

# Adquisición de Conocimiento

Juan Luis Castro Peña

# Preguntas a responder del Tema

- ¿En que consiste la adquisición de conocimiento?
- ¿Cuándo se realiza este proceso?
- ¿De qué fuentes se suele adquirir el conocimiento y en qué medida?
- ¿En qué consiste el proceso de adquisición del conocimiento?
- ¿Cómo se realiza la fase del estudio del problema? ¿Qué debemos de obtener de esta fase?
- ¿Qué es la extracción de conocimiento? ¿Qué técnicas podemos usar para realizarla? ¿Qué debemos obtener de esta fase?
- ¿Qué es la educación del conocimiento? ¿Qué técnicas podemos usar ? ¿En que fases se divide esta parte del proceso? ¿Qué debemos obtener de cada fase?

# Introducción

- Adquisición de Conocimientos (AC): proceso de recolección de información, a partir de cualquier fuente (experto, libros, revistas, informes, ...), necesaria para construir un Sistema Basado en Conocimiento
- Objetivo: proporcionar a cada etapa la información que se requiere en cada momento del desarrollo.
- Cuándo: no es un paso concreto en la metodología de desarrollo de un SBC, sino más bien una tarea que se produce en paralelo a todas las etapas de construcción de estos sistemas (identificación, conceptualización, formalización, validación, mantenimiento, ...).
- Relación con el resto de las fases del desarrollo de un SBC: No es un proceso aislado; bien al contrario, forma parte de cada fase

# Situación actual

- la adquisición de conocimientos es la tarea más importante en el desarrollo de SBC
- Algunas técnicas propias de la IA son aplicables, pero como herramientas
- No existe un método completamente automatizado de Adquisición de Conocimiento.
- Existen unas metodologías utilizadas para adquirir correctamente el conocimiento

# Resumen general

Expertos

Educción

- Entrevistas
- Análisis de incidentes críticos
- Cuestionarios

Documentos

Extracción

- Análisis estructural de textos
- Herramientas de IA para análisis de textos

Datos

Técnicas (semi-)automatizadas

- Rejilla de repertorio
- Aprendizaje de arboles de decisión
- Aprendizaje de reglas

# Fuentes de conocimiento: Personas

- Expertos, directivos y usuarios finales
  - Expertos: mayor parte de los conocimientos
  - Directivos: los objetivos del proyecto, el alcance del sistema, el contexto donde será instalado, ...
  - Usuarios: comprender el tipo de persona que interactuará con el sistema, sus necesidades, requisitos, ....

# Fuentes de conocimiento:

## Documentos

- Resultados de estudios e investigaciones: datos, estudios, informes, resultados estadísticos, etc.
- Publicaciones especializadas: suelen contener las versiones mas actualizadas
- Material utilizado para formación: suele contener conocimientos expuestos de un modo especialmente claro
- Casos: Algunas empresas suelen registrar los casos que se les presentan en forma de órdenes de reparación, fichas de clientes o pacientes, estudios o almacenamiento de casos.

# Fuentes de conocimiento: Experiencia

- Observación (Visitas): útil para clarificar las ideas y entender el esquema general del proceso
- Documentos informales:
  - notas manuscritas, ayudas de trabajo, etc., que circulan dentro de las organizaciones.
  - Proporciona, a menudo, conocimiento heurísticos para resolución de problemas.
  - Suelen reflejar la experiencia de profesionales de la organización a la hora de enfrentarse a ciertos problemas.
  - Proporciona conocimiento semipúblico.



# Extracción y educción

- Extracción de conocimiento: cuando la fuente de conocimiento se presenta en forma escrita o registrada.
- Educción de conocimiento: cuando los conocimientos se obtienen de humanos

# Proceso de adquisición de conocimiento

- Estudio del problema: primeras reuniones y evaluación de viabilidad (de expertos, directivos y usuarios)
- Extracción de conocimientos (de la documentación)
- Educción de conocimientos (del experto)
  - Interrogatorio inicial
  - Investigación profunda

# Estudio del problema

- Determinar los requisitos funcionales del sistema, o, en su caso, las necesidades de los usuarios del futuro sistema, o lo que los usuarios esperan del mismo.
  - Entradas, salidas e interacción con el sistema
- Introducir al IC en el dominio a un nivel tal que sea capaz de desarrollar un estudio de viabilidad del SE donde se determine si la tarea del experto es tratable, o no, mediante la Ingeniería del Conocimiento.
  - Esquema general del proceso

# Estudio del problema: primeras reuniones

- Se buscan conocimientos generales, no de detalle, así como familiarizarse con la terminología del dominio. El ingeniero debe centrarse en el entorno de la tarea y sus usuarios.
- Hay que dirigir la entrevista a definir el problema: entradas, salidas e interacciones del sistema
- Posteriormente hay que conseguir el esquema general del proceso de razonamiento en forma de tareas, subtareas y las relaciones y dependencias entre ellas.
- Hay que analizar la viabilidad de obtener los hechos en que se basa el razonamiento
- Hay que evitar a toda costa que nuestras opiniones e ideas interfieran en los resultados (error muy común que ralentiza esta fase)
- Una o dos reuniones debe ser suficiente

# Estudio del problema: análisis de viabilidad

- ¿Es factible obtener los datos que se usan de entrada?  
¿Cuánto coste tendría?
  - > Posible modificación: de calculados (demasiado complejo o costoso) a introducidos por usuario
- ¿El esquema general del proceso de razonamiento se puede representar?
  - > Solución mas común: sistema modular de tareas
- ¿Cada tarea se podría representar como un SBC simple mediante un modelo de representación del conocimiento básico?
  - > Habría que reiterar el proceso para cada tarea que no resulte básica

# Extracción de conocimiento

- Aprender sobre el dominio tanto como sea posible antes de comenzar las sesiones con el experto.
- Se pretende evitar o, por lo menos, reducir el tiempo que, de otro modo, debería dedicar el experto a fin de iniciar al ingeniero en el tema.

# Extracción de conocimiento: consulta con el experto

- Indique los documentos mas importantes a consultar
- Indique el material relevante de la colección de manuales
- Explique la terminología usada y los conceptos discutidos en la documentación.
- Proporcione detalles omitidos en los documentos, pero cruciales para un entendimiento global de la tarea.
- Señale los puntos donde la práctica real difiere de los procedimientos documentales..
- Explique las anotaciones a mano hechas sobre los documentos de trabajo.

# Extracción de conocimiento: Análisis estructural de textos

- Técnicas en las cuales los términos son determinados por el IC **en tiempo de ejecución**.
  - Ante un determinado texto, establece una serie de términos que deben ser buscados en el texto.
- Los términos a buscar están **preestablecidos** por la técnica y son **dependientes del dominio**.  
(Enfermedad, Medicina, Terapia, Tratamiento, ...)
- Los términos a buscar están también **preestablecidos**, pero son **independientes del dominio**. (Se define como, Está relacionado con, Es una característica de, etc.)



# Análisis estructural de textos: términos independiente del dominio

- Independientes del dominio → Conceptos y relaciones
- Estructuras tipo “definición” para los conceptos
- Estructuras tipo “afirmación relacional” para las relaciones
- Necesitamos:
  1. Estructuras a buscar y tipo de conocimiento que se obtendrá con cada una (concepto, relación, característica, valor)
  2. Modo de detectar esas estructuras

# Tipo de estructuras a buscar

- **Definiciones:** Introducción de un concepto nuevo en el texto. El criterio puede venir definido en base a distintos criterios (uso, partes que lo componen, ...)
- **Afirmaciones:** Una afirmación es una frase que establece una verdad. Para el objetivo de extraer conocimientos básicos, las afirmaciones que interesan son aquellas que expresan relaciones entre conceptos.
- **Leyes:** Las leyes de un dominio establecen sus principios básicos, así como las reglas que fijan el funcionamiento de objetos del dominio.
- **Procedimientos:** Los procedimientos de un dominio establecen los pasos para la resolución de problemas en el dominio. Al igual que en el caso anterior, los conocimientos proporcionados por esta estructura están más allá del objetivo de la extracción de conocimientos a partir de la documentación.

# Modo de detectar las estructuras

- Búsqueda de patrones:
- Algunos ejemplos:
  - Definiciones: A se usa para B; A es un B que C; A esta compuesto de B, C,...
  - Afirmaciones: A es causa de B; A se relaciona con B; A es la finalidad de B; etc
- Títulos de secciones, subsecciones, y remarcados (negrita, cursiva)

# Detectando estructuras

- El trabajo consistirá en una tarea tan simple, pero tediosa, como:
  - Hojear el texto deteniéndose sólo en las definiciones y afirmaciones de relación. Estas frases serán señaladas.
  - Analizar las frases señaladas para extraer los conocimientos buscados: conceptos, relaciones y definiciones de conceptos.

# Una ejemplo de herramienta que puede usarse: word2vec

- Asigna a cada palabra un vector numérico (un punto en el espacio)
- Codifica espacialmente el significado de las palabras y las relaciones entre las palabras
- Aprende a partir de textos
- Desarrollado por Mikolov, Sutskever, Chen, Corrado y Dean en 2013 en el departamento de investigación de Google
- Desarrollada en C y puesta para uso público ([Apache License 2.0](#)). Hay versiones para Python (gensim) y R.

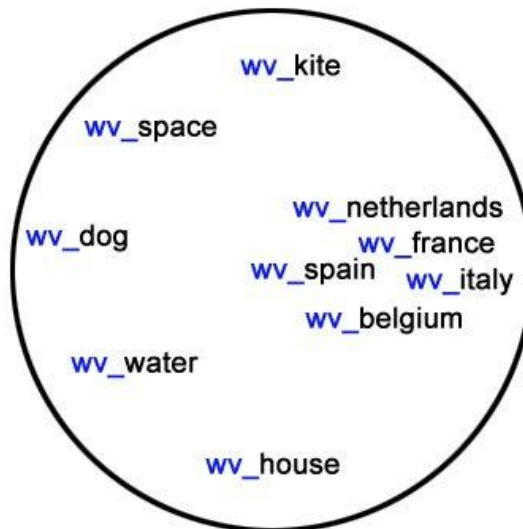
# word2vec

Input:  
text

Lorem ipsum dolor  
sit amet, consete-  
tur sadipscing elitr,  
sed diam nonumy  
eirmod tempor  
invidunt ut labore  
et dolore magna  
aliquyam erat, sed  
diam voluptua. At  
vero eos et

  
train for  
each word  
a word vector

Model:



vector space:  
consists of **word vectors**  
for each word

most\_similar('france'):

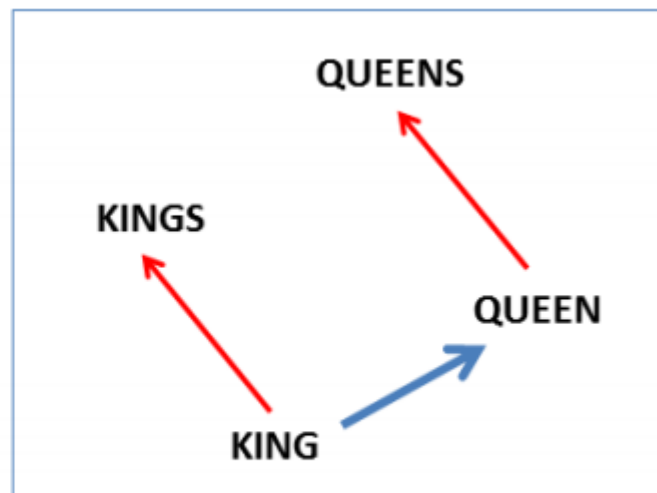
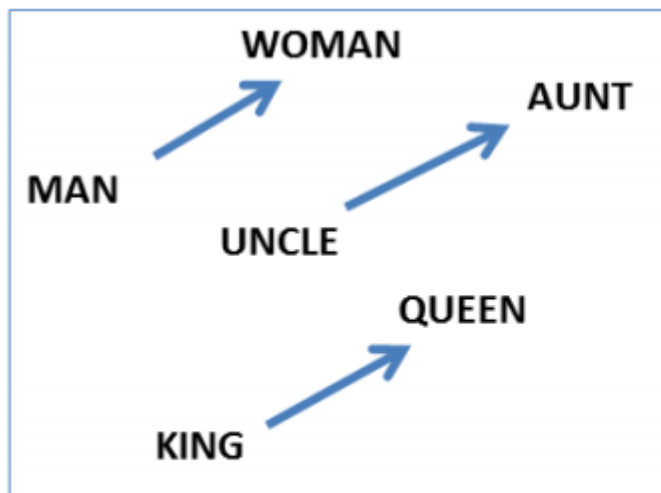
spain	0.678515
belgium	0.665923
netherlands	0.652428
italy	0.633130

highest cosine  
distance values  
in vector space  
of the nearest  
words

# Características de word2vec

- Las palabras cercanas tienen un significado similar:  
Vector(Open\_Source\_Software) cercano Vector(Free\_Software)
- Los vectores entre dos palabras representan la relación entre esas dos palabras

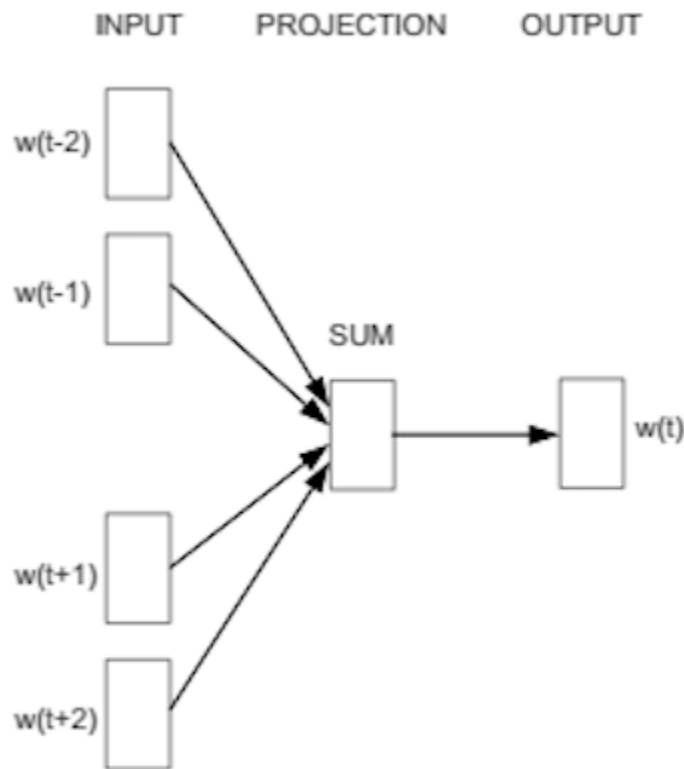
$$\text{Vector(hombre)} - \text{Vector(mujer)} = \text{Vector(rey)} - \text{Vector(reina)}$$



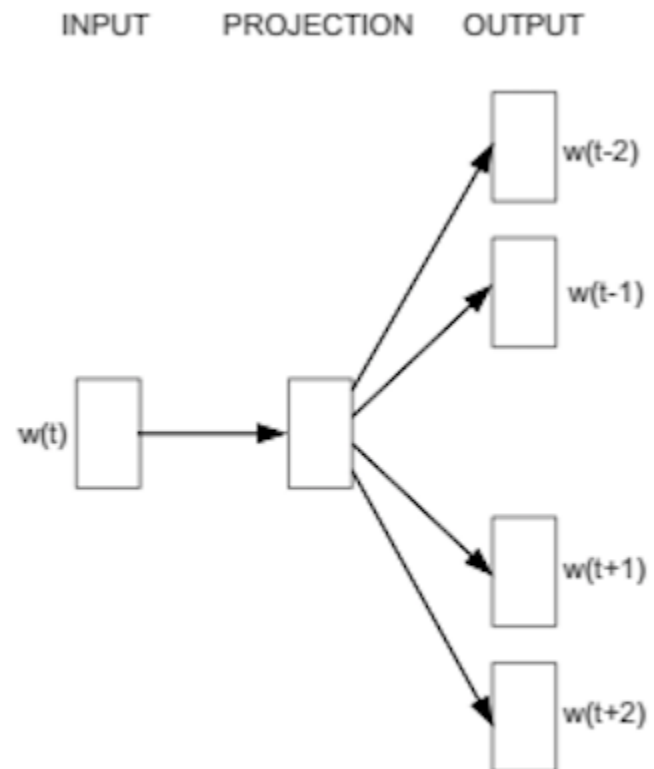
(Mikolov et al., NAACL HLT, 2013)

# Cómo se asigna el vector

- Entrena una red neuronal para predecir el contexto de cada palabra



**CBOW**



**Skip-gram**



# Educción de conocimiento

- IC deberá marcar la perspectiva y profundidad deseada en cada sesión.
- Dada la complejidad intrínseca del proceso, cada sesión debe ser controlada, además, en muchos otros aspectos.

# Ciclo de educación

- Preparación de la sesión
- Sesión
- Transcripción
- Análisis de la sesión
- Evaluación

# Preparación de las sesiones

- Información a tratar
- Amplitud o profundidad de cada una
- Técnica adecuada
- Preparación de las preguntas

# Desarrollo de las sesiones

- Repaso del análisis de la última sesión
- Explicación al experto de los objetivos de la nueva sesión
- Educción
- Resumen y comentarios del experto

# Análisis de la sesión

- Lectura para obtención de una visión general
- Extracción de conocimientos concretos: subdomio, tareas, problemas, conceptos, relaciones
- Lectura para recuperar detalles olvidados
- Crítica para mejorar parte del IC: problemas que ha habido, información que falta, preguntas que se deberían haber realizado, .....

# Evaluación de la sesión

- Las preguntas a responder en este momento son, entre otras:
  - ¿Se han conseguido los objetivos?
  - ¿Es necesario volver sobre el mismo objetivo?
  - Número y tipo de sesiones necesarias para cubrir el área
- Esta fase del ciclo de educación se funde con la primera de preparación de la siguiente. Es decir, evaluando si se han alcanzado los objetivos, si se necesitan más sesiones sobre el mismo área, etc, se responde a las preguntas que el IC debe plantearse en la primera fase del ciclo: cuáles son los objetivos de la nueva sesión, área a tratar, etc.

# Técnicas de educción

- Métodos directos: le preguntan directamente al experto lo que sabe. El experto es la única fuente de información; el IC confía totalmente en lo que el experto le dice.

Dime lo que sabes

- Métodos indirectos: se usan porque no siempre los expertos pueden acceder a los detalles de sus conocimientos o procesos mentales, y para confirmar lo adquirido mediante técnicas directas.

Déjame comprobar lo que sabes

# Técnicas de educación

- Entrevistas
  - abiertas
  - Estructuradas
- Cuestionarios
- Observación de tareas habituales
- Incidentes críticos
- Clasificación de conceptos
- Análisis de protocolos
- Emparillado
- Inducción



# Entrevista abierta

- Entrevista abierta:

- IC plantea, más o menos espontáneamente, preguntas al experto
- Espontáneamente no significa que esta técnica no necesite planificación y control.
- Como en cualquier sesión de educación, el IC habrá fijado un tema o perspectiva a tratar con el experto, así como una profundidad de los conocimientos a educir.
- La falta de conocimientos sobre la perspectiva fijada, y el requisito de un grano grueso en el tema, lleva al IC a seleccionar la Entrevista Abierta como la técnica más adecuada a usar en una determinada sesión de educación.

# Preguntas típicas de una entrevista abierta

## Iniciales

- ¿Cómo resuelves este problema?
- ¿Cuáles son los elementos que influyen cuando resuelves el problema?
- ¿Cuáles son las informaciones que necesitas antes de empezar el tratamiento del problema?

## Alentar

- ¿Qué haces a continuación?
- ¿Puedes describir lo que quieres decir con eso?
- ¿Por qué haces eso?

## Alerta

- ¿En qué se parece y diferencia este problema con los típicos del dominio?
- ¿Qué tipos de datos necesita el problema?
- ¿Qué clase de soluciones son adecuadas para el problema?
- ¿Qué constituye una explicación o justificación adecuada de la solución del problema?

# Entrevista estructurada

- IC planifica todas las preguntas que debe plantear al experto durante la sesión.
- Debe formular y agrupar las cuestiones lógicamente. Normalmente, los grupos conciernen a acciones o procesos que se ha identificado en sesiones previas.
- El IC plantea un grupo de cuestiones sobre determinado objeto o atributo y, una vez resueltas, pasa al grupo siguiente.
- Las preguntas a plantear deberían centrarse sobre los conocimientos de: los conceptos, relaciones e inferencias del experto.

# Preguntas típicas de una entrevista estructurada

- ¿Qué tipo de cosas te gustaría saber acerca del problema cuando empiezas a sopesarlo?
  - ¿Qué hechos o hipótesis intentas establecer cuando piensas sobre el problema?
  - Cuáles son los factores que influyen en la forma en que razonas en el problema?
  - ¿Qué tipo de valores puede tener ese objeto?, ¿qué rango de valores está permitido?
  - ¿Este valor depende de otros factores? En caso afirmativo, ¿cuáles son?
  - ¿Es este factor necesario para resolver todos los problemas en el dominio, o sólo para algunos? En este caso, ¿cuáles?
- 
- ¿Podrías explicarme este concepto en mayor detalle?
  - ¿Qué ocurre en este punto?
  - ¿Por qué te planteaste este problema?
  - ¿Es correcta esta secuencia?
  - ¿Es esto lo que haces en esta situación?
  - ¿Te parece que este diagrama muestra correctamente el orden de tus decisiones?
  - ¿Están incluidos aquí todos los conceptos relacionados con tal tema?

# Limitaciones de las entrevistas

- Puntos fuertes:
  - Sirven para compilar conocimientos básicos de la tarea
  - Sirven para obtener la información conceptual implicada en el problema
  - Sirven para extraer conocimientos de relaciones, valores y acciones
- Puntos débiles:
  - Las entrevistas son consumidoras de tiempo, sobre todo del Ingeniero de Conocimiento
  - Confían en la memoria del experto
  - Existen problemas con el lenguaje

# Problemas con el lenguaje

A) *Eliminación de componentes claves* en un proceso de razonamiento

B) *Palabras con referencias no especificadas.*

C) *Referencias hechas con comparativos.*

D) *Condensación, con palabras, de procesos complejos*

E) *Implicación de conexión causal entre eventos*

# Observación de tareas habituales

- Con frecuencia, la mejor forma de descubrir cómo hace un juicio un experto, efectúa un diagnóstico, o diseña una solución, es observar a un experto trabajar en un problema real habitual.
- La primera decisión que se debe tomar al respecto es cómo registrar las prestaciones del experto. Una posibilidad es sencillamente observar, tomar notas e intentar seguir *al vuelo* el proceso de pensamiento del experto. Otra, es grabar todo el proceso para una posterior revisión con el experto.
- El IC no interfiere en la actuación del experto en la solución de sus tareas reales cotidianas.

# Observación de tareas habituales

- Ventajas:
  - Proporciona al IC una primera idea de los tipos de conocimientos y habilidades implicados en el dominio.
  - Proporciona conocimiento básicos del dominio y ayuda a que el IC comprenda la tarea del experto.
  - Es útil para captar conocimientos procedimentales o para entender las características peculiares de los usuarios del SBC.
- Inconvenientes:
  - Si las tareas habituales no son muy informativas acerca del razonamiento del experto, el método suministra poco conocimiento.
  - Consume mucho tiempo.
  - A veces es inoportuna y fastidiosa.



# Otras técnicas

## Técnica de las veinte preguntas:

El IC presenta un caso inventado al experto. Para la resolución del problema, el experto sólo podrá plantear al IC, veinte preguntas sobre el mismo. De nuevo la técnica sirve para conocer la información relevante que usa el experto en la resolución.

## Incidentes críticos

- El experto que describa casos especialmente interesantes o difíciles que se le hayan presentado. El experto, además, deberá contar cómo los resolvió.
- Esta técnica es muy interesante, ya que los casos especialmente complejos hacen comentar detalles que, en otro momento, hubiera pasado por alto.
- Una variante es plantear casos críticos imaginarios. Se le pide al experto que considere un caso cualquiera y se le convierte en un incidente crítico mediante el planteamiento de situaciones alternativas de la forma: ¿qué pasaría si ...?.

# Imposición de restricciones

- Se pretende que el experto verbalice ciertas deducciones que, en un caso familiar, pasaría por alto, imponiéndole limitaciones en el uso de recursos.
- Los recursos a restringir pueden ser:
  - el tiempo de resolución (se necesita una respuesta rápida),
  - la información disponible (faltan datos de entrada).
- Es útil para educir cierto tipo de conocimientos como pueden ser las prioridades de atención por parte del experto.

# Clasificación de conceptos

- En primer lugar se obtiene, a partir de un simple glosario o texto, un conjunto de conceptos que cubran ampliamente el dominio. A continuación, se transfiere cada concepto a una ficha y se le pide al experto que las clasifique en una serie de grupos, describiendo lo que cada grupo tiene en común. Por último, las fichas de cada grupo se comparan para formar jerarquías.
- Esta técnica es especialmente aconsejable cuando hay un gran número de conceptos en un dominio, de modo que requieren una estructuración para que sean manejables.

# Cuestionarios

- Consiste en realizar una entrevista estructurada al experto de forma indirecta a través de cuestionarios.
- Esta técnica tiene la ventaja, frente a las entrevistas, de ser una forma eficiente de acumular información.
- Los cuestionarios pueden ser particularmente apropiados para describir los conceptos, revelar las relaciones en el dominio y determinar incertidumbre.

# Ejemplo de cuestionario

- ☐ Nombre de la variable: *ventas*
- ☐ Descripción: *la cantidad de una línea de producto realmente*
- ☐ *vendida por un representante en una zona particular*
- ☐ Tipo de valores: *numéricos*
- ☐ Clase de tamaño y valores
  - ☐ rango: *0-100.000*
  - ☐ inciertos: *sí no*
- ☐ ¿Estatus al comienzo?
  - ☐ *Conocido Desconocido*
- ☐ Depende de otras variables
  - ☐ *Sí No*