Association Rules

What is association mining?
  n Ex:
  n If A and B then C
  n If A and not B then C
  n If A and B and C then D etc.

Support & Confidence

Support is defined as the minimum percentage of transactions in the DB containing A and B.

Confidence is defined as the minimum percentage of those transactions containing A that also contain B.

Ex. Suppose the DB contains 1 million transactions andt that 10'000 of those transactions contain both A and B.

We can then say that the support of the association if A then B is:
Supp= 10'000/1'000'000 = 1%.

Likewise, if 50'000 of the transactions contain A and 10'000 out of those 50'000 also contain B then the association rule if A then B has a confidence 10'000/50'000 = 20%.

Confidence is just the conditional probability of B given A.

R: LS ==> RS

Supp(R) **=** supp(LS $\cup$ RS)
 = # Transaction verifying R / (Total # of Transaction)
Conf(R) = supp(LS $\cup$ RS)/supp(LS)

Ex:
R: Milk=> Eggs,
A support(R) of 0.8 means in 80% of transaktion Milk and eggs are together.

The confidence means the correlation, the relation between the LS and the RS.

Exercice 1: Association Mining Based on the following data, find out the support and confidence of the rule : Farine => Sucre

Ticket 1	Ticket 2	Ticket 3	Ticket 4
Farine	Oeufs	Farine	Oeufs
Sucre	Sucre	Oeufs	Chocolat
Lait	Chocolat	Sucre	Thé

Solution:
Farine => Sucre has a **confidence** of 100%, this is the force
of the association and a support of 2/3. <==> number of association farine => Sucre divided by number of ticket where sucre or farine exist.

Exercice 2

Repeat the data used in exercice 1 a, by using the option statistics in DLV to find the support and confidence?

Exercice 3

FishLastWee	SaltConsump	Smoking	DrinkPattern	Gender	Hypertension
1	O	O	1	1	1
O	1	O	1	1	O
1	1	1	1	1	1
O	O	1	O	1	1
O	O	O	1	1	1
O	O	O	1	1	O
1	1	O	O	O	O
1	O	1	O	O	1
O	O	1	1	1	1
1	1	O	O	O	1
O	O	O	O	O	1
1	1	O	1	1	1
O	O	1	1	O	1
1	1	1	1	1	1
O	O	O	1	1	1
1	1	O	1	1	1
1	O	O	1	1	1
O	O	O	O	O	1
1	1	O	1	1	O
O	O	1	1	1	1
1	O	O	O	O	O
1	O	O	1	1	1
1	1	O	1	1	1
O	O	O	1	1	O
O	1	O	1	1	O
1	O	1	1	1	1
1	1	O	1	1	1
O	O	1	O	1	1
1	O	O	O	O	1

What is Support and what is Confidence, having the following Rule:

Rule: *FishlastWeek=1, Hypertension=1.*

$$<==>$$

FishlastWeek => Hypertension.

Calculate the support the <u>Support</u> and the <u>Confidence</u> of this Rule?