overset{*}{\rightarrow})$.

Corolario 4.9. Si $\alpha \neq \lambda$ es una expresión de camino hacia abajo cuyos tests de nodo son normales y β es una expresión de camino horizontal normal, entonces $\alpha\beta$ es equivalente a una expresión hacia abajo cuyos tests de nodo son normales.

Demostración. Supongamos $\alpha = \downarrow [\varphi_1] \dots \downarrow [\varphi_m]$ con φ_i normal para todo $i \in \{1, \dots, m\}$ y $\beta = [\psi_0]o[\psi_1] \dots o[\psi_n]$ con $o \in \{\overset{*}{\rightarrow}, \overset{*}{\leftarrow}\}$ y ψ_j normal para todo $j \in \{0, \dots, n\}$.

Por 3 de la Observación 3.13, $[\varphi_m][\psi_0] = [\varphi_m \wedge \psi_0]$. Notar que $\varphi_m \wedge \psi_0$ es normal. Luego, por el Lema 4.7 podemos concluir que $\downarrow [\varphi_m]\beta \equiv \downarrow [\varphi]$ para cierta φ normal. Por lo tanto, $\alpha\beta \equiv \downarrow [\varphi_0] \dots \downarrow [\varphi_{m-1}] \downarrow [\varphi]$. \square

Proposición 4.10. Para toda expresión de nodo φ y expresión de camino α existe una expresión de nodo φ' normal y una expresión de camino α' , la cual es una unión de expresiones de camino normales, tales que $\varphi \equiv \varphi'$ y $\alpha \equiv \alpha'$

Demostración. Haciendo inducción mutua sobre φ y α , la prueba es similar a la dada para la Proposición 3.17, salvo para el caso de la concatenación.

Supongamos entonces que $\alpha = \beta\gamma$ con $\beta \neq \lambda \neq \gamma$. Veamos primero cuando β y γ son normales para luego pasar al caso general.

- Supongamos $\beta = \beta^{\leftrightarrow}\beta^\downarrow$ y $\gamma = \gamma^{\leftrightarrow}\gamma^\downarrow$ donde β^{\leftrightarrow} y γ^{\leftrightarrow} son expresiones horizontales normales, y β^\downarrow y γ^\downarrow son expresiones hacia abajo cuyos tests de nodo son normales.

Si $\beta^\downarrow = \lambda$ entonces $\beta\gamma = \beta^{\leftrightarrow}\gamma^{\leftrightarrow}\gamma^\downarrow$. Como $\beta^{\leftrightarrow}\gamma^{\leftrightarrow}$ es horizontal con tests de nodo normales, por el Lema 4.6, $\beta^{\leftrightarrow}\gamma^{\leftrightarrow} \equiv \bigcup_{k=1}^n \eta_k^{\leftrightarrow}$ con η_k^{\leftrightarrow} horizontal normal para todo $k \in \{1, \dots, n\}$. Luego, $\alpha \equiv \bigcup_{k=1}^n \eta_k^{\leftrightarrow}\gamma^\downarrow$ y cada $\eta_k\gamma^\downarrow$ es normal.

Si $\beta^\downarrow \neq \lambda$ entonces por el Corolario 4.9, $\beta^\downarrow\gamma^{\leftrightarrow} \equiv \eta^\downarrow$ con η^\downarrow hacia abajo y sus tests de nodo normales. Por lo tanto, $\alpha \equiv \beta^{\leftrightarrow}(\beta^\downarrow\gamma^{\leftrightarrow})\gamma^\downarrow \equiv \beta^{\leftrightarrow}(\eta^\downarrow\gamma^\downarrow)$. Como $\eta^\downarrow\gamma^\downarrow$ es hacia abajo con tests de nodo normales, llegamos a lo que queríamos.

- Por hipótesis inductiva, $\beta \equiv \bigcup_{i=1}^M \beta_i$ con β_i normal para todo $i \in \{1, \dots, M\}$ y $\gamma = \bigcup_{j=1}^N \gamma_j$ con γ_j normal para todo $j \in \{1, \dots, N\}$. Luego, $\alpha \equiv \bigcup_{i=1}^M \bigcup_{j=1}^N \beta_i\gamma_j$. Cada $\beta_i\gamma_j$, por el item anterior, es equivalente a una unión de normales. Por lo tanto, α resulta equivalente a una unión de expresiones de camino normales.

Ejemplo 4.11. Veamos con un ejemplo de expresión de nodo cómo hallar otra que sea normal.

Sea $\varphi := \langle \varepsilon \cup (\downarrow [a] \leftarrow^* [b] \rightarrow^* [c]) \neq \varepsilon \rangle$. Queremos encontrar una φ' normal tal que $\varphi \equiv \varphi'$.

Primero, podemos deshacernos de la unión con la propiedad 4 de la Observación 3.13. Luego, $\varphi \equiv \langle \varepsilon \neq \varepsilon \rangle \vee \langle \downarrow [a] \leftarrow^* [b] \rightarrow^* [c] \neq \varepsilon \rangle$. Como $\langle \varepsilon \neq \varepsilon \rangle$ es una contradicción, podemos simplificar la disyunción y llegamos a que $\varphi \equiv \langle \downarrow [a] \leftarrow^* [b] \rightarrow^* [c] \neq \varepsilon \rangle$.

Si llamamos $\alpha = \downarrow [a] \leftarrow^* [b] \rightarrow^* [c]$ siguiendo con la idea del Lema 4.7 tenemos que $\alpha \equiv \downarrow [[b] \rightarrow^* [a]] \rightarrow^* [c] \equiv \downarrow [[c] \leftarrow^* [[b] \rightarrow^* [a]]]$. Con lo cual, $\varphi \equiv \langle \downarrow [[c] \leftarrow^* [[b] \rightarrow^* [a]]] \neq \varepsilon \rangle$.

Por último, agregamos los tests de nodo $[\top]$ para completar la forma normal, $\varphi' := \langle [\top] \downarrow [[c] \leftarrow^* [[b] \rightarrow^* [a]]] \neq [\top] \rangle \equiv \varphi$.

□

4.3. Bisimilaridad

El objetivo principal de estudiar previamente la existencia de una expresión normal equivalente para toda fórmula es que nos da un idea clara y sencilla de cómo definir bisimulación de una manera adecuada. Similar a lo que ocurría con el fragmento anterior, primero vamos a definir las $\{\downarrow, \rightarrow^*, \leftarrow^*\}$ -trayectorias para luego dar el concepto de $\{\downarrow, \rightarrow^*, \leftarrow^*\}$ -bisimulación.

Definición 4.12. Dado un árbol ordenado \mathcal{T} , diremos que una secuencia $\tilde{v} = v_0 e v_1 e \dots v_n \downarrow v_{n+1} \downarrow \dots \downarrow v_{n+k}$ con $v_i \in T$ para todo $i \in \{0, \dots, n+k\}$ y $e \in \{\rightarrow^*, \leftarrow^*\}$ es una $\{\downarrow, \rightarrow^*, \leftarrow^*\}$ -trayectoria si

$$\begin{aligned} (\mathcal{T}, v_i, v_{i+1}) \models e & \quad \forall i \in \{0, \dots, n-1\} \\ (\mathcal{T}, v_i, v_{i+1}) \models \downarrow & \quad \forall i \in \{n, \dots, k-1\} \end{aligned}$$

Al $n \in \mathbb{N}_0$ lo llamaremos la *longitud horizontal* y al $k \in \mathbb{N}_0$, la *longitud hacia abajo*.

Para cada $o \in \{\rightarrow^*, \leftarrow^*\}$ denotaremos por $\{\downarrow, \rightarrow^*, \leftarrow^*\}$ - $\text{Tray}_{n,k}^o(\mathcal{T}, x)$ al conjunto de $\{\downarrow, \rightarrow^*, \leftarrow^*\}$ -trayectorias \tilde{v} con $e = o$, longitud horizontal n y longitud hacia abajo k tales que $v_0 = x$.

Definición 4.13. Dados \mathcal{T} y \mathcal{T}' árboles ordenados con datos, diremos que una relación $Z \subseteq T \times T'$ es una $\{\downarrow, \rightarrow^*, \leftarrow^*\}$ -bisimulación si para todo $(x, x') \in Z$ se cumplen las siguientes propiedades:

Harmony. $label(x) = label(x')$

Zig. Para todo $o, e \in \{\rightarrow^*, \leftarrow^*\}$, $\tilde{v} \in \{\downarrow, \rightarrow^*, \leftarrow^*\}$ - $\text{Tray}_{n,k}^o(\mathcal{T}, x)$ y $\tilde{w} \in \{\downarrow, \rightarrow^*, \leftarrow^*\}$ - $\text{Tray}_{m,t}^e(\mathcal{T}, x)$ existen $\tilde{v}' \in \{\downarrow, \rightarrow^*, \leftarrow^*\}$ - $\text{Tray}_{n,k}^o(\mathcal{T}', x')$ y $\tilde{w}' \in \{\downarrow, \rightarrow^*, \leftarrow^*\}$ - $\text{Tray}_{m,t}^e(\mathcal{T}', x')$ tales que

- $(v_i, v'_i) \in Z \forall i \in \{1, \dots, n+k\}$

- $(w_i, w'_i) \in Z \forall i \in \{1, \dots, m+t\}$
- $data(v_{n+k}) = data(w_{m+t}) \Leftrightarrow data(v'_{n+k}) = data(w'_{m+t})$

Zag. Para todo $o, e \in \{\rightarrow, \leftarrow\}$, $\tilde{v}' \in \{\downarrow, \rightarrow, \leftarrow\}$ -Tray $^o_{n,k}(\mathcal{T}', x')$ y $\tilde{w}' \in \{\downarrow, \rightarrow, \leftarrow\}$ -Tray $^e_{m,t}(\mathcal{T}', x')$ existen $\tilde{v} \in \{\downarrow, \rightarrow, \leftarrow\}$ -Tray $^o_{n,k}(\mathcal{T}, x)$ y $\tilde{w} \in \{\downarrow, \rightarrow, \leftarrow\}$ -Tray $^e_{m,t}(\mathcal{T}, x)$ tales que

- $(v_i, v'_i) \in Z \forall i \in \{1, \dots, n+k\}$
- $(w_i, w'_i) \in Z \forall i \in \{1, \dots, m+t\}$
- $data(v_{n+k}) = data(w_{m+t}) \Leftrightarrow data(v'_{n+k}) = data(w'_{m+t})$

Definición 4.14. Dados dos modelos punteados (\mathcal{T}, x) y (\mathcal{T}', x') diremos que son $\{\downarrow, \rightarrow, \leftarrow\}$ -bisimilares y lo notaremos $(\mathcal{T}, x) \simeq^{\{\downarrow, \rightarrow, \leftarrow\}} (\mathcal{T}', x')$ si existe una $\{\downarrow, \rightarrow, \leftarrow\}$ -bisimulación $Z \subseteq T \times T'$ tal que $(x, x') \in Z$.

Ejemplo 4.15. Los modelos punteados de la Figura 5 son $\{\downarrow, \rightarrow, \leftarrow\}$ -bisimilares utilizando la misma relación Z definida en el Ejemplo 3.24 para $\{\downarrow, \rightarrow\}$ -bisimulación. Sin embargo, no es cierto que ambas nociones de bisimulación coincidan siempre desde la raíz de un árbol. Por ejemplo, si \mathcal{T} y \mathcal{T}' son los árboles de la Figura 8 entonces para los modelos (\mathcal{T}, x) y (\mathcal{T}', x') existe una única posible $\{\downarrow, \rightarrow\}$ -bisimulación $Z^{\{\downarrow, \rightarrow\}} := \{(x, x'), (y, y'), (z, z'), (w, w')\}$. Por otro lado, no es posible definir una $\{\downarrow, \rightarrow, \leftarrow\}$ -bisimulación $Z^{\{\downarrow, \rightarrow, \leftarrow\}}$ tal que $(x, x') \in Z^{\{\downarrow, \rightarrow, \leftarrow\}}$ pues necesariamente tendríamos que $Z^{\{\downarrow, \rightarrow, \leftarrow\}} = Z^{\{\downarrow, \rightarrow\}}$ pero si consideramos las $\{\downarrow, \rightarrow, \leftarrow\}$ -trayectorias $\tilde{v} = z \rightarrow w$ y $\tilde{w} = z \leftarrow y$ con $data(y) = 1 = data(w)$ entonces como $(z, z') \in Z^{\{\downarrow, \rightarrow, \leftarrow\}}$ por **Zig** las únicas posibilidades son $\tilde{v}' = z' \rightarrow w'$ y $\tilde{w}' = z' \leftarrow y'$ pero $data(y') = 1 \neq 3 = data(w')$.

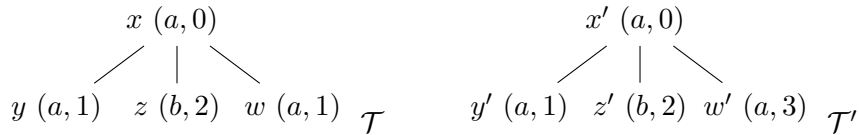


Figura 8: $(\mathcal{T}, x) \simeq^{\{\downarrow, \rightarrow\}} (\mathcal{T}', x)$ pero $(\mathcal{T}, x) \not\simeq^{\{\downarrow, \rightarrow, \leftarrow\}} (\mathcal{T}', x)$

4.4. Bisimilaridad Acotada

Definición 4.16. Extendemos la Definición 3.14 a $\text{XPath}_=(\downarrow, \rightarrow, \leftarrow)$ definiendo $\text{ad}(o\alpha) = \text{ad}(\alpha) + 1$ para todo $o \in \{\downarrow, \rightarrow, \leftarrow\}$.

Definición 4.17. Denotaremos por $\ell\text{-XPath}_=(\downarrow, \rightarrow, \leftarrow)$ al conjunto de expresiones normales η tales que $\text{ad}(\eta) \leq \ell$.

Definición 4.18. Diremos que el par de modelos (\mathcal{T}, x) y (\mathcal{T}', x') es $\{\downarrow, \overset{*}{\rightarrow}, \overset{*}{\leftarrow}\}_\ell$ -equivalente y escribiremos $(\mathcal{T}, x) \equiv_\ell^{\{\downarrow, \overset{*}{\rightarrow}, \overset{*}{\leftarrow}\}} (\mathcal{T}', x')$ siempre que para toda $\varphi \in \ell\text{-XPath}_=(\downarrow, \overset{*}{\rightarrow}, \overset{*}{\leftarrow})$ se cumpla que $(\mathcal{T}, x) \models \varphi$, si y sólo si, $(\mathcal{T}', x') \models \varphi$.

Definición 4.19. Dados árboles ordenados \mathcal{T} y \mathcal{T}' , una familia $\mathcal{Z} = \{Z_j\}_{0 \leq j \leq \ell}$ con $Z_j \subseteq T \times T'$ para todo j es una $\{\downarrow, \overset{*}{\rightarrow}, \overset{*}{\leftarrow}\}_\ell$ -bisimulación si para todo j y $(x, x') \in Z_j$ se cumplen:

Harmony. $label(x) = label(x')$

Zig. Para todo $e, o \in \{\overset{*}{\rightarrow}, \overset{*}{\leftarrow}\}$, $\tilde{v} \in \{\downarrow, \overset{*}{\rightarrow}, \overset{*}{\leftarrow}\}\text{-Tray}_{n,k}^e(\mathcal{T}, x)$ y $\tilde{w} \in \{\downarrow, \overset{*}{\rightarrow}, \overset{*}{\leftarrow}\}\text{-Tray}_{m,t}^o(\mathcal{T}, x)$ con $\max\{n+k, m+t\} \leq j$ existen $\tilde{v}' \in \{\downarrow, \overset{*}{\rightarrow}, \overset{*}{\leftarrow}\}\text{-Tray}_{n,k}^e(\mathcal{T}', x')$ y $\tilde{w}' \in \{\downarrow, \overset{*}{\rightarrow}, \overset{*}{\leftarrow}\}\text{-Tray}_{m,\ell}^o(\mathcal{T}', x')$ tales que

- $(v_i, v'_i) \in Z_{j-i} \ \forall i \in \{1, \dots, n+k\}$
- $(w_i, w'_i) \in Z_{j-i} \ \forall i \in \{1, \dots, m+t\}$
- $data(v_{n+k}) = data(w_{m+t}) \Leftrightarrow data(v'_{n+k}) = data(w'_{m+t})$

Zag. Para todo $e, o \in \{\overset{*}{\rightarrow}, \overset{*}{\leftarrow}\}$, $\tilde{v}' \in \{\downarrow, \overset{*}{\rightarrow}, \overset{*}{\leftarrow}\}\text{-Tray}_{n,k}^e(\mathcal{T}', x')$ y $\tilde{w}' \in \{\downarrow, \overset{*}{\rightarrow}, \overset{*}{\leftarrow}\}\text{-Tray}_{m,j}^o(\mathcal{T}', x')$ con $\max\{n+k, m+t\} \leq j$ existen $\tilde{v} \in \{\downarrow, \overset{*}{\rightarrow}, \overset{*}{\leftarrow}\}\text{-Tray}_{n,k}^e(\mathcal{T}, x)$ y $\tilde{w} \in \{\downarrow, \overset{*}{\rightarrow}, \overset{*}{\leftarrow}\}\text{-Tray}_{m,\ell}^o(\mathcal{T}, x)$ tales que

- $(v_i, v'_i) \in Z_{j-i} \ \forall i \in \{1, \dots, n+k\}$
- $(w_i, w'_i) \in Z_{j-i} \ \forall i \in \{1, \dots, m+t\}$
- $data(v_{n+k}) = data(w_{m+t}) \Leftrightarrow data(v'_{n+k}) = data(w'_{m+t})$

Definición 4.20. Dos modelos punteados (\mathcal{T}, x) y (\mathcal{T}', x') son $\{\downarrow, \overset{*}{\rightarrow}, \overset{*}{\leftarrow}\}_\ell$ -bisimilares si existe una $\{\downarrow, \overset{*}{\rightarrow}, \overset{*}{\leftarrow}\}_\ell$ -bisimulación $\mathcal{Z} = \{Z_j\}_{0 \leq j \leq \ell}$ tal que $(x, x') \in Z_\ell$. En tal caso, escribiremos $(\mathcal{T}, x) \simeq_\ell^{\{\downarrow, \overset{*}{\rightarrow}, \overset{*}{\leftarrow}\}} (\mathcal{T}', x')$.

Definición 4.21. Sean $\tilde{v} \in \{\downarrow, \overset{*}{\rightarrow}, \overset{*}{\leftarrow}\}\text{-Tray}_{n,k}(\mathcal{T}, x)$ y α expresión de camino normal tales que

$$\begin{aligned} \alpha &= [\psi_0]o_1[\psi_1] \dots o_n[\psi_n] \downarrow [\psi_{n+1}] \downarrow \dots \downarrow [\psi_{n+k}] & o_i \in \{\downarrow, \overset{*}{\rightarrow}, \overset{*}{\leftarrow}\} \\ \tilde{v} &= v_0e_1v_1 \dots e_nv_n \downarrow v_{n+1} \downarrow \dots \downarrow v_{n+k} \end{aligned}$$

entonces diremos que el par (\mathcal{T}, \tilde{v}) *satisface* α y escribiremos $(\mathcal{T}, \tilde{v}) \models \alpha$ si se cumple que

$$\begin{aligned} o_i &= e_i & \forall i \in \{1, \dots, n\} \\ (\mathcal{T}, v_i) &\models \psi_i & \forall i \in \{0, \dots, n+k\} \end{aligned}$$

Lema 4.22. Sea $\mathcal{Z} = \{Z_j : 0 \leq j \leq \ell\}$ una $\{\downarrow, \overset{*}{\rightarrow}, \overset{*}{\leftarrow}\}_\ell$ -bisimulación entre (\mathcal{T}, x) y (\mathcal{T}', x') . Para todo $0 \leq j \leq \ell$, α normal y φ normal tales que $\max\{\text{ad}(\alpha), \text{ad}(\varphi)\} \leq j$ se tiene que

1. Si $(x, x') \in Z_j$ entonces $(\mathcal{T}, x) \models \varphi \iff (\mathcal{T}', x') \models \varphi$.
2. Si $\tilde{v} \in \{\downarrow, \overset{*}{\rightarrow}, \overset{*}{\leftarrow}\}\text{-Tray}_{n,k}^o(\mathcal{T}, x)$, $\tilde{v}' \in \{\downarrow, \overset{*}{\rightarrow}, \overset{*}{\leftarrow}\}\text{-Tray}_{n,k}^o(\mathcal{T}', x')$ y $(v_i, v'_i) \in Z_{j-i}$ para todo $i \in \{0, \dots, n+k\}$ con $n+k \leq j$ entonces $(\mathcal{T}, \tilde{v}) \models \alpha \iff (\mathcal{T}', \tilde{v}') \models \alpha$.

Demostración. Lo probaremos por inducción sobre j .

- Si $j = 0$ entonces $\text{ad}(\varphi) = \text{ad}(\alpha) = 0$. Por lo tanto $\alpha = [\psi]$ con $\text{ad}(\psi) = 0$. Aplicando un resultado similar al Lema 3.30 sobre φ y ψ tenemos probado este caso.
- Supongamos $j > 0$. Haremos inducción estructural mutua sobre φ y α .
 2. Supongamos $\alpha = [\psi]$ entonces $\text{ad}(\psi) = \text{ad}(\alpha)$. Si $(\mathcal{T}, \tilde{v}) \models \alpha$ necesariamente $\tilde{v} = x$. Luego, $\tilde{v}' = x'$ y $(x, x') \in Z_j$. Por hipótesis inductiva sobre ψ , aplicando 1, como $(\mathcal{T}, x) \models \psi$ entonces $(\mathcal{T}', x') \models \psi$. Por lo tanto, $(\mathcal{T}', \tilde{v}') \models \alpha$. De manera análoga, si $(\mathcal{T}', \tilde{v}') \models \alpha$ entonces $(\mathcal{T}, \tilde{v}) \models \alpha$.

Supongamos ahora que $\alpha = [\psi]o\beta$ donde $o \in \{\downarrow, \overset{*}{\rightarrow}, \overset{*}{\leftarrow}\}$ y β es una expresión normal de camino. Para simplificar, fijemos $o = \overset{*}{\rightarrow}$ (los otros casos son iguales). Por un lado, si $\tilde{v} = x \overset{*}{\rightarrow} \tilde{w}$ con $\tilde{w} \in \{\downarrow, \overset{*}{\rightarrow}, \overset{*}{\leftarrow}\}\text{-Tray}_{n-1,k}^{\overset{*}{\rightarrow}}(\mathcal{T}, y)$ y $\tilde{v}' = x' \overset{*}{\rightarrow} \tilde{w}'$ con

$\tilde{w}' \in \{\downarrow, \overset{*}{\rightarrow}, \overset{*}{\leftarrow}\}\text{-Tray}_{n-1,k}^{\overset{*}{\rightarrow}}(\mathcal{T}', y')$, como $(v_i, v'_i) \in Z_{j-i}$ para todo $i \in \{0, \dots, n+k\}$ tenemos que $(x, x') \in Z_j$ y $(w_i, w'_i) \in Z_{j-1-i}$ para todo $i \in \{0, \dots, n+k-1\}$. Por el otro, aplicando hipótesis inductiva sobre $j-1$ (para β) y sobre ψ (en el mismo j)

$$(\mathcal{T}, x) \models \psi \iff (\mathcal{T}', x') \models \psi \quad (\mathcal{T}, \tilde{w}) \models \beta \iff (\mathcal{T}', \tilde{w}') \models \beta$$

Por lo tanto, $(\mathcal{T}, \tilde{v}) \models \alpha$, si y sólo si, $(\mathcal{T}', \tilde{v}') \models \alpha$.

1. Los casos en que φ es una etiqueta o combinación booleana son fáciles. Supongamos entonces $\varphi = \langle \beta \odot \gamma \rangle$ con $\odot \in \{=, \neq\}$. Notar que las profundidades de β y γ son menores o iguales a las de φ . Si $(\mathcal{T}, x) \models \varphi$ entonces existen $\tilde{v} \in \{\downarrow, \overset{*}{\rightarrow}, \overset{*}{\leftarrow}\}\text{-Tray}_{n,k}^o(\mathcal{T}, x)$ y $\tilde{w} \in \{\downarrow, \overset{*}{\rightarrow}, \overset{*}{\leftarrow}\}\text{-Tray}_{m,t}^e(\mathcal{T}, x)$ tales que $(\mathcal{T}, \tilde{v}) \models \beta$, $(\mathcal{T}, \tilde{w}) \models \gamma$ y $\text{data}(v_{n+k}) \odot \text{data}(w_{m+t})$. Como $(x, x') \in Z_j$ por **Zig** existen $\tilde{v}' \in \{\downarrow, \overset{*}{\rightarrow}, \overset{*}{\leftarrow}\}\text{-Tray}_{n,k}^o(\mathcal{T}', x')$ y $\tilde{w}' \in \{\downarrow, \overset{*}{\rightarrow}, \overset{*}{\leftarrow}\}\text{-Tray}_{m,t}^e(\mathcal{T}', x')$ tales que $(v_i, v'_i) \in Z_{j-i}$ para todo $i \in \{0, \dots, n+k\}$, $(w_i, w'_i) \in Z_{j-i}$ para todo $i \in \{0, \dots, m+t\}$ y $\text{data}(v'_{n+k}) \odot \text{data}(w'_{m+t})$.

Por hipótesis inductiva sobre β y γ , aplicando [2](#) tenemos que $(\mathcal{T}', v') \models \beta$ y $(\mathcal{T}', w') \models \gamma$. Por lo tanto, $(\mathcal{T}', x') \models \langle \beta \odot \gamma \rangle$. De manera análoga, aplicando **Zag**, si $(\mathcal{T}', x') \models \varphi$ entonces $(\mathcal{T}, x) \models \varphi$.

□

Proposición 4.23. Sean \mathcal{T} y \mathcal{T}' árboles ordenados. Luego,

$$(\mathcal{T}, x) \simeq_{\ell}^{\{\downarrow, \overset{*}{\rightarrow}, \overset{*}{\leftarrow}\}} (\mathcal{T}', x') \implies (\mathcal{T}, x) \equiv_{\ell}^{\{\downarrow, \overset{*}{\rightarrow}, \overset{*}{\leftarrow}\}} (\mathcal{T}', x')$$

Demostración. Sea $\mathcal{Z} = \{Z_j\}_{0 \leq j \leq \ell}$ una $\{\downarrow, \overset{*}{\rightarrow}, \overset{*}{\leftarrow}\}_{\ell}$ -bisimulación entre (\mathcal{T}, x) y (\mathcal{T}', x') . Queremos probar que para toda φ normal tal que $\text{ad}(\varphi) \leq \ell$

$$(\mathcal{T}, x) \models \varphi \iff (\mathcal{T}', x') \models \varphi$$

Supongamos $(\mathcal{T}, x) \models \varphi$. Como $(x, x') \in Z_{\ell}$, por [2](#) del Lema [4.22](#), tenemos que $(\mathcal{T}', x') \models \varphi'$. Análogamente, si $(\mathcal{T}', x') \models \varphi$ entonces $(\mathcal{T}, x) \models \varphi$. □

Lema 4.24. Existe una cantidad finita de expresiones de nodo en ℓ -XPath $_{=}$ ($\downarrow, \overset{*}{\rightarrow}, \overset{*}{\leftarrow}$) salvo equivalencia lógica.

Demostración. La prueba es similar a la del Lema [3.34](#). □

Definición 4.25. Dado $\tilde{v} \in \{\downarrow, \overset{*}{\rightarrow}, \overset{*}{\leftarrow}\}$ -Tray $_{n,k}(\mathcal{T}, x)$ tal que $\tilde{v} = v_0 e_1 v_1 \dots e_n v_n \downarrow v_{n+1} \downarrow \dots \downarrow v_{n+k}$ definimos $\alpha_{\tilde{v}}$ la expresión (normal) de camino asociada a \tilde{v} como

$$\alpha_{\tilde{v}} := [\top]e_1[\top] \dots [\top]e_n[\top]o_1[\top]o_2 \dots o_k[\top]$$

donde $o_j = \downarrow$ para todo $j \in \{1, \dots, k\}$. Notar que $(\mathcal{T}, \tilde{v}) \models \alpha_{\tilde{v}}$.

Proposición 4.26. Sean \mathcal{T} y \mathcal{T}' árboles ordenados. Luego,

$$(\mathcal{T}, x) \equiv_{\ell}^{\{\downarrow, \overset{*}{\rightarrow}, \overset{*}{\leftarrow}\}} (\mathcal{T}', x') \implies (\mathcal{T}, x) \simeq_{\ell}^{\{\downarrow, \overset{*}{\rightarrow}, \overset{*}{\leftarrow}\}} (\mathcal{T}', x')$$

Demostración. Para todo $i \in \{0, \dots, \ell\}$ definimos las relaciones $Z_i \subseteq T \times T'$ de la siguiente manera

$$(y, y') \in Z_i \quad \stackrel{\text{def}}{\iff} \quad (\mathcal{T}, y) \equiv_i^{\{\downarrow, \overset{*}{\rightarrow}, \overset{*}{\leftarrow}\}} (\mathcal{T}', y')$$

Probaremos que $\mathcal{Z} := \{Z_j : 0 \leq j \leq \ell\}$ es una ℓ -bisimulación entre (\mathcal{T}, x) y (\mathcal{T}', x') . Notar que por hipótesis tenemos que $(x, x') \in Z_{\ell}$.

Supongamos que $(y, y') \in Z_h$.

Harmony. Supongamos que $\text{label}(y) = a$ y llamemos $\varphi = a$. Como

$(\mathcal{T}, y) \models \varphi$ y $(\mathcal{T}, y) \equiv_h^{\{\downarrow, \overset{*}{\rightarrow}, \overset{*}{\leftarrow}\}} (\mathcal{T}', y')$, debido a que $\text{ad}(\varphi) = 0 \leq h$ tenemos que $(\mathcal{T}', y') \models \varphi$.

Zig. Supongamos $\max\{n+k, m+t\} \leq h$. Sea $\tilde{v} \in \{\downarrow, \overset{*}{\rightarrow}, \overset{*}{\leftarrow}\}$ - $\text{Tray}_{n,k}^o(\mathcal{T}, y)$ y $\tilde{w} \in \{\downarrow, \overset{*}{\rightarrow}, \overset{*}{\leftarrow}\}$ - $\text{Tray}_{m,t}^e(\mathcal{T}, y)$ tales que $\text{data}(v_{n+k}) \odot \text{data}(w_{m+t})$ con $\odot \in \{=, \neq\}$. Consideremos el siguiente conjunto

$$P := \{(\tilde{v}', \tilde{w}') \in \{\downarrow, \overset{*}{\rightarrow}, \overset{*}{\leftarrow}\}\text{-Tray}_{n,k}^o(\mathcal{T}', y') \times \{\downarrow, \overset{*}{\rightarrow}, \overset{*}{\leftarrow}\}\text{-Tray}_{m,t}^e(\mathcal{T}', y') : \text{data}(v'_{n+k}) \odot \text{data}(w'_{m+t})\}$$

Veamos que P es no vacío. Si $\alpha_{\tilde{v}}$ y $\alpha_{\tilde{w}}$ son las expresiones de camino asociadas a \tilde{v} y \tilde{w} , entonces como $(\mathcal{T}, y) \models \langle \alpha_{\tilde{v}} \odot \alpha_{\tilde{w}} \rangle$ y $\text{ad}(\langle \alpha_{\tilde{v}} \odot \alpha_{\tilde{w}} \rangle) \leq h$, dado que $(\mathcal{T}, y) \equiv_h^{\{\downarrow, \overset{*}{\rightarrow}, \overset{*}{\leftarrow}\}} (\mathcal{T}', y')$ tenemos que $(\mathcal{T}', y') \models \langle \alpha_{\tilde{v}} \odot \alpha_{\tilde{w}} \rangle$. Por lo tanto, existen $\tilde{v}' \in \{\downarrow, \overset{*}{\rightarrow}, \overset{*}{\leftarrow}\}$ - $\text{Tray}_{n,k}^o(\mathcal{T}', y')$ y $\tilde{w}' \in \{\downarrow, \overset{*}{\rightarrow}, \overset{*}{\leftarrow}\}$ - $\text{Tray}_{m,t}^e(\mathcal{T}', y')$ tales que $(\mathcal{T}', \tilde{v}') \models \alpha_{\tilde{v}}$, $(\mathcal{T}', \tilde{w}') \models \alpha_{\tilde{w}}$ y $\text{data}(v'_{n+k}) \odot \text{data}(w'_{m+t})$. O sea, $(\tilde{v}', \tilde{w}') \in P \neq \emptyset$.

Supongamos que no vale **Zig** en (y, y') . Luego, por definición de \mathcal{Z} , para todo $(\tilde{v}', \tilde{w}') \in P$ tenemos que alguna de las siguientes no se cumple

$$\exists i \in \{0, \dots, n+k\} : (\mathcal{T}, v_i) \not\equiv_{h-i}^{\{\downarrow, \overset{*}{\rightarrow}, \overset{*}{\leftarrow}\}} (\mathcal{T}', v'_i) \quad (6)$$

$$\exists j \in \{0, \dots, m+t\} : (\mathcal{T}, w_j) \not\equiv_{h-j}^{\{\downarrow, \overset{*}{\rightarrow}, \overset{*}{\leftarrow}\}} (\mathcal{T}', w'_j) \quad (7)$$

Ahora, para cada $(\tilde{v}', \tilde{w}') \in P$ definimos expresiones de nodo

$$\begin{array}{ll} \varphi_{(\tilde{v}', \tilde{w}')}^0, \varphi_{(\tilde{v}', \tilde{w}')}^1, \dots, \varphi_{(\tilde{v}', \tilde{w}')}^{n+k} & \psi_{(\tilde{v}', \tilde{w}')}^0, \psi_{(\tilde{v}', \tilde{w}')}^1, \dots, \psi_{(\tilde{v}', \tilde{w}')}^{m+t} \\ \text{ad}(\varphi_{(\tilde{v}', \tilde{w}')}^i) \leq h-i & \text{ad}(\psi_{(\tilde{v}', \tilde{w}')}^j) \leq h-j \\ \forall i \in \{0, \dots, n+k\} & \forall j \in \{0, \dots, m+t\} \end{array}$$

de la siguiente manera:

- Supongamos que vale **6** entonces consideremos el mínimo i tal que $(\mathcal{T}, v_i) \not\equiv_i^{\{\downarrow, \overset{*}{\rightarrow}, \overset{*}{\leftarrow}\}} (\mathcal{T}', v'_i)$. Luego, existe $\varphi_{(\tilde{v}', \tilde{w}')}^i$ con $\text{ad}(\varphi_{(\tilde{v}', \tilde{w}')}^i) \leq h-i$ tal que $(\mathcal{T}, v_i) \models \varphi_{(\tilde{v}', \tilde{w}')}^i$ pero $(\mathcal{T}', v'_i) \not\models \varphi_{(\tilde{v}', \tilde{w}')}^i$. Para el resto de los $k \in \{0, \dots, n+k\} \setminus \{i\}$ y $j \in \{0, \dots, m+t\}$ definimos $\varphi_{(\tilde{v}', \tilde{w}')}^k := \top$ y $\psi_{(\tilde{v}', \tilde{w}')}^j := \top$.
- Si no vale **6** entonces vale **7**. Consideremos el mínimo j tal que $(\mathcal{T}, w_j) \not\equiv_j^{\{\downarrow, \overset{*}{\rightarrow}, \overset{*}{\leftarrow}\}} (\mathcal{T}', w'_j)$. Luego, existe $\psi_{(\tilde{v}', \tilde{w}')}^j$ con $\text{ad}(\psi_{(\tilde{v}', \tilde{w}')}^j) \leq h-j$ tal que $(\mathcal{T}, w_j) \models \psi_{(\tilde{v}', \tilde{w}')}^j$ pero $(\mathcal{T}', w'_j) \not\models \psi_{(\tilde{v}', \tilde{w}')}^j$. Para el resto de los $k \in \{0, \dots, m+t\} \setminus \{j\}$ e $i \in \{0, \dots, n+k\}$ definimos $\psi_{(\tilde{v}', \tilde{w}')}^k := \top$ y $\varphi_{(\tilde{v}', \tilde{w}')}^i := \top$.

Luego, para cada $i \in \{0, \dots, n+k\}$ y $j \in \{0, \dots, m+t\}$ definimos

$$\Phi^i := \bigwedge_{(\tilde{v}', \tilde{w}') \in P} \varphi_{(\tilde{v}', \tilde{w}')}^i \quad \Psi^j := \bigwedge_{(\tilde{v}', \tilde{w}') \in P} \psi_{(\tilde{v}', \tilde{w}')}^j$$

Notar que P puede ser infinito, pero por el Lema 4.24, tanto Φ^i como Ψ^j están bien definidas, módulo equivalencia lógica. Finalmente, si $\tilde{v} = v_0 e_1 v_1 \dots e_n v_n \downarrow v_{n+1} \downarrow \dots \downarrow v_{n+k}$ y $\tilde{w} = w_0 o_1 w_1 \dots o_m w_m \downarrow w_{m+1} \downarrow \dots \downarrow w_{m+t}$ definimos

$$\begin{aligned} \alpha &:= [\Phi^0] e_1 [\Phi^1] \dots e_n [\Phi^n] \downarrow [\Phi^{n+1}] \downarrow \dots \downarrow [\Phi^{n+k}] & (\mathcal{T}, \tilde{v}) \models \alpha \\ \beta &:= [\Psi^0] o_1 [\Psi^1] \dots o_m [\Psi^m] \downarrow [\Psi^{m+1}] \downarrow \dots \downarrow [\Psi^{m+t}] & (\mathcal{T}, \tilde{w}) \models \beta \end{aligned}$$

Por lo tanto, $(\mathcal{T}, y) \models \langle \alpha \odot \beta \rangle$. Como $\text{ad}(\langle \alpha \odot \beta \rangle) \leq h$ y $(\mathcal{T}, y) \equiv_h^{\{\downarrow, \overset{*}{\rightarrow}, \overset{*}{\leftarrow}\}} (\mathcal{T}', y')$ entonces $(\mathcal{T}', y') \models \langle \alpha \odot \beta \rangle$. Luego, existe $(\tilde{v}', \tilde{w}') \in P$ tal que $(\mathcal{T}', \tilde{v}') \models \alpha$ y $(\mathcal{T}', \tilde{w}') \models \beta$. Pero, por construcción, $(\mathcal{T}', \tilde{v}') \not\models \alpha$ o $(\mathcal{T}', \tilde{w}') \not\models \beta$. En cualquiera de los casos, llegamos a un absurdo, el cual vino de suponer que no valía **Zig**.

Zag. Análogo a **Zig**. □

Teorema 4.27. Sean \mathcal{T} y \mathcal{T}' árboles ordenados. Luego,

$$(\mathcal{T}, x) \simeq_{\ell}^{\{\downarrow, \overset{*}{\rightarrow}, \overset{*}{\leftarrow}\}} (\mathcal{T}', x') \iff (\mathcal{T}, x) \equiv_{\ell}^{\{\downarrow, \overset{*}{\rightarrow}, \overset{*}{\leftarrow}\}} (\mathcal{T}', x')$$

Demostración. Es una consecuencia de la Proposición 4.23 y 4.26. □

4.5. Caracterización al estilo Hennessy-Milner

Proposición 4.28. Sean \mathcal{T} y \mathcal{T}' árboles ordenados finitamente ramificados. Entonces

$$(\mathcal{T}, x) \equiv^{\{\downarrow, \overset{*}{\rightarrow}, \overset{*}{\leftarrow}\}} (\mathcal{T}', x') \implies (\mathcal{T}, x) \simeq^{\{\downarrow, \overset{*}{\rightarrow}, \overset{*}{\leftarrow}\}} (\mathcal{T}', x')$$

En general, no es válida la vuelta.

Demostración. Probaremos que $Z := \{(y, y') \in T \times T' : (\mathcal{T}, y) \equiv^{\{\downarrow, \overset{*}{\rightarrow}, \overset{*}{\leftarrow}\}} (\mathcal{T}', y')\}$ es una $\{\downarrow, \overset{*}{\rightarrow}, \overset{*}{\leftarrow}\}$ -bisimulación entre (\mathcal{T}, x) y (\mathcal{T}', x') . Notar que por hipótesis tenemos que $(x, x') \in Z$.

Sea $(y, y') \in Z$.

Harmony. Supongamos que $\text{label}(y) = a$ y llamemos $\varphi = a$. Como $(\mathcal{T}, y) \models \varphi$ y $(\mathcal{T}, y) \equiv^{\{\downarrow, \overset{*}{\rightarrow}, \overset{*}{\leftarrow}\}} (\mathcal{T}', y')$ tenemos que $(\mathcal{T}', y') \models \varphi$. Por lo tanto, $\text{label}(y) = \text{label}(y')$.

Zig. Supongamos $\tilde{v} \in \{\downarrow, \overset{*}{\rightarrow}, \overset{*}{\leftarrow}\}$ -Tray $_{n,k}^o(\mathcal{T}, y)$ y $\tilde{w} \in \{\downarrow, \overset{*}{\rightarrow}, \overset{*}{\leftarrow}\}$ -Tray $_{m,t}^e(\mathcal{T}, y)$ tales que $data(v_{n+k}) \odot data(w_{m+t})$ con $\odot \in \{=, \neq\}$. Consideremos el siguiente conjunto

$$P := \{(\tilde{v}', \tilde{w}') \in \{\downarrow, \overset{*}{\rightarrow}, \overset{*}{\leftarrow}\}\text{-Tray}_{n,k}^o(\mathcal{T}', y') \times \{\downarrow, \overset{*}{\rightarrow}, \overset{*}{\leftarrow}\}\text{-Tray}_{m,t}^e(\mathcal{T}', y') : data(v'_{n+k}) \odot data(w'_{m+t})\}$$

Veamos que P es no vacío. Si $\alpha_{\tilde{v}}$ y $\alpha_{\tilde{w}}$ son las expresiones de camino asociadas a \tilde{v} y \tilde{w} , entonces como $(\mathcal{T}, y) \models \langle \alpha_{\tilde{v}} \odot \alpha_{\tilde{w}} \rangle$ y $(\mathcal{T}, y) \equiv \{\downarrow, \overset{*}{\rightarrow}, \overset{*}{\leftarrow}\}$ (\mathcal{T}', y') tenemos que $(\mathcal{T}', y') \models \langle \alpha_{\tilde{v}} \odot \alpha_{\tilde{w}} \rangle$. Por lo tanto, existen $\tilde{v}' \in \{\downarrow, \overset{*}{\rightarrow}, \overset{*}{\leftarrow}\}\text{-Tray}_{n,k}^o(\mathcal{T}', y')$ y $\tilde{w}' \in \{\downarrow, \overset{*}{\rightarrow}, \overset{*}{\leftarrow}\}\text{-Tray}_{m,t}^e(\mathcal{T}', y')$ tales que $(\mathcal{T}', \tilde{v}') \models \alpha_{\tilde{v}}$, $(\mathcal{T}', \tilde{w}') \models \alpha_{\tilde{w}}$ y $data(v'_{n+k}) \odot data(w'_{m+t})$. O sea, $(\tilde{v}', \tilde{w}') \in P \neq \emptyset$.

Supongamos que no vale **Zig** en (y, y') . Luego, por definición de \mathcal{Z} , para todo $(\tilde{v}', \tilde{w}') \in P$ tenemos que alguna de las siguientes no se cumple

$$\exists i \in \{0, \dots, n+k\} : (\mathcal{T}, v_i) \not\equiv \{\downarrow, \overset{*}{\rightarrow}, \overset{*}{\leftarrow}\} (\mathcal{T}', v'_i) \quad (8)$$

$$\exists j \in \{0, \dots, m+t\} : (\mathcal{T}, w_j) \not\equiv \{\downarrow, \overset{*}{\rightarrow}, \overset{*}{\leftarrow}\} (\mathcal{T}', w'_j) \quad (9)$$

Ahora, para cada $(\tilde{v}', \tilde{w}') \in P$ definimos expresiones de nodo

$$\varphi_{(\tilde{v}', \tilde{w}')}^0, \varphi_{(\tilde{v}', \tilde{w}')}^1, \dots, \varphi_{(\tilde{v}', \tilde{w}')}^{n+k} \quad \psi_{(\tilde{v}', \tilde{w}')}^0, \psi_{(\tilde{v}', \tilde{w}')}^1, \dots, \psi_{(\tilde{v}', \tilde{w}')}^{m+t}$$

de la siguiente manera:

- Supongamos que vale **8** entonces consideremos el mínimo i tal que $(\mathcal{T}, v_i) \not\equiv \{\downarrow, \overset{*}{\rightarrow}, \overset{*}{\leftarrow}\} (\mathcal{T}', v'_i)$. Luego, existe $\varphi_{(\tilde{v}', \tilde{w}')}^i$ tal que $(\mathcal{T}, v_i) \models \varphi_{(\tilde{v}', \tilde{w}')}^i$ pero $(\mathcal{T}', v'_i) \not\models \varphi_{(\tilde{v}', \tilde{w}')}^i$. Para el resto de los $k \in \{0, \dots, n+k\} \setminus \{i\}$ y $j \in \{0, \dots, m+t\}$ definimos $\varphi_{(\tilde{v}', \tilde{w}')}^k := \top$ y $\psi_{(\tilde{v}', \tilde{w}')}^j := \top$.
- Si no vale **8** entonces vale **9**. Consideremos el mínimo j tal que $(\mathcal{T}, w_j) \not\equiv \{\downarrow, \overset{*}{\rightarrow}, \overset{*}{\leftarrow}\} (\mathcal{T}', w'_j)$. Luego, existe $\psi_{(\tilde{v}', \tilde{w}')}^j$ tal que $(\mathcal{T}, w_j) \models \psi_{(\tilde{v}', \tilde{w}')}^j$ pero $(\mathcal{T}', w'_j) \not\models \psi_{(\tilde{v}', \tilde{w}')}^j$. Para el resto de los $k \in \{0, \dots, m+t\} \setminus \{j\}$ e $i \in \{0, \dots, n+k\}$ definimos $\psi_{(\tilde{v}', \tilde{w}')}^k := \top$ y $\varphi_{(\tilde{v}', \tilde{w}')}^i := \top$.

Luego, para cada $i \in \{0, \dots, n+k\}$ y $j \in \{0, \dots, m+t\}$ definimos

$$\Phi^i := \bigwedge_{(\tilde{v}', \tilde{w}') \in P} \varphi_{(\tilde{v}', \tilde{w}')}^i \quad \Psi^j := \bigwedge_{(\tilde{v}', \tilde{w}') \in P} \psi_{(\tilde{v}', \tilde{w}')}^j$$

Como P es finito por ser \mathcal{T}' finitamente ramificado, tanto Φ^i como Ψ^j están bien definidas. Finalmente, si $\tilde{v} = v_0 e_1 v_1 \dots e_n v_n \downarrow v_{n+1} \downarrow \dots \downarrow$

v_{n+k} y $\tilde{w} = w_0 o_1 w_1 \dots o_m w_m \downarrow w_{m+1} \downarrow \dots \downarrow w_{m+t}$ definimos

$$\begin{aligned}\alpha &:= [\Phi^0]e_1[\Phi^1] \dots e_n[\Phi^n] \downarrow [\Phi^{n+1}] \downarrow \dots \downarrow [\Phi^{n+k}] & (\mathcal{T}, \tilde{v}) \models \alpha \\ \beta &:= [\Psi^0]o_1[\Psi^1] \dots o_m[\Psi^m] \downarrow [\Psi^{m+1}] \downarrow \dots \downarrow [\Psi^{m+t}] & (\mathcal{T}, \tilde{w}) \models \beta\end{aligned}$$

Por lo tanto, $(\mathcal{T}, y) \models \langle \alpha \odot \beta \rangle$. Como $(\mathcal{T}, y) \equiv^{\{\downarrow, \overset{*}{\rightarrow}, \overset{*}{\leftarrow}\}} (\mathcal{T}', y')$ entonces $(\mathcal{T}', y') \models \langle \alpha \odot \beta \rangle$. Luego, existe $(\tilde{v}', \tilde{w}') \in P$ tal que $(\mathcal{T}', \tilde{v}') \models \alpha$ y $(\mathcal{T}', \tilde{w}') \models \beta$. Pero, por construcción, $(\mathcal{T}', \tilde{v}') \not\models \alpha$ o $(\mathcal{T}', \tilde{w}') \not\models \beta$. En cualquiera de los casos, llegamos a un absurdo, el cual vino de suponer que no valía **Zig**.

Zag. Análogo a **Zig**.

Como contraejemplo, se pueden considerar los modelos de la Figura 4. \square

Observación 4.29. Dada una $\{\downarrow, \overset{*}{\rightarrow}, \overset{*}{\leftarrow}\}$ -bisimulación Z , la familia $\mathcal{Z} = \{Z_j : 0 \leq j \leq \ell\}$ con $Z_j = Z$ para todo j es una $\{\downarrow, \overset{*}{\rightarrow}, \overset{*}{\leftarrow}\}_\ell$ -bisimulación. Por lo tanto,

$$(\mathcal{T}, x) \simeq^{\{\downarrow, \overset{*}{\rightarrow}, \overset{*}{\leftarrow}\}} (\mathcal{T}', x') \implies (\mathcal{T}, x) \simeq_\ell^{\{\downarrow, \overset{*}{\rightarrow}, \overset{*}{\leftarrow}\}} (\mathcal{T}', x')$$

Teorema 4.30. Sean \mathcal{T} y \mathcal{T}' árboles ordenados. Luego,

$$(\mathcal{T}, x) \simeq^{\{\downarrow, \overset{*}{\rightarrow}, \overset{*}{\leftarrow}\}} (\mathcal{T}', x') \implies (\mathcal{T}, x) \equiv^{\{\downarrow, \overset{*}{\rightarrow}, \overset{*}{\leftarrow}\}} (\mathcal{T}', x')$$

Demostración. Supongamos $(\mathcal{T}, x) \models \varphi$. Por la Proposición 4.10 existe φ' normal tal que $\varphi \equiv \varphi'$. Luego, $\varphi' \in \ell\text{-XPath}_{=}(\downarrow, \overset{*}{\rightarrow}, \overset{*}{\leftarrow})$ para cierto ℓ . La Observación 4.29 nos dice que $(\mathcal{T}, x) \simeq_\ell^{\{\downarrow, \overset{*}{\rightarrow}, \overset{*}{\leftarrow}\}} (\mathcal{T}', x')$, y por el Teorema 4.27 llegamos a que $(\mathcal{T}, x) \equiv_\ell^{\{\downarrow, \overset{*}{\rightarrow}, \overset{*}{\leftarrow}\}} (\mathcal{T}', x')$. Por lo tanto, $(\mathcal{T}', x') \models \varphi'$ y esto implica que $(\mathcal{T}', x') \models \varphi$. De manera análoga, si $(\mathcal{T}', x') \models \varphi$ entonces $(\mathcal{T}, x) \models \varphi$. \square

5. Conclusiones

A lo largo de esta tesis hemos estudiado distintos fragmentos de XPath. En la Sección 2 presentamos lo hecho para $\text{XPath}_=(\downarrow)$ en [7] respecto a la caracterización al estilo Hennessy-Milner. En la Sección 3 y 4 adaptamos esto a los fragmentos $\text{XPath}_=(\downarrow, \overset{*}{\rightarrow})$ y $\text{XPath}_=(\downarrow, \overset{*}{\rightarrow}, \overset{*}{\leftarrow})$. Específicamente, probamos que las nociones acotadas de equivalencia lógica y bisimulación son la misma sobre modelos cuyos árboles tienen definido un orden sobre los hijos de cada nodo. Esto resulta algo novedoso debido a que los fragmentos considerados no fueron estudiados con anterioridad. Lo más semejante es lo hecho por D. Figueira en [9] para $\text{XPath}_=(\downarrow_*, \overset{*}{\rightarrow}, \overset{*}{\leftarrow})$, aunque los problemas tratados tienen que ver con la satisfacibilidad y decidibilidad de las expresiones de nodo. No obstante, la forma normal que allí se presenta sirvió de referencia para definir las expresiones normales en $\text{XPath}_=(\downarrow, \overset{*}{\rightarrow}, \overset{*}{\leftarrow})$.

Trabajos a futuro

- La caracterización al estilo van Benthem dada en la Sección 2 para $\text{XPath}_=(\downarrow)$ no fue estudiada para los nuevos fragmentos considerados. Queda pendiente adaptar este resultado para $\text{XPath}_=(\downarrow, \overset{*}{\rightarrow})$ y $\text{XPath}_=(\downarrow, \overset{*}{\rightarrow}, \overset{*}{\leftarrow})$.
- Aunque aquí no fueron mencionados, otros problemas estudiados para $\text{XPath}_=(\downarrow)$ en [8] por S. Figueira, Descotte y Abriola tienen que ver con la definibilidad y separabilidad de clases de modelos. A futuro, se pueden pensar estos mismos problemas para los nuevos fragmentos.
- Otras preguntas que surgen para lógicas en general, es si cumplen la *propiedad de interpolación de Craig*. Esto quiere decir, si para todo par de expresiones φ y ψ tales que $\varphi \models \psi$ existe θ tal que $\varphi \models \theta \models \psi$. Sería interesante analizar esta propiedad dentro de los fragmentos estudiados.

Referencias

- [1] P. Blackburn, J. van Benthem and F. Wolter. *Handbook of Modal Logic*. Elsevier, 2006.
- [2] D. Marker. *Model Theory: An Introduction*. Springer, 2002.
- [3] J. van Benthem. *Modal Correspondence Theory*. PhD thesis, Universiteit van Amsterdam, 1976.
- [4] M. Otto. *Elementary proof of the van Benthem-Rosen characterisation theorem*. Tech. rep. 2342, Fachbereich Mathematik, Technische Universität Darmstadt. 2004.
- [5] J. Clark and S. DeRose. *XML path language (XPath)*. Website, 1999. W3C Recommendation.
- [6] M. Marx and M. de Rijke. *Semantic characterizations of navigational XPath*. SIGMOD Record, 34(2):41–46, 2005.
- [7] Diego Figueira, Santiago Figueira and Carlos Areces. *Model Theory of XPath on Data Trees. Part I: Bisimulation and Characterization*. Journal of Artificial Intelligence Research 53:271-314, 2015.
- [8] Sergio Abriola, María Emilia Descotte and Santiago Figueira. *Model Theory of XPath on Data Trees. Part II: Binary Bisimulation and Definability*. Information and Computation. 255:195-223, 2017.
- [9] Diego Figueira. *On XPath with Transitive Axes and Data Tests*. PODS'13, 2013.