**Nome e Cognome:**

**Matricola:**

## Ricerca dell'Informazione nel Web

**Compito di esame**, *tempo a disposizione: 90 minuti*
5 punti/problema

## Problema 1

1. How should the Boolean query `x AND y AND NOT z` be handled? Why is naive evaluation of this query normally very expensive? Write out a postings merge algorithm that evaluates this query efficiently.
2. Describe what structure we need to be able to answer phrase queries such as: "In un piatto poco cupo, poco pepe cape." Give an example of such an index by constructing some sample documents and presenting the corresponding index.

## Problema 2

1. Describe what tiered indexes are, how they work, and give an example of a 3-tier system by creating some sample documents and showing the corresponding indexes.
2. Describe what is cluster prunning and how it works. How does it help us at query time? What are the dangers of using it?

## Problema 3

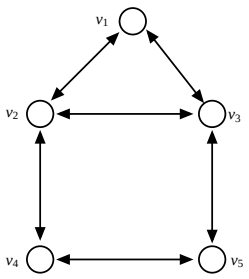
The following list of R's and N's represents relevant (R) and nonrelevant (N) returned documents in a ranked list of a collection of 30 documents. The top of the ranked list (the document the system thinks is most likely to be relevant) is on the left of the list.

R R N R N  N R R N N  N R N N N  N R N N R  N N N N N  N R N R N

1. What is the precision of the system on the top 20?
2. What is the recall on the top 20?
3. What is the $F_1$ measure on the top 20?
4. Draw the precision-recall curve.
5. What is the interpolated precision at 33% recall?

## Problema 4

1. We are given the following graph. Compute the pagerank score of each node for teleporting probability $\alpha = 0$.
2. Compute the pagerank score of each node for $\alpha = 1/2$.



**Acconsento alla pubblicazione dei risultati sul Web**

**Nome e Cognome in stampatello.....**

**Firma**