

Desarrollando una ontología sencilla

**Curso de Doctorado:
Sistemas Multiagente**

Dpt. Informática – Curso 2002-03

Agenda

- ❑ Porqué desarrollar una ontología
- ❑ Qué es una ontología
- ❑ Una metodología de KE sencilla para ontologías
- ❑ Pasos para la creación de una ontología
- ❑ Elementos a tener en cuenta

Porqué desarrollar una ontología

□ Gran interés en

- el WEB
 - Para categorizar lugares web
 - W3C desarrolla el RDF
 - W3C y ARPA desarrollan DAML
- Recuperación de información
- Librerías digitales
- Comercio electrónico
- NLP
 - Medicina: SNOWMED, UMLS.
 - UNSPSC para terminología de productos y servicios (www.eccma.org/unspsc/).

Qué permite una ontología

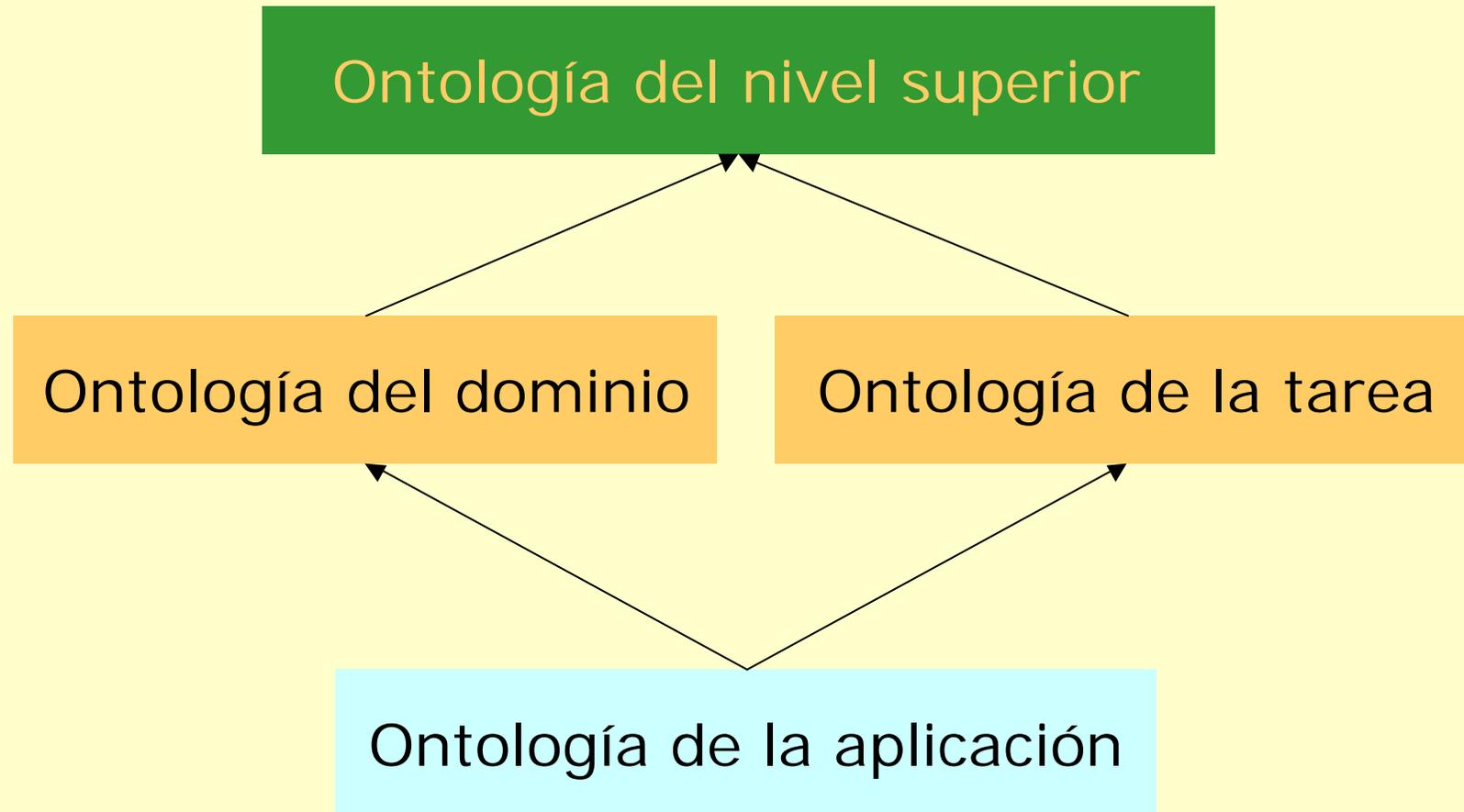
- ❑ Compartir conocimiento común sobre la estructura de las cosas
- ❑ Permite reusar el conocimiento del dominio
- ❑ Explicitar suposiciones sobre el dominio
- ❑ Separar el conocimiento del dominio del conocimiento operacional
- ❑ Posibilitar el análisis del conocimiento del dominio



Qué es una ontología

- Descripción formal explícita de ciertas entidades de un dominio de discurso:
 - **Clases** – conceptos del dominio
 - **Slots** – propiedades de cada concepto describiendo las características y atributos (también llamados roles o propiedades)
 - **Facets** – restricciones sobre los Slots (también llamadas restricciones de roles)
- Base de conocimiento: ontología + ejemplares de las clases.

Tipos de Ontologías



Clases

- Describen conceptos del dominio
- Una subclase representa conceptos más específicos que su superclase
- Ej:
 - Vino:
 - Tinto, Rosado, Blanco
 - Vino:
 - Con burbujas, Sin burbujas

Slots

- ❑ Describe propiedades de las clases e instancias
- ❑ Ej:
 - Un “Protos Gran Reserva” tiene un color cereza intenso;
 - se produce en la bodega “Protos”.
 - Este vino se describe con dos slots:
 - bodega: Protos
 - olor: cereza intenso
- ❑ A nivel de clase podemos incluir otros slots más, como: temperatura, cepas, ...



El desarrollo de una ontología incluye...

- ❑ Definir clases en la ontología
- ❑ Encontrar una jerarquía taxonómica de clases
- ❑ Definir los slots y sus valores permitidos.
- ❑ Llenar los valores de los slots en las instancias.

Una metodología de KE sencilla

- Puede emplearse una aproximación iterativa:
 - Proponer una ontología sencilla
 - Refinarla en épocas sucesivas aplicando criterios de modelado

Reglas para la metodología

- ❑ No hay una única forma de modelar un dominio.
 - La mejor solución suele depender de la aplicación y las extensiones previstas
- ❑ El desarrollo de la ontología es necesariamente un proceso iterativo
- ❑ Los conceptos ontológicos deben ser cercanos a los objetos y a las relaciones del dominio de interés.

Hay que fijarse en el lenguaje común

- Nombres (objetos)
- Verbos (relaciones)

Criterios de diseño

- Intuitiva
- Extensible
- Mantenable
- Debe ser un modelo de la realidad

Pasos para la construcción de la ontología

1. Dominio y alcance de la ontología
2. Utilización de ontologías existentes
3. Enumerar los conceptos importantes
4. Definir los conceptos y su jerarquía
5. Definir las propiedades de las clases (slots)
6. Definir los facets de los slots
7. Crear las instancias



1.- Dominio y alcance de la ontología

□ Cuestiones básicas

- Cuál es el dominio que cubre la ontología
- Para qué se usará la ontología
- Para qué tipo de cuestiones
- Quién usará o mantendrá la ontología

1.- Ejemplo Vinos

- ❑ Dominio: combinaciones de vinos y comida.
- ❑ Conceptos:
 - Tipos de vinos
 - Tipos de comidas
 - Combinación vino-comida
 - No es nuestro objetivo incluir inventarios de vinos.
 - Si se empleara en una revista de vinos habrá que incluir sinónimos y frases hechas.
 - Si se empleara para aconsejar compras habrá que incluir precios y lugares de compra.

1.- Dominio y alcance de la ontología (ii)

□ Cuestionario de *competencia*.

Una forma de determinar el alcance de la ontología es bosquejar un cuestionario de competencia, que deberá ser de responder (test):

- La ontología deberá permitir responder al cuestionario.
- Permite saber si es preciso mejorar el nivel de detalle de cierta parte de la ontología.

1.- Ejemplo Vinos (Competencia)

- ❑ ¿Qué características del vino debo tener en cuenta al elegir un vino?
- ❑ ¿Es el Burdeos un vino tinto o blanco?
- ❑ ¿Va bien el Cabernet Sauvignon con el pescado?
- ❑ ¿Cuál es la mejor elección para la carne a la brasa?
- ❑ ¿Qué características del vino afectan a su adecuación a cada plato?
- ❑ ¿Cambian sus características en boca con la añada?
- ❑ ¿Cuáles fueron las mejores añadas para el Ribera de Duero?



2.- Plantéese reutilizar una ontología existente

- ❑ Facilita la integración con otros sistemas.
- ❑ Es posible importar y exportar de diversos formalismos (automáticamente o a mano)
- ❑ Hay ontologías públicas: Ontolíngua, DAML, UNSPSC, RosettaNet, DMOZ, CyC...
- ❑ Ejemplo: Existen listas de propiedades de vinos (www.wines.com) que se usan.



3.- Enumere términos importantes

- ❑ Escriba la lista de términos que se usarían para construir sentencias sobre el tema.
 - Qué términos nos gustaría usar
 - Qué propiedades tienen estos términos.
 - Qué podemos decir sobre estos términos.
- ❑ Ejemplo:
 - Vino, uva, bodega, ubicación, color, cuerpo...
 - Pescado, carne roja,
- ❑ No debe importar el solapamiento de términos, si son relaciones o términos o si son clases o slots.



4.- Defina las clases y la jerarquía

□ Aproximaciones:

- **Top-down:** Primero conceptos generales, después las especializaciones.
 - Ejemplo: Primero "vino", "comida". Después "vino tinto", "vino rosado", ... Después "Burdeos", "Cabernet Sauvignon", ...
- **Bottom-up:** Primero las clases más específicas. Las más generales se construyen por agrupamiento.
- **Híbrida:** de las dos anteriores.

4.- Defina las clases y la jerarquía (ii)

- ❑ Ninguna de ellas es necesariamente mejor
- ❑ Habitualmente los conceptos de nivel intermedio son los más manejables:
 - No son demasiado abstractos
 - Presentan alguna estructura. No así los de bajo nivel.
- ❑ Depende de la experiencia del diseñador.
- ❑ De la lista de términos se seleccionan aquellos que describen objetos con existencia independiente.

4.- Jerarquía

Si una clase A es una superclase de la clase B, cualquier ejemplar de B lo es también de A.

- La clase B representa un concepto ISAKO A. (una especie de A)



5.- Definir las propiedades de las clases – slots

- La mayoría de “términos” restantes alude a propiedades de estas clases
 - Ejemplo: el *color*, *sabor*, *cuero* de un vino.
La *ubicación* de una bodega.
- Debemos determinar a que clase describe cada término
 - Estas propiedades reflejan **slots** ligados a las clases.
 - Ojo, porque esta es la aproximación de Protege.

5.- Definir las propiedades de las clases–slots (ii)

□ Tipos de slots

- Intrínsecos
 - Ej: sabor, color del vino,...
- Extrínsecos
 - Ej: el nombre del vino, el área,...
- Partes
 - Ej: Los diferentes ingredientes de una comida,...
- Relaciones entre instancias de la clase y otros elementos.
 - Ej: embotellador, la uva,...

5.- Definir las propiedades de las clases-slots (ii)

- ❑ Las subclases heredan los slots.
- ❑ Se puede aumentar su número.
 - Ej: en el caso del vino tinto: nivel de tanino.
- 👉 El slot debe agregarse en el nivel más externo posible.



6.- Definir los facets de cada slot

- Los facets describen el
 - **tipo** del valor (string, número, booleano, enumerado, instancia de otra clase)
 - Ej: produce: es de tipo instancia de vino
 - La **cardinalidad** del facet,
 - Simple, múltiple (acotado, libre)
 - Ej: bodega tiene cardinalidad múltiple
 - **rango** del slot: clases permitidas para los slots de tipo instancia
 - Ej: el rango de produce es vino

6.- Definir los facets de cada slot (ii)

- **dominio** del slot: clases descritas por el slot
 - Ej: el dominio de produce es "bodega"
- 👉 El dominio|rango deberá ser lo más general posible.
 - 👉 Pero sin pasarse: todas las clases del dominio de un slot se describirán con él y las instancias son potenciales "usuarios" del slot.
- 👉 Si un rango|dominio incluye una clase y sus subclases, elimine las subclases.
 - Ej: el rango de produce es "vino", no "vino tinto"

6.- Definir los facets de cada slot (iii)

- ☞ Si un rango|dominio contiene todas las subclases de una superclase, póngase solo la superclase.
 - Ej: el dominio rango de produce es "vino", no "vino tinto", "rosado", ...
- ☞ Si un rango|dominio incluye casi todas las subclases de una superclase, considere afectar a toda la superclase.



7.- Cree las instancias

- Proceso:
 1. Elegir la clase
 2. Crear una instancia de la clase
 3. Completar los slots



Elementos a tener en cuenta

- ❑ Definiendo clases y una jerarquía de clases
- ❑ Definiendo las propiedades
- ❑ Convenciones de nombre



Definiendo clases y una jerarquía de clases

- La jerarquía se articula sobre la relación “is-a”
 - Una subclase de una clase representa un concepto que es “cierto tipo de” el concepto que representa la superclase
 - Ej: un solo vino no es una subclase de todos los vinos
- “Vino no es una subclase de Vinos”
 - Violaría la regla anterior

Definiendo clases y una jerarquía de clases

- ❑ La relación jerárquica es transitiva
- ❑ La jerarquía está sujeta a evolución de nuevas clases.
 - Ej: Cigales ha sido siempre un vino rosado, pero ahora hay nuevos vinos
- ❑ Las clases representan conceptos del dominio y no las palabras que denotan los conceptos
 - No utilizar sinónimos para clases diferentes
 - Algunas herramientas permiten listas de sinónimos
- ❑ Evítense los ciclos de clases

Analizando los hermanos

- ❑ Los hermanos son subclases directas de la misma clase
 - Todos los hermanos deben estar al mismo nivel de *generalidad*, salvo los raíz.
 - Ej: "Vino blanco" y "Chardonay" no deben ser hermanos.

- ❑ Cuántos?
 - La mayoría de las buenas ontologías suelen tener entre dos y una docena de hermanos.

Analizando los hermanos

□ Cuántos (ii)

- No debe haber “hijos únicos”
- Si hay más de una docena de hermanos puede que haya que pensar en categorías adicionales.

Herencia múltiple

- ❑ La mayoría de sistemas de representación permiten la herencia múltiple
- ❑ Los slots se heredan.

Cuando introducir nuevas clases (o no)

- ❑ Hay que llegar a un compromiso entre anidamiento y número de hermanos
- ❑ Reglas básicas.
Las subclases de una clase generalmente:
 - Tienen propiedades adicionales que la superclase no posee, o
 - Tienen restricciones diferentes, o
 - Participa en relaciones diferentes.
- ❑ En las jerarquías terminológicas las clases no introducen nuevas propiedades
 - Permiten organizar términos y hacerlos más específicos.

Cuando introducir nuevas clases (o no)

- ❑ A veces se incluyen nuevas clases para reflejar distinciones que efectúan los expertos aunque no se desee modelar tal distinción.
- ❑ No hay por qué crear subclases para cada restricción adicional.
 - Ej: No es preciso introducir clases para hablar de vinos delicados, o populares,

¿Una nueva clase o un valor de una propiedad?

- Hay que fijarse en la relevancia de la distinción.
 - Ej: ¿Vino blanco y vino tinto, o una propiedad para el color?
 - Para un catador es necesario.
 - Para casar comidas y vinos es necesario.
 - Para un impresor de etiquetas de botellas, no.
 - Si los conceptos con diferentes valores de slots se convierten en restricciones para diferentes slots de otras clases entonces crearemos una nueva clase.

¿Una nueva clase o un valor de una propiedad?

- Si una distinción es importante en el dominio y pensamos en los objetos con valores diferentes como diferentes tipos de objetos, debemos crear una clase.
 - Es beneficioso pensar en las potenciales instancias.
- Usualmente la distinción entre clases no obedece a propiedades extrínsecas.

¿Una instancia o una clase?

- Depende del uso potencial de la ontología: fija el nivel de granularidad.
 - Los conceptos más específicos deberán ser instancias individuales de la base de conocimiento.
 - Ej: Para un bodeguero, las instancias son las existencias concretas. A un catador no le interesa cada botella concreta.

Límite de la ontología

- ❑ La ontología está completa en el momento en que se conoce todo lo que se necesita:
 - La ontología no debe contener toda la posible información del dominio. 😊
- ❑ La ontología no debe contener todas las propiedades y distinciones posibles entre clases.

Subclases disjuntas

- Algunos sistemas nos permiten especificar que diversas clases son disjuntas (no tienen instancias en común).
 - Esto permite validar mejor la ontología.



Definiendo propiedades

❑ Slots inversos

- Un valor puede depender de un valor de otro slot, y es bueno que aparezcan en ambas cada clase donde aparece el slot.
 - Ej: "vino.embotellador" y "bodega.produce"

❑ Valores por defecto

- Es posible incluir valores por defecto para dar cuenta de la mayoría:
 - Ej: "vinos.de postre" suelen ser dulces.



Convenciones de nombre

- ❑ Es preciso un convenio de nombres para clases y slots.
- ❑ Factores que afectan
 - Hay un espacio de nombres único.
 - Es sensible a mayúsculas/minúsculas
 - Se permiten delimitadores

Convenciones de nombres

- ❑ Reglas habituales
 - Utilizar espacios para separar palabras
 - Separar palabras con mayúsculas
 - Separar con guión bajo.
- ❑ ¿Singular o plural?
 - No hay regla, pero hay que establecer un acuerdo.
- ❑ Prefijos y sufijos: Hacer legible el texto
 - Usualmente: tiene-*, *-de
 - Ej: Tiene-productor, productor-de
- ❑ Convenio entre infinitivos y verbos conjugados.

Convenciones de nombres

□ Otras consideraciones:

- No añadir cadenas como "clase", "propiedad", "slot" y demás (tampoco en otro idioma)
- Conviene evitar abreviaturas (son innecesarias).
- Los nombres de las subclases directas deberán tener (o no) el nombre de la superclase (fijar convenio).
 - Ej: Blanco y Tinto, Vino Blanco y Vino Tinto, pero no mezclas.



Recursos soft.

- ❑ Protege2000
- ❑ OIL
- ❑ Chimaera
- ❑ Ontolingua
- ❑ ...