

Tema 2: Representación del conocimiento y Razonamiento

Ejemplo tema 1: Asignación

¿Cómo enfocaríamos la resolución del problema "asignar alumnos a aulas de clase"?

Analizamos:

- Los alumnos: pueden ir a cualquier clase (niveles)? ¿van separados o en bloques (zonas horarias)?
- Las clases: ¿hay tamaño máximo o mínimo? ¿tienen que tener el mismo tamaño? ¿hay un número fijo de aulas?
- La asignación ¿puedo medir lo buena o mala que es y por tanto optimizar? ¿O es sólo una cuestión de satisfacer restricciones?

Fijamos estrategia:

- Partiremos de una tabla de clases/aulas y un listado de alumnos.
 - Las Aulas, tendrán ciertos atributos, y los alumnos también.
1. Agrupamos los alumnos en bloques y asignamos bloques enteros (**restricciones**)
 2. Preferiremos minimizar número de aulas (**métrica de evaluación**)
 3. Es mejor asignar pronto los bloques de mayor número (**heurística**)

En RyP implementamos: Operador

- **Asignar**(grupo_alumnos, clase)

R1: grupo no asignado

R2: clase tiene espacio

R3: clase es del nivel adecuado al grupo

Resultado: Anota id clase en grupo, calcula capacidad clase.

Con esto, vamos asignando hasta no poder más:

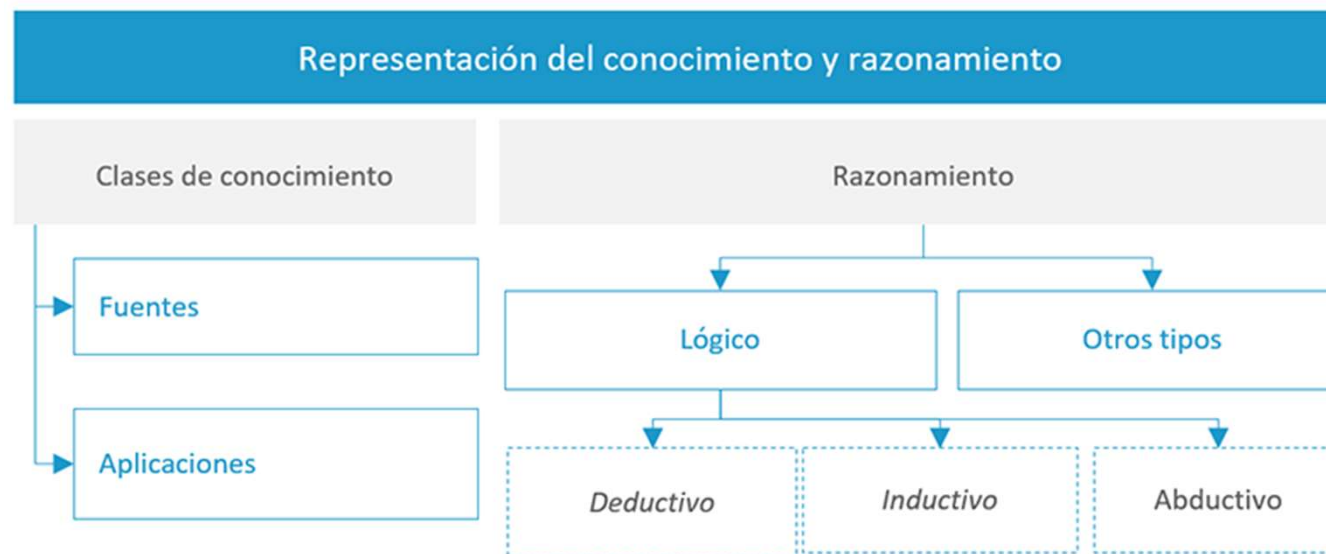
búsqueda.

Otros operadores posibles:
Desasignar, intercambiar,
Abrir_clase, Cerrar_clase

Índice de la clase

Tema 2

- ▶ Razonamiento y tipos de razonamiento
- ▶ Representación y tipos de representación
- ▶ Ejemplos de representación para algunos problemas





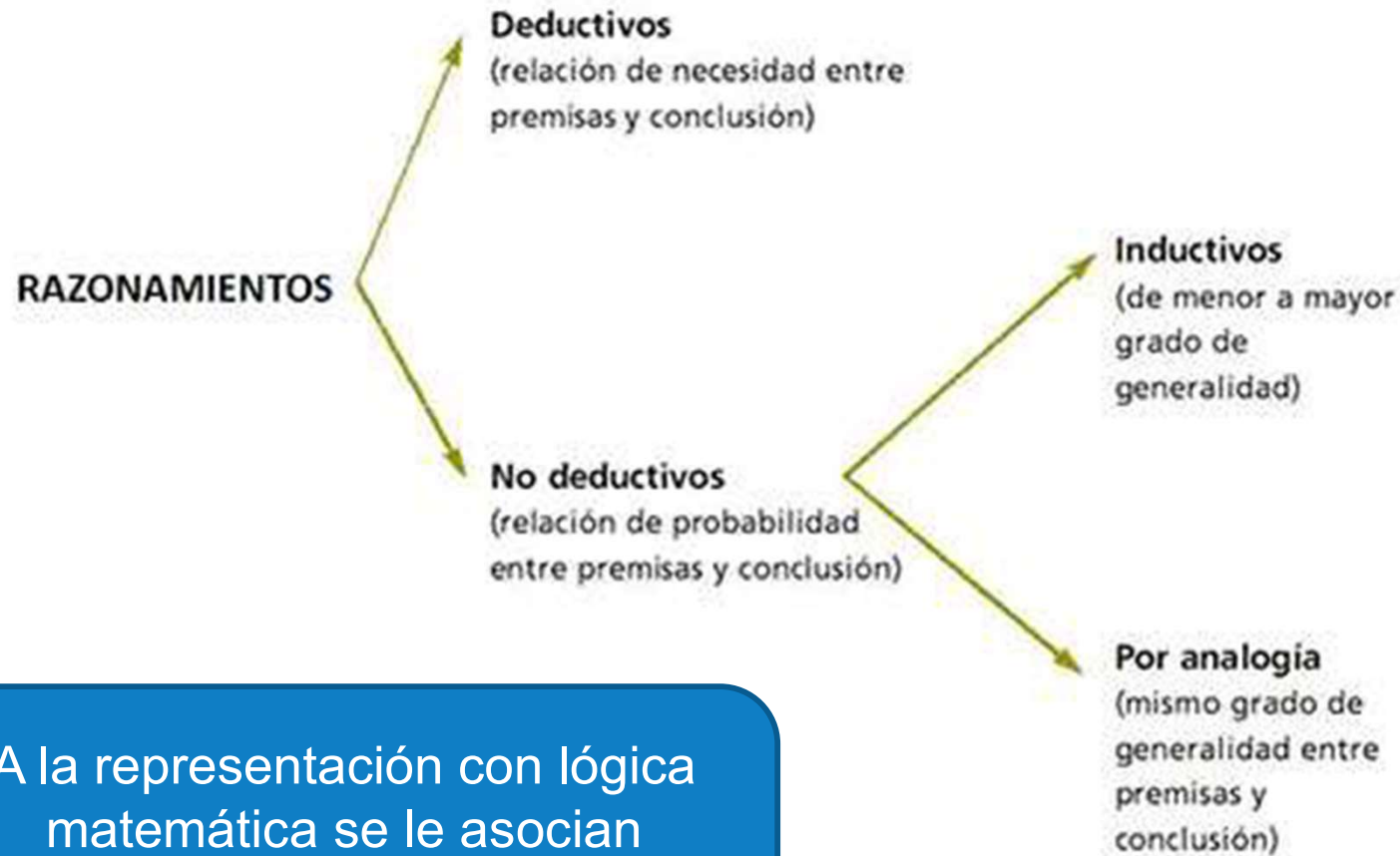
Razonamiento

Problema para la sesión

¿Pueden los modelos de IIA basada en grandes modelos de lenguaje como ChatGPT (y otros) realizar razonamientos? ¿De qué tipo?

Un ejemplo: *Todos los cuervos vuelan, y todos los pájaros vuelan; en consecuencia, ¿podemos decir que el cuervo es un pájaro?*

Tipos de razonamiento



A la representación con lógica matemática se le asocian mecanismos de razonamiento deductivo

Problema para la sesión (cont)



Todos los cuervos vuelan, y todos los pájaros vuelan; en consecuencia, ¿podemos decir que el cuervo es un pájaro?

- ☐ Se nos está preguntando si este razonamiento es válido. Esto depende de qué tipo de razonamiento sea: cada tipo de razonamiento tiene su propio criterio de validación.

RESPUESTA 1

ChatGPT (2/2024): Por lo tanto, podemos concluir que los cuervos son una subclase de los pájaros, pero no todos los pájaros son cuervos. En resumen, mientras que todos los cuervos son pájaros, no todos los pájaros son cuervos. (ver completa)

Dice que no es válido, pero no por las razones correctas. En este caso premisas y conclusión son verdaderas, pero la forma de razonamiento NO es válida. ¿Podemos poner un caso formalmente idéntico en el que no se cumpla esa condición de validez?

Problema para la sesión (cont)



RESPUESTA 2

BING (11/2024): Interesante pregunta. Apliquemos un poco de lógica.

- **Premisa 1:** Todos los cuervos vuelan.
- **Premisa 2:** Todos los pájaros vuelan.

La conclusión que se propone es que, si ambos vuelan, entonces los cuervos son pájaros. Pero, ¿realmente la conclusión se desprende de estas premisas?

En este caso, no necesariamente. Ambas premisas comparten una característica común (volar), pero esto no es suficiente para afirmar que los cuervos son pájaros, aunque en la realidad sí lo sean. Por lo tanto, faltaría una premisa adicional:

- **Premisa 3:** Todos los cuervos son pájaros.

Esta tercera premisa nos permite concluir que, efectivamente, los cuervos son pájaros. Sin esta información, solo sabemos que ambos vuelan, pero no tenemos suficiente para clasificar taxonómicamente a los cuervos.

Si necesitas más aclaraciones, estoy aquí para ayudarte



Problema para la sesión (cont)

¿Puedes decirme qué regla de la lógica de predicados puede extraer la conclusión 3 a partir de las premisas 1 y 2 anteriores?

RESPUESTA 3



BING (11/2024): Para extraer la conclusión 3 a partir de las premisas 1 y 2, podemos utilizar la regla del silogismo hipotético en lógica de predicados. Esta regla establece que si tenemos dos implicaciones del tipo $(A \rightarrow B)$ y $(B \rightarrow C)$, podemos concluir $(A \rightarrow C)$

...



En resumen, con las premisas dadas, no podemos directamente concluir que “todo cuervo es un pájaro” sin una premisa adicional que conecte el hecho de volar con ser un pájaro. La regla del silogismo hipotético sería la adecuada si tuviéramos esa premisa adicional.

Razonamiento lógico deductivo

- **Deductivo**: Sobre la representación lógica se define el razonamiento deductivo



Premisa(regla): Si X es un ave, X vuela
Premisa (hecho): Tito es un gorrión
Premisa (regla): Si X es un gorrión, X es un ave
Conclusión (hecho): Tito vuela

Contenido o significado: La lógica representa un hecho del mundo (o una creencia). A este significado la podremos asignar el **valor de verdad** que se cumpla en la situación representada (valor de Verdadero o Falso).

Forma: en lógica la **forma** determina si un **razonamiento es válido o inválido**:

Ejemplo: "Si A implica B, y no se da B, entonces seguro que no se da A"
(*Modus Tollens, o Tollendo Tollens*)

Los razonamientos son válidos o inválidos **independientemente del significado de sus elementos**.

Dedución: si el valor de todas las premisas es Verdadero, sabemos con certeza que el de las **conclusiones válidas** es también Verdadero

Problema para la sesión (cont)

Extrae una conclusión válida razonando sobre el siguiente caso: el alumno 1 aprueba, el alumno 2 aprueba, el alumno 3 aprueba.

- ❑ Se nos está preguntando cuál es la respuesta a un razonamiento de tipo inductivo.

RESPUESTA 2

ChatGPT (2/24): Dado que todos los alumnos (1, 2 y 3) aprueban, podemos extraer la conclusión de que "todos los alumnos aprueban". Esto se puede expresar formalmente de la siguiente manera:

$\forall x (\text{Alumno}(x) \rightarrow \text{Aprueba}(x))$

Esto significa que para cualquier valor de x (que representaría a cualquier alumno), si x es un alumno, entonces x aprueba.

En inducción, no tenemos una regla de validez, sino una regla que invalida: el razonamiento no es válido si existe algún contraejemplo.

¿Podríamos pensar en un contraejemplo? **El razonamiento no es válido porque puede existir un contraejemplo.**

Razonamiento lógico inductivo

- **Inductivo**: Las observaciones particulares se pueden **generalizar** a veces en leyes. Suele entenderse que se usa un criterio probabilístico.
- **Contraejemplo**: la existencia de un solo contraejemplo invalida la hipótesis



Falsabilidad (Popper, 1934): las hipótesis falsificables (sobre las cuales podemos diseñar un experimento que pueda derivar en un contraejemplo) son preferibles

Problema para la sesión (cont)

Extrae una conclusión válida razonando sobre el siguiente caso: Pedro aprueba RyP, José suspende RyP, María suspende RyP, Paloma aprueba RyP

RESPUESTA 3

ChatGPT: Dado el caso presentado, podemos extraer la conclusión de que no todos los estudiantes aprueban la asignatura "RyP" (Razonamiento y Programación). Esto se puede expresar formalmente de la siguiente manera:

Aquí la respuesta es correcta desde el punto de vista deductivo; pero se podría extraer una hipótesis que explique los aprobados. Hay múltiples hipótesis que pueden explicar esas observaciones, podemos pensar en una: si el nombre de X empieza por P, X aprueba RyP.

No podemos garantizar que nuestra hipótesis (absurda) sea válida en general, pero podemos admitirla provisionalmente (con poco fundamento).

Razonamiento lógico abductivo

- Hay cierto debate sobre si es un tipo de inductivo
- **Hipótesis razonada**: ante unos hechos, se busca la hipótesis que mejor los explica

Pedro aprueba RyP
José suspende RyP
María suspende RyP
Paloma aprueba RyP



Conjetura (regla generalizada): si el nombre de X empieza por P, X aprueba RyP

Conjunto de observaciones



Conjeturas



Mejor hipótesis (¿criterio?)

Observación: aparece un sandwich medio comido en la cocina.

Hipótesis: tu hija lo preparó esta mañana, iba tarde al trabajo y se lo dejó al salir corriendo.

Criterio de evaluación: ¿navaja de Occam?

Ver: <https://dyingwords.net/using-occams-razor-without-cutting-your-own-throat/>