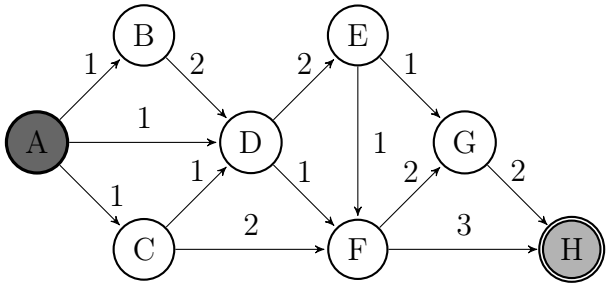
# AI Final Exam

**(January 17$^{th}$ 2012)**

**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?




| Nodo | A | B | C | D | E | F | G | H |
|---|---|---|---|---|---|---|---|---|
| h(nodo) | 4 | 4 | 3 | 3 | 2 | 3 | 2 | 0 |

2. (2 points) The company Twitter[1] wants to build applications to process all the content that is published in this service and as a first step an ontology describing the concepts involved is needed.

   The first concept they want to define is the user, that can be specialized in people and entities. The second is the message (or *tweet*) that can be long (if it has more that 70 characters) or short. The third concept is the topic (or *hashtag*), that is the term that the users use to tag their messages, classified in personal, commercial and news. A tweet can have several hashtags.

   (a) Twitter is considering to use the following relations, the relation *follower* among users, that represents that a user follows the tweets of other users, the relation *tweeted* between a user and his tweets and the relation *about* between a tweet and its hashtags. Define completely these relations using the frame language used in class.

   (b) In Twitter is called to *retweet* that a user repeats the messages of other users, copying or referencing its content in its own messages. You can also retweet another retweet. Define the relation *is_a_retweet* as a relation among messages. Define the slot *echo* that returns how many times a user *retweets* messages from other users.

   (c) Define a method in message that returns the user if it is an original message (it is not a retweet) or the user that originated the message if it is not original. Have in mind that a retweet does not have to be of the original message.

   (d) A *trending topic* is a topic that appears in tweets over a period of time much more than the rest of topics. Could a slot *trending topic* be defined in the frame topic that returns the ten most commented topics over a period of time? If this is not the best solution, explain the cause and define an alternative, adding to the representation anything you need. You can assume that you have a clever datastructure that stores topics, counts the number of times a topic has been added and that allows to obtain the $n$ most frequent topics.

3. (3 points) This time of the year *christmas gift baskets* are a usual gift. There are many companies that allow their clients to customize the contents of these baskets. We want to use a KBS to help the clients about the contents of the basket depending on some characteristics.

[1]For those that do not know Twitter, it is a *microblogging* service where users can post messages up to 140 characters long about anything.

The clients of this service can be business, that can be classified in small enterprises, medium enterprises, large enterprises and transnationals (the larger the business the more they can afford). We have also information about how many gift baskets they want to buy (depending on the purchase volume we can put more content for the same price) and what is the price range, *mr. Scrooge* (less than 20 euros), *normal* (up to 40 euros), generous (up to 60 euros), *totally a bribe* (more than 60 euros).

About the receivers of the gift, we now if they are *clients* (business usually treat clients better), having only a business relation or a personal relation, or *employees*, being those administrative personnel, middle management or upper management (everybody knows how this works). We have also some information about general or particular preferences and constraints of the receivers about the different products.

As designers of the system we have decided to use the following abstract characteristics as the base of our recommendation:

- **Quantity**, that measures the approximated volume of products that the basket should contain. It is evaluated as *a few* (between 5 and 8 products), *medium* (between 9 and 12 products), *a lot* (between 13 and 18 products), *very generous* (up to 25 products).
- **Quality**, that measures the average quality of the products in the basket. It is evaluated as *equilibrated* (all the products of a similar quality), *something good* (some products of a higher quality than the rest), *something really good* (some products of exceptional quality).
- **Variety**, that measures how many different products the basket will contain. it is evaluated as *low* (products from two different categories), *medium* (products from four categories) and *high* (products from all the categories).
- **Food/drinks**, measures the ratio between food products and drinks in the basket. It is evaluated as *all the same*, *more food* and *more drinks*.

From evaluating these criteria we want to obtain a priority for each one of the different kinds of products that a christmas gift basket can have: sweets, wine, cava[2], liquor, cold meats/cheese and canned food. The priority will be obtained as a natural number from 0 to 2, 0 meaning that that product should not be in the basket.

We know that the abstract characteristics have different relations with these products, for example, the baskets of higher quality usually give more priority to drinks and cold meats, when quantity is more important than quality usually there is more canned food, sweets and cava, or if drinks are more important than food usually there is more wine and liquor than cava, ...

(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.

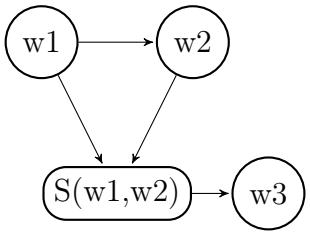
(b) The characteristics used to define the priorities are not independent. For instance, the quantity has influence over the quality, the variety also influences the quality and the ratio between food and drinks influences variety. The product priorities are also related, because there are combinations that make better sense than others. For instance, the importance of cava influences the priority of the sweets and if there is more cold meat/cheese this will influence the priority of the wine, ...
Define the heuristic association step as a bayesian network. Represent in the network at least the relations that have been explained and also all the necessary relations so the heuristic association problem is solved. Give an example of probability table of a node, you can made up the probabilities, but these have to represent how the values of the parent nodes influence the probability of the values of the descendant nodes.

(c) 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 commands, we would have cases of solved problems. Explain

[2]Here in Catalonia we have cava, not champagne.

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) Bayesian networks can be used for any decision problem, for example, in natural language processing, we can use this formalism for determining the probability of the different syntactic roles for a word. Lets suppose that we have three consecutive words $w1$, $w2$ y $w3$ and also the syntactic group the two first words belong to $S(w1, w2)$, the following bayesian network describes the probability of the syntactic category of the word $w3$ as a function of the rest of the variables:




| $w1$ | $P(w1)$ |
|---|---|
| adj | 0.8 |
| adv | 0.2 |

| $w1$ | $P(w2\|w1)$ | |
|---|---|---|
| | name | verb |
| adj | 0.7 | 0.3 |
| adv | 0.1 | 0.9 |

| $w1$ | $w2$ | $P(S(w1,w2)\|w1,w2)$ | |
|---|---|---|---|
| | | SN | nSN |
| adj | name | 0.9 | 0.1 |
| adj | ver | 0.5 | 0.5 |
| adv | name | 0.8 | 0.2 |
| adv | ver | 0.3 | 0.7 |

| $S(w1,w2)$ | $P(w3\|S(w1,w2))$ | |
|---|---|---|
| | name | adj |
| SN | 0.7 | 0.3 |
| nSN | 0.2 | 0.8 |

Define the joint probability distribution of $P(w1, w2, w3, S(w1, w2))$ represented by the bayesian network and use the variable elimination algorithm to compute the probability of $P(w3|w1 = adj)$.

5. (2.5 points) The restaurant chain *Fancy Food for Fancy Fee* is automatizing to take the food orders from the clients. We want to develop a natural language processing system able to recognize the order in which the food has to be served and the dishes from the menu. The system has to recognize food orders like these:

   - Starters two soups, main two of chicken and no dessert
   - No starters, main one of turkey and one of salmon and dessert two chocolate cakes
   - Starters two soups and a salad, main one of beef, one of turkey and one of chicken and dessert an orange and two apple pies

   To recognize these sentences we have developed the following grammar:

```
command --> order, comma, order, and, order.

order --> no, position.
order --> position, food.
order --> position, food, morefood.

morefood --> and, food.
morefood --> comma, food, morefood.

food --> number, namefood.

namefood --> of, namedish.
namefood --> namedish.
```

```
and --> [and].
no --> [no]
number --> [X], {member(X,[one, a, an, two, three, four, five])}.
```

and the following lexicon:

```
pos(starters).
pos(main).
pos(dessert).
dish(soup).
dish(soups).
dish(salad).
dish(salads).
dish(beef).
dish(chicken).
dish(turkey).
dish(salmon).
dish(orange).
dish(oranges).
dish(apple,pie).
dish(apple,pies).
dish(chocolate,cake).
dish(chocolate,cakes).
```

(a) Add to the grammar the rules corresponding to the terminals that are needed for the grammar to recognize correctly the sentences

(b) Add to the grammar the necessary rules so a food order is only recognized if it is given in its natural order, first starters, then main and finally dessert.

(c) Modify the grammar so there is agreement in number between the quantity ordered of a dish and the name of the dish.

(d) Modify the grammar so the dishes correspond with the position in the food order, so for instance it can not be ordered a salad as main or dessert, etc.

(e) Modify the grammar to generate as output the predicate `kitchen(X,Y,Z)` where each position has a list with instances of the predicate `dsh(X,Y)` where `X` is the number of orders of a dish and `Y` is the dish that is ordered.

It is not necessary to copy all the grammar in each question, only specify the rules that have to be modified.

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