1. Why are there both forward indexes and inverted indexes?  What does each do?

Forward indexes let you find, or contain a list of words by document, like if you were scanning and finding the words that were included in the document to understand it, summarize it, vectorize it, etc. Inverse indexes actually sort the websites based on how words appear in the documents, like if you were querying, invert indexes allow you to find the document that map to words you input.

1. Why does anchor text go into the forward index?

This way I think we can perform page rank by looking as to what pages are pointed to by other pages. So, under the Google strategy we store in some way the URLs of the websites of which any given document points to, this also helps to crawl because we just have to follow the anchor text between websites, and we can continue traversing the web.

1. Why is a "hit list" a list of occurrences, rather than just the number of times the word occurred?

Because we want to distinguish between the relevance of the words found in the document, for example we want to give high weight to something that is clearly a little or an anchor which will surely give a hint that there might be a strong match. It is easy to see how if a website has as title “Computer Science” (like the Wikipedia article about Computer Science) will have more importance that a website that just mentions Computer Science.

1. At runtime (when a query comes in), which is consulted first, the inverted index or the forward index?  Or are they consulted simultaneously?

My understanding is that inverted index will be consulted first, to see how our input query maps to websites of our interest. As the paper mentions “We chose a compromise between these options, keeping two sets of inverted barrels -- one set for hit lists which include title or anchor hits and another set for all hit lists. This way, we check the first set of barrels first and if there are not enough matches within those barrels, we check the larger ones.” So, the issue of answering queries is one that deals with how you merge the “preferred” data for a word in the query like anchor text and titles with the preferred data among all the other words.

1. Of the following "type weights", which do you think is probably smallest?  {title, anchor, URL, plain text large font, plain text small font}

Because of the reading I am inclined to say anything that is a large font and small font, anything that is simple text.

1. Given the query "pets on step" and two documents, "step on pets" and "pets on step", how did Google in 1998 rank the first over the second, or couldn't it?