

Seminar 1

Exercise 1/1

Recommend a query processing strategy for (*tangerine* OR *trees*) AND (*marmalade* OR *skies*) AND (*kaleidoscope* OR *eyes*) with respect to the following postings list sizes:

eyes 213312

kaleidoscope 87009

marmalade 107913

skies 271658

tangerine 46653

trees 316812

Exercise 1/2

What is the best order for processing the query *ostrich* AND *hippo* AND *giraffe* if we know that the number of occurrences of the animals are 100, 500, 300, respectively?

Exercise 1/3

Create an inverted index composed of the following collection of documents:

Doc 1: new home sales top forecasts

Doc 2: home sales rise in July

Doc 3: increase in home sales in July

Doc 4: July new home sales rise

Exercise 1/4

Create an inverted index composed of the following collection of documents:

Doc 1: hippo ostrich ostrich giraffe

Doc 2: lion frog giraffe hippo

Doc 3: ostrich frog bat giraffe lion frog
