

**Vector Model Review Questions
(and Springboard for Discussion)**

1. Should a document containing 120 sentences of useful advice about child-proofing have a higher or lower ranking than one with 12 sentences?

How is this handled by the vector model as described in the textbook?

2. Suppose there is a document which contains only the text “the best way to get rich quick”. Based on the vector model (as seen so far), is this guaranteed to be the top-ranked document for the query “the best way to get rich quick”?

Why or why not?

3. Suppose that website **a** has the top ranking for the lucrative query “non-conductive hardened mil spec grommets”, and that website **b** has the second ranking. Now suppose that the webmaster of site **b** has hired you to increase the ranking of his site. You are unable to alter either site **a** or site **b**, but you have the resources to create millions of web documents containing anything you like. Assuming that the target search engine uses the pure vector model, what can you do?

Is this legal? Is this ethical? Why or why not?

4. It is sometimes said that google queries are “AND” by default, since google typically returns only documents in which *all* of the terms in the query are present. Does this mean that google is really doing Boolean queries? Or is it really using the “extended Boolean” model?