Tema 9: Modelos de Herbrand

José A. Alonso Jiménez
Andrés Cordón Franco

Dpto. de Ciencias de la Computación e Inteligencia Artificial

Universidad de Sevilla

Reducción de la LPO básica a proposicional

- **Observación:**
  - En este tema sólo se consideran lenguajes de primer orden sin igualdad.
- **Reducción de la LPO básica a proposicional**
  - Def.: Una fórmula básica es una fórmula sin variables ni cuantificadores.
  - Prop.: Sea S un conjunto de fórmulas básicas. Son equivalentes:
    1. S es consistente en el sentido de la lógica de primer orden.
    2. S es consistente en el sentido de la lógica proposicional.

Reducción de la LPO básica a proposicional

- **Ejemplos:**
  - $\{P(a) \lor P(b), \neg P(b) \lor P(c), P(a) \rightarrow P(c)\}$
    es consistente en el sentido de la lógica de primer orden (con modelos $\mathcal{I}_4, \mathcal{I}_6, \mathcal{I}_8$).
  - $\{P(a) \lor P(b), \neg P(b) \lor P(c), P(a) \rightarrow P(c), \neg P(c)\}$
    es inconsistente en el sentido de la lógica de primer orden.

	P^I	$P(a) \lor P(b)$	$\neg P(b) \lor P(c)$	$P(a) \rightarrow P(c)$	$\neg P(c)$
$\mathcal{I}_1$	$\emptyset$	0	1	1	1
$\mathcal{I}_2$	$\{c^I\}$	0	1	1	0
$\mathcal{I}_3$	$\{b^I\}$	1	0	1	1
$\mathcal{I}_4$	$\{b^I, c^I\}$	1	1	1	0
$\mathcal{I}_5$	$\{a^I\}$	1	1	0	1
$\mathcal{I}_6$	$\{a^I, c^I\}$	1	1	1	0
$\mathcal{I}_7$	$\{a^I, b^I\}$	1	0	0	1
$\mathcal{I}_8$	$\{a^I, b^I, c^I\}$	1	1	1	0

Reducción de la LPO básica a proposicional

- **Ejemplos:**
  - $\{P(a) \lor P(b), \neg P(b) \lor P(c), P(a) \to P(c)\}$
    es consistente en el sentido proposicional (con modelos v_4, v_6, v_8).
  - $\{P(a) \lor P(b), \neg P(b) \lor P(c), P(a) \to P(c), \neg P(c)\}$
    es inconsistente en el sentido proposicional.

Se consideran los cambios $P(a)/p, P(b)/q, P(c)/r$

	p	q	r	$p \lor q$	$\neg q \lor r$	$p \to r$	$\neg r$
v_1	0	0	0	0	1	1	1
v_2	0	0	1	0	1	1	0
v_3	0	1	0	1	0	1	1
v_4	0	1	1	1	1	1	0
v_5	1	0	0	1	1	0	1
v_6	1	0	1	1	1	1	0
v_7	1	1	0	1	0	0	1
v_8	1	1	1	1	1	1	0

Interpretaciones de Herbrand

- **Notación:**
  - L **representa un lenguaje de primer orden sin igualdad.**
  - $\mathcal{C}$ **es el conjunto de constantes de** L.
  - $\mathcal{F}$ **es el conjunto de símbolos de función de** L.
  - $\mathcal{R}$ **es el conjunto de símbolos de relación de** L.
  - $\mathcal{F}_n$ **es el conjunto de símbolos de función** n**–aria de** L.
  - $\mathcal{R}_n$ **es el conjunto de símbolos de relación** n**–aria de** L.
  - f/n **indica que** f **es un símbolo de función** n**–aria de** L.
  - P/n **indica que** f **es un símbolo de relación** n**–aria de** L.

Interpretaciones de Herbrand: Universo de Herbrand

- **Universo de Herbrand de L:**
  - **Def.: El universo de Herbrand de L es el conjunto de los términos básicos de L. Se representa por UH(L).**
  - **Prop.: UH(L) $= \bigcup_{i \geq 0} H_i(L)$, donde $H_i(L)$ es el nivel i del UH(L) definido por**

$$
\begin{aligned}
H_0(L) &= \begin{cases} \mathcal{C}, & \text{si } \mathcal{C} \neq \emptyset; \\ \{a\}, & \text{en caso contrario. } (a \text{ es una nueva constante}). \end{cases} \\
H_{i+1}(L) &= H_i(L) \cup \{f(t_1, \dots, t_n) : f \in \mathcal{F}_n \text{ y } t_1, \dots, t_n \in H_i(L)\}
\end{aligned}
$$

  - **Prop.: UH(L) es finito syss L no tiene símbolos de función.**
  - **Ejemplos:**
    - **Si $\mathcal{C} = \{a, b, c\}$ y $\mathcal{F} = \emptyset$, entonces**

$$
\begin{aligned}
H_0(L) &= \{a, b, c\} \\
H_1(L) &= \{a, b, c\} \\
&\vdots \\
\mathrm{UH}(L) &= \{a, b, c\}
\end{aligned}
$$

Interpretaciones de Herbrand: Universo de Herbrand

- **Ejemplos:**
  - Si $\mathcal{C} = \emptyset$ y $\mathcal{F} = \{f/1\}$, entonces

$$
\begin{array}{ll}
H_0(L) & = \{a\} \\
H_1(L) & = \{a, f(a)\} \\
H_2(L) & = \{a, f(a), f(f(a))\} \\
\vdots & \\
\mathrm{UH}(L) & = \{a, f(a), f(f(a)), \dots\} = \{f^i(a) : i \in \mathbb{N}\}
\end{array}
$$

  - Si $\mathcal{C} = \{a, b\}$ y $\mathcal{F} = \{f/1, g/1\}$, entonces

$$
\begin{array}{lll}
H_0(L) & = \{a, b\} & \\
H_1(L) & = \{a, b, f(a), f(b), g(a), g(b)\} & \\
H_2(L) & = \{a, b, f(a), f(b), g(a), g(b), & f(f(a)), f(f(b)), f(g(a)), f(g(b)), \\
 & & g(f(a)), g(f(b)), g(g(a)), g(g(b))\ \} \\
\vdots & &
\end{array}
$$

  - Si $\mathcal{C} = \{a, b\}$ y $\mathcal{F} = \{f/2\}$, entonces

$$
\begin{array}{ll}
H_0(L) & = \{a, b\} \\
H_1(L) & = \{a, b, f(a, a), f(a, b), f(b, a), f(b, b)\} \\
H_2(L) & = \{a, b, f(a, a), f(a, b), f(b, a), f(b, b), f(a, f(a, a)), f(a, f(a, b)), \dots\} \\
\vdots &
\end{array}
$$

Interpretaciones de Herbrand: Base de Herbrand

- **Base de Herbrand:**
  - **Def.:** La base de Herbrand de L es el conjunto de los átomos básicos de L. Se representa por BH(L).
  - **Prop.:** $\mathrm{BH}(L) = \{P(t_1, \dots, t_n) : P \in \mathcal{R}_n \text{ y } t_1, \dots, t_n \in \mathrm{UH}(L)\}$.
  - **Prop.:** BH(L) es finita syss L no tiene símbolos de función.
  - **Ejemplos:**
    - Si $\mathcal{C} = \{a, b, c\}$, $\mathcal{F} = \emptyset$ y $\{\mathcal{R} = P/1\}$, entonces

$$\begin{aligned} \mathrm{UH}(L) &= \{a, b, c\} \\ \mathrm{BH}(L) &= \{P(a), P(b), P(c)\} \end{aligned}$$

    - Si $\mathcal{C} = \{a\}$, $\mathcal{F} = \{f/1\}$ y $\mathcal{R} = \{P/1, Q/1, R/1\}$, entonces

$$\begin{aligned} \mathrm{UH}(L) &= \{a, f(a), f(f(a)), \dots\} = \{f^i(a) : i \in \mathbb{N}\} \\ \mathrm{BH}(L) &= \{P(a), Q(a), R(a), P(f(a)), Q(f(a)), R(f(a)), \dots\} \end{aligned}$$

Interpretaciones de Herbrand

- **Interpretaciones de Herbrand:**
  - Def.: Una interpretación $\mathcal{I} = (U, I)$ de L es de Herbrand si
    - U es el universo de Herbrand de L;
    - $I(c) = c$, para cada constante c de L;
    - $I(f) = f$, para cada símbolo de función f de L.
  - Prop.: Sea $\mathcal{I}$ una interpretación de Herbrand de L. Si t es un término básico de L, entonces $\mathcal{I}(t) = t$.
  - Prop.: Una interpretación de Herbrand queda determinada por un subconjunto de la base de Herbrand, el conjunto de átomos básicos verdaderos en esa interpretación.
  - Ejemplo: Si $\mathcal{C} = \{a, b, c\}$, $\mathcal{F} = \emptyset$ y $\mathcal{R} = \{P/1\}$, entonces las interpretaciones de Herbrand de L son

n	1	2	3	4	5	6	7	8
$I_n(P)$	$\emptyset$	$\{c\}$	$\{b\}$	$\{b, c\}$	$\{a\}$	$\{a, c\}$	$\{a, b\}$	$\{a, b, c\}$
IH_n	$\emptyset$	$\{P(c)\}$	$\{P(b)\}$	$\{P(b), P(c)\}$	$\{P(a)\}$	$\{P(a), P(c)\}$	$\{P(a), P(b)\}$	$\{P(a), P(b), P(c)\}$

Modelos de Herbrand

- **Modelos de Herbrand:**
  - **Nota: Las definiciones de universo de Herbrand, base de Herbrand e interpretación de Herbrand definidas para un lenguaje se extienden a fórmulas y conjuntos de fórmulas considerando el lenguaje formado por los símbolos no lógicos que aparecen.**
  - **Def.: Un modelo de Herbrand de una fórmula F es una interpretación de Herbrand de F que es modelo de F.**
  - **Def.: Un modelo de Herbrand de un conjunto de fórmulas S es una interpretación de Herbrand de S que es modelo de S.**
  - **Ejemplo: Los modelos de Herbrand de $\{P(a) \vee P(b), \neg P(b) \vee P(c), P(a) \to P(c)\}$ son $\{P(b), P(c)\}, \{P(a), P(c)\}$ y $\{P(a), P(b), P(c)\}$ (ver página ??).**
  - **Ejemplo: Sea $S = \{(\forall x)(\forall y)[Q(b,x) \to P(a) \vee R(y)], P(b) \to \neg(\exists z)(\exists u)Q(z,u)\}$. Entonces,**

    $$\begin{aligned} \mathrm{UH}(S) &= \{a, b\} \\ \mathrm{BH}(S) &= \{P(a), P(b), Q(a,a), Q(a,b), Q(b.a), Q(b,b), R(a), R(b)\} \end{aligned}$$

    Un modelo de Herbrand de S es $\{P(a)\}$.

Interpretación de Herbrand correspondiente

- **Interpretación de Herbrand correspondiente a una interpretación:**
  - Sea $S = \{\{\neg Q(b,x), P(a), R(y)\}, \{\neg P(b), \neg Q(z,u)\}\}$ e $\mathcal{I} = (\{1,2\}, I)$ con $a^I = 1, b^I = 2, P^I = \{1\}, Q^I = \{(1,1),(2,2)\}, R^I = \{2\}$. Entonces, $\mathcal{I} \models S$.

    Cálculo de la interpretación de Herbrand $\mathcal{I}^*$ correspondiente a $\mathcal{I}$:

$$
\begin{array}{ll}
\mathcal{I}^* & = (\mathrm{UH}(S), I^*) \\
\mathrm{UH}(S) & = \{a, b\} \\
\mathrm{BH}(S) & = \{P(a), P(b), Q(a,a), Q(a,b), Q(b.a), Q(b,b), R(a), R(b)\} \\
I^*(P(a)) & = P^I(a^I) = P^I(1) = \mathrm{V} \\
I^*(P(b)) & = P^I(b^I) = P^I(2) = \mathrm{F} \\
I^*(Q(a,a) & = Q^I(a^I, a^I) = Q^I(1,1) = \mathrm{V} \\
I^*(Q(a,b) & = Q^I(a^I, b^I) = Q^I(1,2) = \mathrm{F} \\
I^*(Q(b,a) & = Q^I(b^I, a^I) = Q^I(2,1) = \mathrm{F} \\
I^*(Q(b,b) & = Q^I(b^I, b^I) = Q^I(2,2) = \mathrm{V} \\
I^*(R(a)) & = R^I(a^I) = R^I(1) = \mathrm{F} \\
I^*(R(b)) & = R^I(b^I) = R^I(2) = \mathrm{V}
\end{array}
$$

    $I^* = \{P(a), Q(a,a), Q(b,b), R(b)\}$

    $\mathcal{I}^* \models S$.

Interpretación de Herbrand correspondiente

- **Interpretación de Herbrand correspondiente a una interpretación:**
  - Sea S el conjunto de cláusulas $\{\{P(a)\}, \{Q(y, f(a))\}\}$ e $\mathcal{I} = (\{1,2\}, I)$ con $a^I = 1, f^I = \{(1,2),(2,1)\}, P^I = \{1\}, Q^I = \{(1,2),(2,2)\}$. Entonces, $\mathcal{I} \models S$.

    Cálculo de la interpretación de Herbrand $\mathcal{I}^*$ correspondiente a $\mathcal{I}$:

$$\begin{array}{ll}
\mathcal{I}^* & = (\mathrm{UH}(S), I^*) \\
\mathrm{UH}(S) & = \{a, f(a), f(f(a)), \dots\} = \{f^i(a) : i \in \mathbb{N}\} \\
\mathrm{BH}(S) & = \{P(f^n(a)) : n \in \mathbb{N}\} \cup \{Q(f^n(a), f^m(a)) : n, m \in \mathbb{N}\} \\
I^*(P(a)) & = P^I(a^I) = P^I(1) = \mathrm{V} \\
I^*(P(f(a))) & = P^I(f^I(a^I)) = P^I(f^I(1)) = P^I(2) = \mathrm{F} \\
I^*(P(f(f(a)))) & = P^I(f^I(f^(a^I))) = P^I(1) = \mathrm{V} \\
I^*(P(f^n(a))) & = \begin{cases} \mathrm{V}, & \text{si } n \text{ es par;} \\ \mathrm{F}, & \text{en caso contrario.} \end{cases} \\
I^*(Q(f^n(a), f^m(a))) & = \begin{cases} \mathrm{V}, & \text{si } m \text{ es impar;} \\ \mathrm{F}, & \text{en caso contrario.} \end{cases}
\end{array}$$

    $I^* = \{P(f^{2n}(a)) : n \in \mathbb{N}\} \cup \{Q(f^n(a), f^{2m+1}(a)) : n, m \in \mathbb{N}\}$

    $\mathcal{I}^* \models S$.

Consistencia mediante modelos de Herbrand

- **Consistencia mediante modelos de Herbrand:**
  - **Prop.: Sea S un conjunto de fórmulas básicas. Son equivalentes:**
    1. **S es consistente.**
    2. **S tiene un modelo de Herbrand.**
  - **Prop.: Sea S un conjunto de cláusulas. Si $\mathcal{I}^*$ es una interpretación de Herbrand correspondiente a un modelo $\mathcal{I}$ de S, entonces $\mathcal{I}^*$ es un modelo de S.**
  - **Prop.: Sea S un conjunto de cláusulas. Son equivalentes:**
    1. **S es consistente.**
    2. **S tiene un modelo de Herbrand.**
  - **Prop.: Sea S un conjunto de cláusulas. Son equivalentes:**
    1. **S es inconsistente.**
    2. **S no tiene ningún modelo de Herbrand.**
  - **Prop.: Existen conjuntos de fórmulas consistentes sin modelos de Herbrand.**

Consistencia mediante modelos de Herbrand

- **Ejemplo de conjunto consistente sin modelos de Herbrand:**
  - Sea $S = \{(\exists x)P(x), \neg P(a)\}$. Entonces,
    - S es consistente.

      $\mathcal{I} \models S$ con $\mathcal{I} = (\{1,2\}, I)$, $a^I = 1$ y $P^I = \{2\}$.
    - S no tiene modelos de Herbrand

      $\mathrm{UH}(S) = \{a\}$

      $\mathrm{BH}(S) = \{P(a)\}$

      Las interpretaciones de Herbrand de S son $\emptyset$ y $\{P(a)\}$.

      $\emptyset \not\models S$

      $\{P(a)\} \not\models S$

Extensiones de Herbrand

- **Instancias básicas de una cláusula**
  - **Def.:** Una sustitución σ (de L) es una aplicación $\sigma : \text{Var} \to \text{Térm}(L)$.
  - **Def.:** Sea $C = \{L_1, \dots, L_n\}$ una cláusula de L y σ una sustitución de L. Entonces, $C\sigma = \{L_1\sigma, \dots, L_n\sigma\}$ es una instancia de C.
  - **Ejemplo:** Sea $C = \{P(x, a), \neg P(x, f(y))\}$.

    $$C[x/a, y/f(a)] = \{P(f(a), a), \neg P(f(a), f(f(a)))\}$$
  - **Def.:** $C\sigma$ es una instancia básica de C si todos los literales de $C\sigma$ son básicos.
  - **Ejemplo:** Sea $C = \{P(x, a), \neg P(x, f(y))\}$.

    $\{P(f(a), a), \neg P(f(a), f(f(a)))\}$ es una instancia básica de C.
    $\{P(f(a), a), \neg P(f(f(a)), f(a))\}$ no es una instancia básica de C.
    $\{P(x, a), \neg P(f(f(a)), f(a))\}$ no es una instancia básica de C.

Extensiones de Herbrand

- **Extensiones de Herbrand**
  - **Def.:** La extensión de Herbrand de un conjunto de cláusulas S es el conjunto de fórmulas

$$\mathrm{EH}(S) = \{C\sigma : C \in S \text{ y, para toda variable } x \text{ en } C,\ \sigma(x) \in \mathrm{UH}(S)\}.$$

  - **Prop.:** $\mathrm{EH}(L) = \bigcup_{i \geq 0} \mathrm{EH}_i(L)$, donde $\mathrm{EH}_i(L)$ es el nivel i de la $\mathrm{EH}(L)$ definido por

$$\mathrm{EH}_i(S) = \{C\sigma : C \in S \text{ y, para toda variable } x \text{ en } C,\ \sigma(x) \in \mathrm{UH}_i(S)\}.$$

  - **Ejemplo:** Sea $S = \{\{P(x)\}, \{\neg P(f(x))\}\}$ (p. 8.17). Entonces,

$$\begin{aligned}
\mathrm{EH}_0(S) &= \{\{P(a)\}, \{\neg P(f(a))\}\} \\
\mathrm{EH}_1(S) &= \mathrm{EH}_0(S) \cup \{\{P(f(a))\}, \{\neg P(f(f(a)))\}\} \\
\mathrm{EH}_2(S) &= \mathrm{EH}_1(S) \cup \{\{P(f(f(a)))\}, \{\neg P(f(f(f(a))))\}\}
\end{aligned}$$

  - **Ejemplo:** Sea $S = \{\{\neg P(x), Q(x)\}, \{P(a)\}, \{\neg Q(z)\}\}$ (p. 8.21).
    Entonces, $\mathrm{EH}(S) = \{\{\neg P(a), Q(a)\}, \{P(a)\}, \{\neg Q(a)\}\}$.
  - **Ejemplo:** Sea $S = \{\{\neg P(x), Q(x)\}, \{\neg Q(y), R(y)\}, \{P(a)\}, \{\neg R(a)\}\}$ (p. 8.21).
    Entonces, $\mathrm{EH}(S) = \{\{\neg P(a), Q(a)\}, \{\neg Q(a), R(a)\}, \{P(a)\}, \{\neg R(a)\}\}$.

Teorema de Herbrand

- **Teorema de Herbrand:**
  - **Teorema de Herbrand:** Sea S un conjunto de cláusulas. Son equivalentes:
    1. S es consistente.
    2. EH(S) es consistente (en el sentido proposicional).
  - **Prop.:** Sea S un conjunto de cláusulas. Entonces, son equivalentes
    1. S es inconsistente.
    2. EH(S) tiene un subconjunto finito inconsistente (en el sentido proposicional).
    3. Para algún i, $\mathrm{EH}_i(S)$ es inconsistente (en el sentido proposicional).

Teorema de Herbrand

- **Semidecisón de la consistencia mediante el teorema de Herbrand:**
  - **Entrada: Un conjunto de cláusulas S.**
  - **Procedimiento:**
    1. **Hacer $i := 0$.**
    2. **Calcular $\mathrm{EH}_i(S)$.**
    3. **Si $\mathrm{EH}_i(S)$ es inconsistente (en el sentido proposicional), parar e indicar que S es inconsistente.**
    4. **Si $\mathrm{EH}_i(S)$ es consistente (en el sentido proposicional), hacer $i := i + 1$ y volver al paso 2.**

Teorema de Herbrand

- **Ejemplo:** $S = \{\{\neg P(x), Q(x)\}, \{P(a)\}, \{\neg Q(z)\}\}$ (p. ??) es inconsistente.

  $\text{EH}_0(S) = \{\{\neg P(a), Q(a)\}, \{P(a)\}, \{\neg Q(a)\}\}$ es inconsistente.

1	$\{\neg P(a), Q(a)\}$	
2	$\{P(a)\}$	
3	$\{\neg Q(a)\}$	
4	$\{Q(a)\}$	Res 1, 2
5	$\square$	Res 3, 4

- **Ejemplo:** $S = \{\{\neg P(x), Q(x)\}, \{\neg Q(y), R(y)\}, \{P(a)\}, \{\neg R(a)\}\}$ es inconsistente.

  $\text{EH}_0(S) = \{\{\neg P(a), Q(a)\}, \{\neg Q(a), R(a)\}, \{P(a)\}, \{\neg R(a)\}\}$.

1	$\{\neg P(a), Q(a)\}$	
2	$\{\neg Q(a), R(a)\}$	
3	$\{P(a)\}$	
4	$\{\neg R(a)\}$	
5	$\{Q(a)\}$	Res 1, 3
6	$\{R(a)\}$	Res 5, 2
7	$\square$	Res 6, 4

Teorema de Herbrand

- **Ejemplo:** $S = \{\{P(x)\}, \{\neg P(f(x))\}\}$ **es inconsistente (p. ??).**
  - $EH_0(S) = \{\{P(a)\}, \{\neg P(f(a))\}\}$ **es consistente**

    $\mathcal{I} = \{P(a)\} \models EH_0(S)$
  - $EH_1(S) = EH_0(S) \cup \{\{P(f(a))\}, \{\neg P(f(f(a)))\}\}$ **es inconsistente.**

1	$\{P(f(a))\}$	
2	$\{\neg P(f(a))\}$	
3	$\square$	Res $1, 2$

- **Ejemplo:** $S = \{\{\neg P(x), Q(f(x), x)\}, \{P(g(b))\}, \{\neg Q(y, z)\}\}$ **es inconsistente.**
  - $S' = \{\{\neg P(g(b)), Q(f(g(b)), g(b))\}, \{P(g(b))\}, \{\neg Q(f(g(b)), g(b))\}\} \subset EH(S)$ **es inconsistente.**

1	$\{\neg P(g(b)), Q(f(g(b)), g(b))\}$	
2	$\{P(g(b))\}$	
3	$\{\neg Q(f(g(b)), g(b))\}$	
4	$\{Q(f(g(b)), g(b))\}$	Res $1, 2$
5	$\square$	Res $3, 3$

Bibliografía

- M.L. Bonet *Apuntes de LPO.* (Univ. Politécnica de Cataluña, 2003) pp. 31–34.
- C.L. Chang y R.C.T. Lee *Symbolic logic and mechanical theorem proving* (Academic Press, 1973) pp. 51–62.
- M. Ojeda e I. Pérez *Lógica para la computación (Vol. 2: Lógica de Primer Orden)* (Ágora, 1997) pp. 59–74.
- E. Paniagua, J.L. Sánchez y F. Martín *Lógica computacional* (Thomson, 2003) pp. 160–169.
- L. Paulson *Logic and proof* (U. Cambridge, 2002) pp. 47–50.
- U. Schöning *Logic for computer scientists* (Birkäuser, 1989) pp. 70–78.

Bibliografía complementaria

- **A. Leitsch *Resolution calculus and proof complexity* (Technische Universität Wien, Austria, 1994) pp. 12–15.**