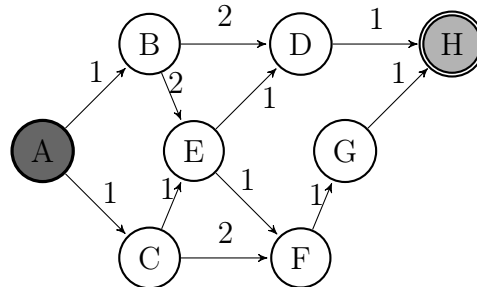


# Examen Final de IA

(18 de enero de 2013)

Duración: 3 horas

1. (1.5 puntos) Dado el siguiente grafo, donde cada arco indica su coste, y la tabla que indica la estimación del coste  $h$  hasta la solución, indica cual sería el árbol de búsqueda que se obtendría mediante el algoritmo de  $A^*$  e  $IDA^*$  para encontrar el camino entre el nodo A y el nodo H. Haz la generación de los nodos siguiendo el orden alfabético e indica claramente las reexpansiones de los nodos y los cambios de coste que aparezcan. ¿Es la función heurística admisible?



Nodo	A	B	C	D	E	F	G	H
$h(\text{nodo})$	6	4	4	4	4	3	3	0

2. (2.5 puntos) La compañía de subastas on-line eBay quiere crear aplicaciones para poder tratar todas las transacciones que realizan sus usuarios y como paso inicial quiere definir una ontología que describa toda la información que maneja.

El primer concepto que quiere usar es el de usuario, que puede especializarse en vendedor y comprador (un usuario puede pertenecer a ambas categorías). El segundo es el de producto, que puede clasificarse en electrónica (ordenadores, cámaras, consolas, móviles, ...), coleccionables (monedas, sellos, comics, ...), entretenimiento (cine, música, ...), motor (coches, motos, barcos, ...), juguetes (radio control, trenes, infantiles, ...), ... El tercero es la subasta, que puede especializarse en subastas rápidas (menos de 24 horas) y subastas largas. El último es la puja, que representa que un usuario comprador ha pujado en una subasta cierta cantidad de dinero.

- Están considerando las siguientes relaciones, la relación *abre* entre vendedor y subasta, la relación *vende* entre subasta y producto, la relación *envía* entre comprador y puja y la relación *oferta* entre puja y subasta. En una subasta se vende un único producto que, si no se adjudica a nadie, no se puede volver a subastar. Define completamente estas relaciones utilizando el lenguaje de frames visto en clase.
- Cuando un vendedor quiere vender un producto, ha de crear su descripción en eBay, es decir, primero ha de ponerlo a la venta dando una descripción para luego poder subastarlo. Define la relación *en venta* entre vendedor y producto. Define el slot *pendientes* en vendedor que retorne el número de productos que un vendedor aún no ha subastado.
- Una subasta tiene una fecha de inicio, una fecha de final y un precio inicial. Una puja tiene una cantidad ofrecida y una fecha. Asumiendo que el método de subasta es al alza, define lo necesario para que se compruebe que una puja sea correcta, es decir, que una puja ofrezca una cantidad mayor que la última que se ofreció.
- Los usuarios vendedores tienen un slot que indica su índice de fiabilidad (cómo es el producto real respecto a lo que dice en su oferta) que es un valor entero entre 0 y 5. Modifica lo necesario para que una subasta herede el índice de fiabilidad de su vendedor. ¿Sería posible redefinir la relación *vende* para heredar este slot en producto? Explica tu respuesta.
- Implementa un método *envíos* que nos retorne una lista con el identificador de los compradores que han ganado las subastas de un día concreto junto con el identificador de los productos que han comprado. Indica donde se definiría este método, añadiendo los elementos que sean necesarios a la representación.

3. (2.5 puntos) Deseamos crear un sistema capaz de aconsejar a un usuario como invertir de la mejor manera los ahorros de los que dispone. Para ello queremos recoger toda una serie de informaciones que nos puedan indicar cuáles son sus características como ahorrador/inversor y a partir de ellas darle unas prioridades sobre qué productos podrían ser más interesantes para el. La información que le pedimos al usuario incluye:

- Su edad.
- Si su perspectiva de ahorro futuro es favorable o desfavorable.
- Si tiene deudas, cuanta proporción de su salario supone pagar esas deudas.
- Su actitud respecto a fluctuaciones negativas interanuales de sus inversiones de un 5 %, un 10 %, un 20 % o más de un 20 % (moderada, negativa, muy negativa).
- Su actitud respecto a la oportunidad de ganancia contra la posibilidad de pérdida (seguridad ante todo, ganar un poco más compensa si la posibilidad de pérdida es pequeña, la ganancia ante todo).
- Objetivo del ahorro (garantizar el capital, mantener la inflación, obtener ganancias sobre la inflación, obtener grandes ganancias).
- Experiencia en la administración de inversiones (nada, poca, bastante).
- El ahorro de impuestos (indiferente, importante o muy importante).

Hemos decidido abstraer la información obtenida del usuario en cinco características:

- **Riesgo**, que mide el riesgo que puede o quiere asumir el cliente, valorado como bajo, medio, alto y muy alto. El riesgo está asociado a la edad (cuanto más viejo, menos riesgo se puede asumir), a la actitud a las fluctuaciones (cuanto peor, menos riesgo), a la proporción que suponen las deudas respecto al salario (cuantas más deudas, menos riesgo), a la actitud respecto a la oportunidad de ganancia, el objetivo de ahorro (cuantas más ganancias más riesgo) y la experiencia en la administración de inversiones (cuanta más experiencia, más riesgo).
- **Diversificación**, que mide la variedad de valores en los que invertir, valorado como poca, media y mucha. La diversificación está asociada a la actitud a la oportunidad de ganancia, (cuanta más seguridad se quiera respecto a la ganancia, mayor ha de ser la diversificación), la actitud respecto a las fluctuaciones (cuanto más negativa, más diversificación) y la experiencia en la administración de inversiones (cuanta menos experiencia, más diversificación).
- **Plazo**, que mide el tiempo que se debe mantener la inversión, valorado como corto, medio o largo. El plazo depende de la actitud respecto a las fluctuaciones (cuanto menos negativa, el plazo puede ser más largo), la perspectiva del futuro ahorro (cuanto mejor sea, más plazo), el objetivo de ahorro (cuantas más ganancias, mayor ha de ser el plazo) y el ahorro de impuestos (se pagan más impuestos cuanto más corto sea el periodo de la inversión).
- **Administración activa**, que mide lo involucrado que debe estar el cliente en la evolución y manejo de sus inversiones, valorada en constante, moderada y poca. Esta característica depende de la experiencia del cliente (cuanta menos experiencia, menos involucrado), el objetivo del ahorro (cuanta mayor la ganancia, más deberá estar involucrado) y la actitud respecto a las fluctuaciones (cuanto más negativa sea a mayores pérdidas, más deberá estar involucrado).
- **Liquidez de la inversión**, que mide la facilidad para convertir la inversión en dinero a corto plazo, valorada como alta, moderada y baja. La liquidez dependerá de la proporción de las deudas (cuantas más deudas, más liquidez), la experiencia en administración de inversiones (cuanto menos experiencia, menos liquidez), la actitud respecto a la pérdida (cuanto menos negativa, menos liquidez es necesaria) y el objetivo de ahorro (cuanto menos ganancia, menos liquidez).

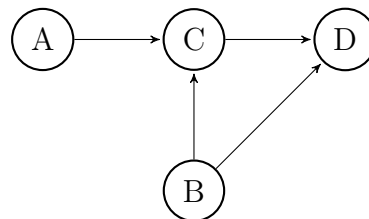
A partir de evaluar estos criterios queremos obtener una prioridad para los diferentes tipos de productos que se pueden ofrecer al cliente: fondos de pensiones, acciones, fondos de inversión (en acciones, renta fija o mixtos) y depósitos (a corto plazo o a largo plazo). Esta prioridad debería obtenerse como un valor entero entre 0 y 5, indicando la importancia de incluir ese tipo de valor en la inversión.

La experiencia del dominio nos dice que:

- Los fondos de pensiones son más convenientes cuanto menos riesgo se quiera, son mejores cuando se quiere una buena diversificación, tienen mejor resultado cuanto más largo es el plazo, no requieren una administración activa y no tienen liquidez.
- Las acciones son más recomendables para inversiones con más riesgo, siempre que se pueda hacer una administración activa, permiten menos diversificación, se puede invertir en ellas a cualquier plazo, pero son más recomendables a corto y tienen total liquidez.
- Los fondos de inversión permiten una mayor diversificación, requieren poca administración y suelen tener liquidez alta o moderada. Los de acciones son más recomendables a largo plazo cuando se quiere menos riesgo, los de renta fija son más recomendables cuanto menos riesgo y los mixtos equilibran el riesgo cuando el plazo es más corto.
- Los depósitos tienen menor liquidez cuanto mayor sea el plazo, no requieren ninguna administración, son recomendables cuanto menos riesgo y diversificación se quiera

- a) El problema descrito es un problema de análisis. Explica detalladamente cómo lo resolverías usando clasificación heurística. Da ejemplos de reglas para cada una de las fases de esta metodología, al menos una por cada característica abstracta y una por cada solución.
- b) Una forma alternativa de solucionar el problema sería utilizar el razonamiento basado en casos. En lugar de plantear el problema de priorizar los tipos de productos como un razonamiento a partir de las características de los clientes, tendríamos ejemplos ya resueltos. Explica como se resolvería el problema con esta metodología: que serían los casos, como se describirían, que sería una solución en un caso, como se recuperarían los casos y como se obtendría la solución de un nuevo caso a partir de los casos recuperados.

4. (1 punto) Dada la siguiente red bayesiana y las probabilidades asociadas a cada nodo:



A	P(A)	B	P(B)	A		P(C A, B)		B	C	P(D B, C)	
				C	F	C	F			C	F
C	0.4	C	0.3	C	C	0.3	0.7	C	C	0.2	0.8
F	0.6	F	0.7	C	F	0.4	0.6	C	F	0.6	0.4
				F	C	0.6	0.4	F	C	0.3	0.7
				F	F	0.1	0.9	F	F	0.4	0.6

Define la distribución de probabilidad conjunta de  $P(A, B, C, D)$  que define la red y usa el algoritmo de eliminación de variables para calcular la probabilidad de  $P(D|a)$ .

5. (2.5 puntos) Queremos implementar una aplicación de análisis de opinión de los mensajes de *twitter* y como paso previo queremos extraer información sobre si los tweets son positivos o negativos y en que grado. Hemos decidido solo analizar expresiones como las siguientes:

- <marca> es buena
- <marca> es muy mala
- <marca> es la mejor
- <marca> es la peor
- <marca> hace muy buenos <producto>
- <marca> fabrica buenas <producto>

<marca> produce malos <producto>

<marca> fabrica las peores <producto>

Para reconocer estas frases se ha desarrollado la siguiente gramática:

oracion --> o\_es.

oracion --> o\_hace.

o\_es --> marca, verb, sent\_ah.

o\_hace --> marca, verb, sent\_ah, prod.

sent\_ah --> adj.

sent\_ah --> adv, adj.

sent\_ah --> det, super.

Con el siguiente lexicón:

marcas(cocacola).

marcas(apple).

marcas(ibm).

productos(bebidas).

productos(telefonos).

productos(ordenadores).

verbo(es).

verbo(hace).

verbo(fabrica).

verbo(produce).

adjetivo(buena).

adjetivo(mala).

adverb(muy).

determ(el).

determ(la).

superlat(mejor).

superlat(peor).

- a) Añade a la gramática lo que falta para que se pueda reconocer las frases.
- b) Modifica la gramática para que pueda haber coordinación en género y número donde corresponda.
- c) Modifica la gramática para que los verbos de las frases sean coherentes con la estructura de la frase. Es decir, la gramática no debe admitir una frase de 'ser' con un verbo del tipo 'hacer'.
- d) Modifica la gramática para que genere como salida el predicado  $opinion(X, Y)$ , donde la primera variable es el tipo de opinión, que ha de ser **positiva** o **negativa** y la segunda variable es el grado de la opinión que puede ser **normal**, **alta** o **muy alta**.

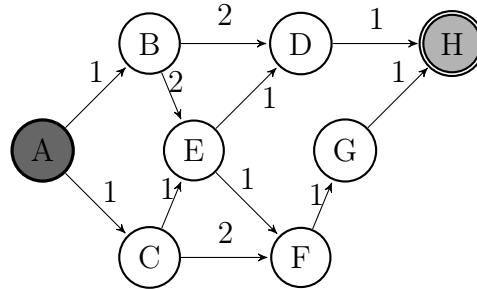
Las notas se publicaran el día **23 de enero**.

# AI Final Exam

(January 18<sup>th</sup> 2013)

Time: 3 hours

1. (1.5 points) Given the following graph, where each edge indicates its cost, and the table that gives the estimated cost  $h$  from each node to the solution node, draw the search tree that can be obtained using the A\* and IDA\* algorithms to find the path from node A to node H. Use the alphabetical order as generation order and indicate clearly the node reexpansions and the cost changes during the search. Is the heuristic function admissible?



Node	A	B	C	D	E	F	G	H
$h(\text{node})$	6	4	4	4	4	3	3	0

2. (2.5 points) The on-line auction company eBay wants to develop new applications to process all the transactions of the users of its site and as an initial step wants to define an ontology that describes all the available information.

The first concept to define is user, that can be specialized in seller and buyer (a user can be from both categories). The second one is product, that can be specialized in electronics (computers, cameras, phones, ...), collectibles (coins, stamps, comics, ...), entertainment (movies, music, ...), motor (cars, motorcycles, boats, ...), toys (radio control, trains, children, ...), ... The third one is auction, that can be specialized in short auctions (less than 24 hours) and long auctions. The last one is bid, representing that a buyer bids in an auction a specific amount of money.

- We want to consider the following relations, the relation *opens* between seller and auction, the relation *sells* between auction and product, the relation *posts* between buyer and bid and the relation *offer* between bid and auction. An auction only sells a product and if it is not won by anyone it can not be auctioned again. Define completely these relations using the frame language used in class.
- When a seller wants to sell a product, he has to create its description in eBay, from that moment the product can be auctioned. Define the relation *for sale* between seller and product. Define the slot *pending* in seller returning the number of products that a seller has not auctioned yet.
- An auction has a starting date, a final date and an initial price. A bid has an amount and a date. Assuming that the auction goes to the highest bidder, define the necessary in order to check that a bid is correct, that is, a bid has to offer a higher amount than the last bid.
- The sellers have a slot that indicate how trustworthy they are (how is really the product compared with what they say about it) that is an integer value from 0 to 5. Modify all the necessary so this slot is inherited by auction from seller. Would it be possible to redefine the relation *sells* to inherit this slot in product? Elaborate your answer.
- Implement a method *deliveries* that returns a list with all the identifiers of the buyers that have won the auctions of a specific day along with the identifier of the product that they have bought. Explain where has to be located and add all the necessary elements to the representation needed to implement the method.

3. (2.5 points) We want to implement a system able to give advice to a user about how to better invest his savings. In order to do that we want to obtain some information that can give us an idea about his investor profile and, from that, to advice about the priorities to give to the different products we can offer him. The information that we ask to the client includes:

- Age.
- If the perspective of having more savings in the future is good or not.
- If in debt, what is the portion of the salary devoted to repay it.
- Attitude towards interannual fluctuations of investments of a 5 %, a 10 %, a 20 % or more than a 20 % (moderated, negative, very negative).
- Attitude towards opportunity of earnings against possibility of losses (always secure, a little bit more earnings if possibility of losses is small, earnings first).
- Investment goal (guarantee main capital, maintain inflation, earn more than inflation, earn a lot of money).
- Experience managing own investments (none, a little bit, a lot).
- Tax savings (indifferent, important, very important).

We have decided to abstract all these information into five characteristics:

- **Risk**, that measures the risk that the client wants or can assume, valued as low, medium, high and very high. Risk is associated with age (the older, the lower risk can be assumed), the attitude towards fluctuations (the more negative, the lower has to be the risk), the portion of the salary that takes the debt (the more debt, the lower the risk has to be), the attitude towards opportunity of earnings, the investment goal (the more he wants to gain, the higher the risk) and the experience managing own investments (the more experience, the higher the risk).
- **Diversification**, that measures the amount of different items the investments have to have, valued as small, medium and high. Diversification is associated to the attitude towards opportunity of earnings (the more security is needed, the more has to be the diversification), the attitude towards fluctuations (the more negative, the more the diversification) and the experience managing own investments (the less the experience, the more the diversification).
- **Term**, that measures how much time the investment has to be maintained, valued as short, medium and long. The term depends on the attitude towards the fluctuations (the more negative, the longer has to be the term), the perspective of future savings (the better, the longer the term), the investment goal (the more earnings are wanted, the longer has to be the term) and the tax savings (the shorter the term, more taxes have to be payed).
- **Active management**, that measures how invested has to be the client in the management and the evolution of his investments, valued as constant, moderate and none. This characteristic depends on the experience of the client (the less the experience, the lower the management), the investment goal (the more the client wants to earn, the more the management) and the attitude towards fluctuations (the more negative to higher losses, the more the management).
- **Liquidity**, that measures how easy is to quickly cash the investment, valued as high, moderate and low. Liquidity depends on the proportion of debt (the more the debt, the higher the liquidity), the experience managing own investments (the less the experience, the lower the liquidity), the attitude towards losses (the less negative, the less the liquidity) and the savings goal (the less gains are wanted, the less the liquidity).

From all these criteria we want to obtain a priority for all the different types of investment products that can be offered to the client: retirement funds, stocks, investment funds (in stocks, bonds or mixed) and fixed term deposits (short and long term). That priority has to be an integer from 0 to 5, indicating the importance of having that product in the investment plan.

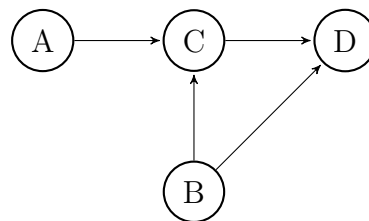
The experience in the domain tells us that:

- Retirement funds are more convenient when low risk is indicated, they are better when a good diversification is needed, they have better yield at long term, they do not require active management and they have low liquidity.

- Stocks are more recommended for higher risk investment as long as an active management is followed, they do not allow high diversification, investment can be to any term and liquidity is high.
- Investment funds allow higher diversification, they require low active management and liquidity is high to moderate. Stock funds are more recommended for long term when low risk is needed, bonds funds are recommended for low risk and mixed funds have a lower risk for short term.
- Fixed term deposits have less liquidity the longer the term, they do not require active management, they are recommended when lower risk and diversification is needed.

- a) The problem described is an analysis problem. Explain how would you solve it using heuristic classification. Give some examples of rules for each one of the steps of the methodology, at least one for each abstract characteristic and one for each solution.
- b) An alternative way to solve this problem would be to use case based reasoning. Instead of defining the problem of obtaining a priority for the different types of products as a reasoning from the characteristics of the clients, we would have cases of solved problems. Explain how this problem can be solved using this methodology: what is a case, how has to be described, what is the solution of a case, how cases can be retrieved and how the solution of a new case can be obtained from the retrieved cases.

4. (1 point) Given the following bayesian network and the probabilities associated to each node:



A	P(A)	B	P(B)	A	B	P(C A, B)		B	C	P(D B, C)	
						C	F			C	F
C	0.4	C	0.3	C	C	0.3	0.7	C	C	0.2	0.8
F	0.6	F	0.7	C	F	0.4	0.6	C	F	0.6	0.4
				F	C	0.6	0.4	F	C	0.3	0.7
				F	F	0.1	0.9	F	F	0.4	0.6

Write the expression of the joint probability distribution  $P(A, B, C, D)$  that represents the network and calculate the probability of  $P(D|a)$  using the variable elimination algorithm.

5. (2.5 points) We want to implement a system for sentiment analysis of *twitter* messages and as a first step we want to extract information about if a tweet is positive or negative and in what grade. We have decided to analyze only expressions like the following:

- <brand> is good
- <brand> is very bad
- <brand> is the best
- <brand> is the worst
- <brand> makes very good <product>
- <brand> makes good <product>
- <brand> produces bad <product>
- <brand> builds the worst <product>

To recognize these sentences the following grammar has been developed:

```
sentence --> s_is.  
sentence --> s_makes.  
  
s_is --> brand, verb, sent.  
s_makes --> brand, verb, sent, prod.  
  
sent --> adj.  
sent --> adv, adj.  
sent --> det, super.
```

With the following lexicon:

```
brands(cocacola).  
brands(apple).  
brands(ibm).  
  
products(sodas).  
products(phones).  
products(computers).  
  
verbs(is).  
verbs(makes).  
verbs(builds).  
verbs(produces).  
  
adjective(good).  
adjective(bad).  
  
adverb(very).  
  
determ(the).  
  
superlat(best).  
superlat(worst).
```

- a) Add to the grammar all necessary so all the sentences are correctly recognized.
- b) Modify the grammar so the correct verb is used for each type of sentence (to be or verbs of making).
- c) Modify the grammar so the predicate `sentiment(X,Y)` is obtained as output, being the first variable the type of the sentiment, that is `positive` or `negative` and the second variable being the grade of the sentiment that is `normal`, `high` or `very high`.

Grades will be published **January 23rd**.