

Joaquín García Benítez y Clara Torre García-Barredo.

Problemas de Sistemas de Información.



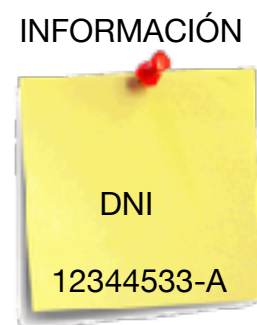
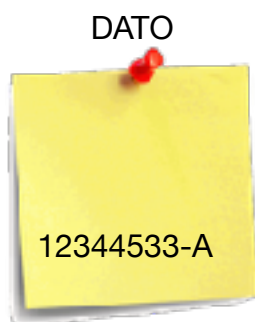
Tema 1.

En este tema los ejercicios propuestos son sencillos e introducen al contenido de la asignatura.

Lo primero que se ha de tener claro para hacer estos ejercicios es la diferencia entre dato e información. Una vez que este concepto se tiene claro, se puede resolver el tipo de ejercicios cuyo enunciado es:

EJERCICIO:

Clasificar los siguientes elementos como datos o información, justificando la respuesta.



En este ejemplo se ve que la misma cadena de caracteres puede ser un dato, si no va acompañada de un contexto que permita identificar a qué se refiere.

El resto de ejercicios del tema 1 no son relevantes para la asignatura.

Temas 2 y 3.

En este caso, hay varios tipos de ejercicios.

El primer tipo es aquel con este enunciado:

EJERCICIO:

Clasifique las siguientes decisiones según su tipo y justifique su elección.

Para la resolución de este ejercicio hay que recordar la diferencia entre una decisión operativa, una decisión táctica y una decisión estratégica.

- **Decidir el presupuesto anual.**

Es una decisión de tipo táctica porque es el dinero que va a tener la empresa para manejar durante todo un año.

- **Alquilar una máquina durante un mes porque se ha roto la que posee la empresa.**

Es una decisión de tipo operativa porque soluciona momentáneamente un problema existente.

EJERCICIO:

Indique cuál de los siguientes sistemas ERP, CRM, SCM o KMS, es el encargado de proporcionar soporte a cada una de las siguientes operaciones. Justifique su respuesta.

- **Mantenimiento de datos de los empleados.**

Es un sistema de tipo ERP porque todos tienen áreas de recursos humanos.

- **Gestión y tratamiento de pedidos.**

Es un sistema de tipo CRM si se piensa en pedidos de venta, y SCM si se piensa en pedidos de compra.

- **Incluir un blog con noticias de la empresa.**

Es un sistema de tipo KMS porque es propio conocimiento sobre la empresa.

Hay ejercicios de tipo test, en los que la teoría es la única manera de resolver las preguntas. No son excesivamente complicadas, porque están destinadas a ser un pequeño repaso de los temas 2 y 3.

Un ejemplo podría ser este:

TEST:

2) Señale cuál de las siguientes las siguientes operaciones es característica de cualquier sistema de información:

- a) Convertir información en datos.**
- b) Convertir datos en información.**
- c) Convertir información no cifrada en información cifrada.**
- d) Convertir datos no cifrados en datos cifrados.**

La respuesta correcta es la b.

El último ejercicio importante de estos dos temas es el que pregunta si los requisitos de un proyecto son de tipo funcional o no funcional. De nuevo, esto se basa exclusivamente en la teoría, y un ejemplo de resolución sería el siguiente:

EJERCICIO:

Clasifique los siguientes requisitos en funcionales y no funcionales:

- **El sistema web debe cumplir con los criterios de accesibilidad WAI-AA.**

Este requisito es no funcional.

Tema 4.

En este tema hay tres tipos de ejercicios:

1. Modelo vectorial de recuperación de la información.
2. Métricas del modelo de recuperación.
3. Diagrama de Flujo de Datos.

En el primero de los tres tipos de ejercicios, lo más importante es hacer una tabla en la que en el eje horizontal estén todas las palabras que se quieren analizar, y en el eje vertical cada documento que se quiere analizar, de esta manera:

	PALABRA 1	PALABRA 2	PALABRA 3	PALABRA 4
DOCUMENTO 1				
DOCUMENTO 2				
DOCUMENTO 3				
DOCUMENTO 4				

Para rellenar la tabla lo único que hay que hacer es contar cuántas veces aparece cada palabra en un documento. Así, si la PALABRA 1 aparece dos veces en el DOCUMENTO 1 y una vez en el DOCUMENTO 2, la tabla quedaría así:

	PALABRA 1	PALABRA 2	PALABRA 3	PALABRA 4
DOCUMENTO 1	2			
DOCUMENTO 2	1			
DOCUMENTO 3	0			
DOCUMENTO 4	0			

Al acabar la tabla hay que calcular la frecuencia inversa de cada palabra, siguiendo esta fórmula:

$$F_{inversa} = \log_{10}(n_d/d_p)$$

En esta fórmula n_d significa el número de documentos totales (en el ejemplo 4), y d_p es el número de documentos en los que aparece la palabra (en el ejemplo 2).

Se repite la tabla colocando las frecuencias inversas de cada palabra, y multiplicándolas por el número de veces que aparece esa palabra en cada documento. Así, si la PALABRA 1 tiene una frecuencia inversa de 0.301 y aparece 2 veces en el DOCUMENTO 1, la tabla quedaría como sigue:

	PALABRA 1	PALABRA 2	PALABRA 3	PALABRA 4
DOCUMENTO 1	0,602			
DOCUMENTO 2	0,301			
DOCUMENTO 3	0			
DOCUMENTO 4	0			

Una vez terminada la tabla, se calcula la semejanza, para determinar cuál de los documentos tiene mayor relevancia, con la fórmula proporcionada en los apuntes.

$$\text{semejanza}(\vec{d}_j, \vec{q}) = \frac{\vec{d}_j \cdot \vec{q}}{|\vec{d}_j| \times |\vec{q}|} = \frac{\sum_{i=1}^t w_{ij} \times w_{iq}}{\sqrt{\sum_{i=1}^t w_{ij}^2} \times \sqrt{\sum_{i=1}^t w_{iq}^2}}$$

El cálculo de los pesos se efectúa así:

- w_{ij} : peso del término j en el documento i
- $f_{t_{ij}}$: frecuencia del término j en el documento i
- n: número de términos incluidos en el vocabulario
- f_{d_j} : número de documentos que incluyen el término j
- d: número de documentos
- $f_i d_j$: $\log_{10}(d/f_{d_j})$
 - ⦿ Su valor mínimo es cero cuando $f_{d_j} = d$ (t_j aparece en todos los documentos).
 - ⦿ Su valor máximo es $\log(d)$ cuando $f_{d_j} = 1$ (t_j solo aparece en un documento).
- $w_{ij} = f_{t_{ij}} * f_i d_j$

Para las consultas se calcula w_{qj} de modo análogo a como se hace con w_{ij} .

El segundo tipo de ejercicio es el que consiste en calcular las métricas del modelo de recuperación. En concreto son cuatro métricas:

1. Índice de recuperación.
2. Índice de precisión.
3. Índice de relevancia.
4. Índice de fallos.

EJERCICIO:

Suponiendo que solamente el DOCUMENTO 2 y el DOCUMENTO 3 son relevantes, calcule los índices.

En este caso, el índice de recuperación se calcula siguiendo la siguiente fórmula:

$$I_{rec} = d_{rel}/d_{rec-rel}$$

En esta fórmula, se divide el número de documentos relevantes (en este ejemplo 2) entre el número de documentos relevantes recuperados. Suponiendo que hemos recuperado solamente el DOCUMENTO 2, este índice sería 0.5.

El segundo índice es el de precisión, que valora la siguiente fórmula:

$$I_{prec} = d_{rec-rel}/d_{rec}$$

En este caso, el índice de precisión mide el número de documentos relevantes que había entre los documentos recuperados. En nuestro caso, el índice sería 1, porque todos los recuperados eran relevantes.

El tercer índice es el de relevancia, que mide la siguiente fórmula:

$$I_{rel} = d_{rel}/d_{total}$$

En este caso, el índice de relevancia mide el porcentaje de documentos relevantes que hay en el número total de documentos. En nuestro caso, sería el 0.5.

El último índice es el de fallos, que se pregunta cuántos documentos no relevantes han sido recuperados:

$$I_{prec} = d_{rec}/d_{rec-rel}$$

En nuestro caso sería 0, porque solo hemos devuelto documentos relevantes.

Y, por último, el tercer tipo de ejercicios: El diagrama de flujo de datos.

EJERCICIO:

Se requiere el diagrama de flujo de datos.

Se da información sobre cómo funciona el negocio del que se intenta realizar el diagrama, y a partir de esa misma información se realiza el dibujo.

Un ejemplo de cómo podría quedar un diagrama sería este:



donde los círculos son acciones, los rectángulos son personas con las que se interacciones y las flechas son las actividades que se pueden realizar sobre un círculo con las personas implicadas.

Tema 5.

En este tema hay un solo tipo de ejercicio: sugerir un producto nuevo en base a productos valorados anteriormente.

Este tipo de ejercicios se resuelven de la siguiente manera:

EJERCICIO:

Aplique una técnica de filtrado colaborativo para determinar qué producto debería sugerir al usuario A en el catálogo electrónico de este sistema de comercio on-line.

El enunciado suele venir con una tabla con las valoraciones de los usuarios a determinados productos, de la siguiente manera:

	PRODUCTO A	PRODUCTO B	PRODUCTO C	PRODUCTO D
CLIENTE A	3	5	?	?
CLIENTE B	4	2	3	5
CLIENTE C	4	4	2	4

Se calculan las similitudes o porcentaje de similitud entre dos usuarios con la fórmula dada:

$$\text{sim}(a, b) = \frac{\sum_{p \in P} (r_{a,p} - \bar{r}_a)(r_{b,p} - \bar{r}_b)}{\sqrt{\sum_{p \in P} (r_{a,p} - \bar{r}_a)^2} \sqrt{\sum_{p \in P} (r_{b,p} - \bar{r}_b)^2}}$$

- ◆ **a, b** son los usuarios.
- ◆ $r_{a,p}$ es la valoración del usuario a al producto p.
- ◆ **P** es el conjunto de productos valorados por a y b.
- ◆ La **medida de similitud** toma valores entre -1 y 1.
- ◆ También predice cuál será la satisfacción del usuario con el producto ofertado a partir de los datos relativos a otros usuarios similares.
- ◆ **a, b** son los usuarios.
- ◆ \bar{r}_a son las valoraciones promedio del usuario a.
- ◆ r_b es la valoración del usuario b al producto p.
- ◆ **sim(a,b)** es la similitud entre el usuario a y el usuario b.

Aquellos usuarios con resultados negativos se descartan, y los usuarios que queden con resultados de similitud positivos se utilizan para predecir la opinión que tendrá el USUARIO A sobre el producto a sugerir con la siguiente fórmula:

$$pred(a, p) = \bar{r}_a + \frac{\sum_{b \in N} sim(a, b) * (r_{b,p} - \bar{r}_b)}{\sum_{b \in N} sim(a, b)}$$

El producto con mayor valoración será aquel que haya que sugerir.