Implement and demonstrate the FIND-S algorithm for finding the most specific hypothesis based on a given set of training data samples. Read the training data from a .CSV file

In [23]:
```
import pandas as pd
import numpy as np

data=np.array(pd.read_csv('enjoysport.csv',header=None))
n = 6

print("\n The initial hypothesis is : ")
hypothesis = ['0']*n
print(hypothesis)

for i in range(n):
    hypothesis[i] = data[0][i]

for i in range(len(data)):
    if data[i][n] == 'Yes':
        for j in range(n):
            if hypothesis[j] != data[i][j]:
                hypothesis[j] = '?'
    print('For Traing Instance {0} the Hypothesis is : {1}'.format(i,hypothesis))

print("\n The Maximally specific hypothesis for the training instance is ")
print(hypothesis)
```

```
 The initial hypothesis is :
['0', '0', '0', '0', '0', '0']
For Traing Instance 0 the Hypothesis is : ['Sunny', 'Warm', 'Normal', 'Strong', 'Warm', 'Same']
For Traing Instance 1 the Hypothesis is : ['Sunny', 'Warm', '?', 'Strong', 'Warm', 'Same']
For Traing Instance 2 the Hypothesis is : ['Sunny', 'Warm', '?', 'Strong', 'Warm', 'Same']
For Traing Instance 3 the Hypothesis is : ['Sunny', 'Warm', '?', 'Strong', '?', '?']

 The Maximally specific hypothesis for the training instance is
['Sunny', 'Warm', '?', 'Strong', '?', '?']
```