# 第8章 怎样研究算法:排序算法示例

1、排序算法是最基本的算法，很多复杂算法都是以排序为基础进行构造的。关于排序算法，下列说法不正确的是\_\_\_\_\_。

(A)大规模数据集合中查找有无某些元素的问题，有序数据集合比无序数据集合的查找要快得多；

(B)大规模数据集合中按元素分组进行计算的问题，有序数据集合比无序数据集合的计算要快得多；

(C)对无序数据集合，两个算法 X和Y：X采用无序数据处理，Y采用先将无序数据排序成有序数据，然后进行处理；则对前述(A)、(B)两类问题，Y算法一定比X算法慢；

(D)上述说法有不正确的；

答案：C

解释：

本题考核排序算法的研究

在大规模数据集合中查找，有序数据集合有利算法进行和判断，要比无序数据集合查找的快，对于（C）选项，Y算法尽管需要排序后再处理，但排序处理后的数据查找更加快捷，因此可能Y算法比X算法更快。

具体内容请参考排序算法以及第八章课件。

2、下列三个算法是关于“大规模数据集合中查找有无某些元素”问题的算法：针对一个“学生”数据表，如下示意，找出“成绩”为某一分数的所有学生。

学生

|  |  |  |
| --- | --- | --- |
| 学号 | 姓名 | 成绩 |
| 120300107 | 闫宁 | 95 |
| 120300103 | 李宁 | 94 |
| 120300101 | 李鹏 | 88 |
| 120300106 | 徐月 | 85 |
| 120300102 | 王刚 | 79 |
| 120300109 | 江海 | 77 |
| 120300110 | 周峰 | 73 |
| 120300104 | 赵凯 | 69 |
| 120300105 | 张伟 | 66 |
| 120300108 | 杜岩 | 44 |

**【算法A1】**

Start of algorithm A1

Step 1. 从数据表的第1条记录开始，直到其最后一条记录为止，读取每一条记录，做Step 2。

Step 2. 对每一条记录，判断成绩是否等于给定的分数：如果是，则输出；如果不是，则不输出。

End of algorithm A1

**【算法A2】**

Start of algorithm A2

Step 1. 从数据表的第1条记录开始，直到其最后一条记录为止，读取每一条记录，做Step 2和Step 3。

Step 2. 对每一条记录，判断成绩是否等于给定的分数：如果等于，则输出；如果不等于，则不输出。

Step 3. 判断该条记录的成绩是否小于给定的分数：如果不是，则继续；否则，退出循环，算法结束。

End of algorithm A2

**【算法A3】**

Start of algorithm A3

Step 1. 假设数据表的最大记录数是n，待查询区间的起始记录位置Start为1，终止记录位置Finish为n；

Step 2. 计算中间记录位置I = (Start+Finish)/2，读取第I条记录。

Step 3. 判断第I条记录的成绩与给定查找分数：

(3.1)如果是小于关系，则调整Finish = I-1；如果Start >Finish则结束，否则继续做Step 2；(3.2)如果是大于关系，则调整Start = I+1；如果Start>Finish则结束，否则继续做Step 2；

(3.3)如果是等于关系，则输出，继续读取I周围所有的成绩与给定查找条件相等的记录并输出，直到所有相等记录查询输出完毕则算法结束。

End of algorithm A3

针对上述三个算法，回答下列问题：

(1)关于算法A1, A2, A3的快慢问题，下列说法正确的是\_\_\_\_\_。

(A)算法A1快于算法A2， 算法A2快于算法A3；

(B)算法A2快于算法A1， 算法A2快于算法A3；

(C)算法A3快于算法A2， 算法A2快于算法A1；

(D)算法A1快于算法A3， 算法A3快于算法A2；

(E)上述都不正确。

答案：C

解释：

本题考核排序算法的研究

首先，数据是有序排列的， 从大到小。

算法A1依次搜索，穷举。

算法A2与A1一样，穷举，不同的是它利用数据是从大到小排序的特点，因此，如果当前数据比如果小于目标数，那么说明只有的也一定小于，则目标不在序列中。因此，A2比A1快。

算法A3利用数据有序特点，采用二分查找，每次将目标数与中间值比较，缩小搜索范围，因此A3比A2快。

综上，答案选（C）。

具体内容请参考排序算法以及第八章课件。

(2)关于算法A3，下列说法正确的是\_\_\_\_\_。

(A)对数据表中的任何数据，算法A3都适用；

(B)对数据表中任何已排序的数据，算法A3都适用；

(C)对已按成绩排序的数据表，算法A3都适用；

(D)对已按成绩进行降序排列的数据表，算法A3都适用；

(E)上述都不正确。

答案：D

解释：

本题考核排序算法的研究

算法A3需求的数据应该是排序的，而且是对成绩排序的，因此排除（A）（B）。其次，按照算法A3，应该是降序排序的数据才可用，因为升降序决定了算法二分查找的重订搜索范围实在中间值的左边还是右边，对于算法A3,要求是降序的。因此选择（D）。

具体内容请参考排序算法以及第八章课件。

(3)关于算法A3和算法A1，下列说法正确的是\_\_\_\_\_。

(A)如果数据表中记录数越多，则算法A3相比算法A1的优势越明显，即查找时间越短；

(B)如果数据表中记录数越多，则算法A1相比算法A3的优势越明显；即查找时间越短；

(C)算法A3和算法A1的执行时间差异不会随数据表中记录数多少而变化；

(D)上述都不正确。

答案：A

解释：

本题考核排序算法的研究

数据越多，算法A1穷举，查找时间正比增加，算法A3二分查找，每次可缩小一半的查找范围，数据越多，优势越明显，算法A1时间复杂度是O(n), 算法A3时间复杂度是O(log2n)，因此选择（A）。

具体内容请参考排序算法以及第八章课件。

(4)关于三个算法的复杂性，下列说法正确的是\_\_\_\_\_。

(A)算法A1、A2和A3的时间复杂性都为O(n)；

(B)算法A1和A2的时间复杂性为O(1)，算法A3的时间复杂性为O(n)；

(C)算法A1的时间复杂性为O(n)，算法A2的时间复杂性为O(n/2)，算法A3的时间复杂性为O(n/4)；

(D)算法A1和A2的时间复杂性为O(n)，算法A3的时间复杂性为O(log2 n)；

(E)上述都不正确。

答案：D

解释：

本题考核排序算法的研究

数据越多，算法A1穷举，线性查找，平均比较次数(n+1)/2，时间复杂度O(n)。算法A2同样是线性查找，只是多了对剩余数据的判断，时间复杂度和A1一样，同样是O(n)。算法A3二分查找，每次可缩小一半的查找范围，算法A3时间复杂度是O(log2n)，因此选择（D）。

具体内容请参考排序算法以及第八章课件。

(5)针对按成绩降序排列的数据表，假设记录数为n，关于算法A2，下列说法正确的是\_\_\_\_\_。

(A)算法A2在任何情况下都需要读取n条记录，才能得到结果；

(B)算法A2在任何情况下都需要读取n/2条记录，才能得到结果；

(C)算法A2在最好的情况下是读取1条记录，在最差的情况是读取n条记录，才能得到结果；

(D)算法A2在任何数据分布情况下，平均要读取n/2条记录才能得到结果；

(E)上述都不正确。

答案：C

解释：

本题考核排序算法的研究

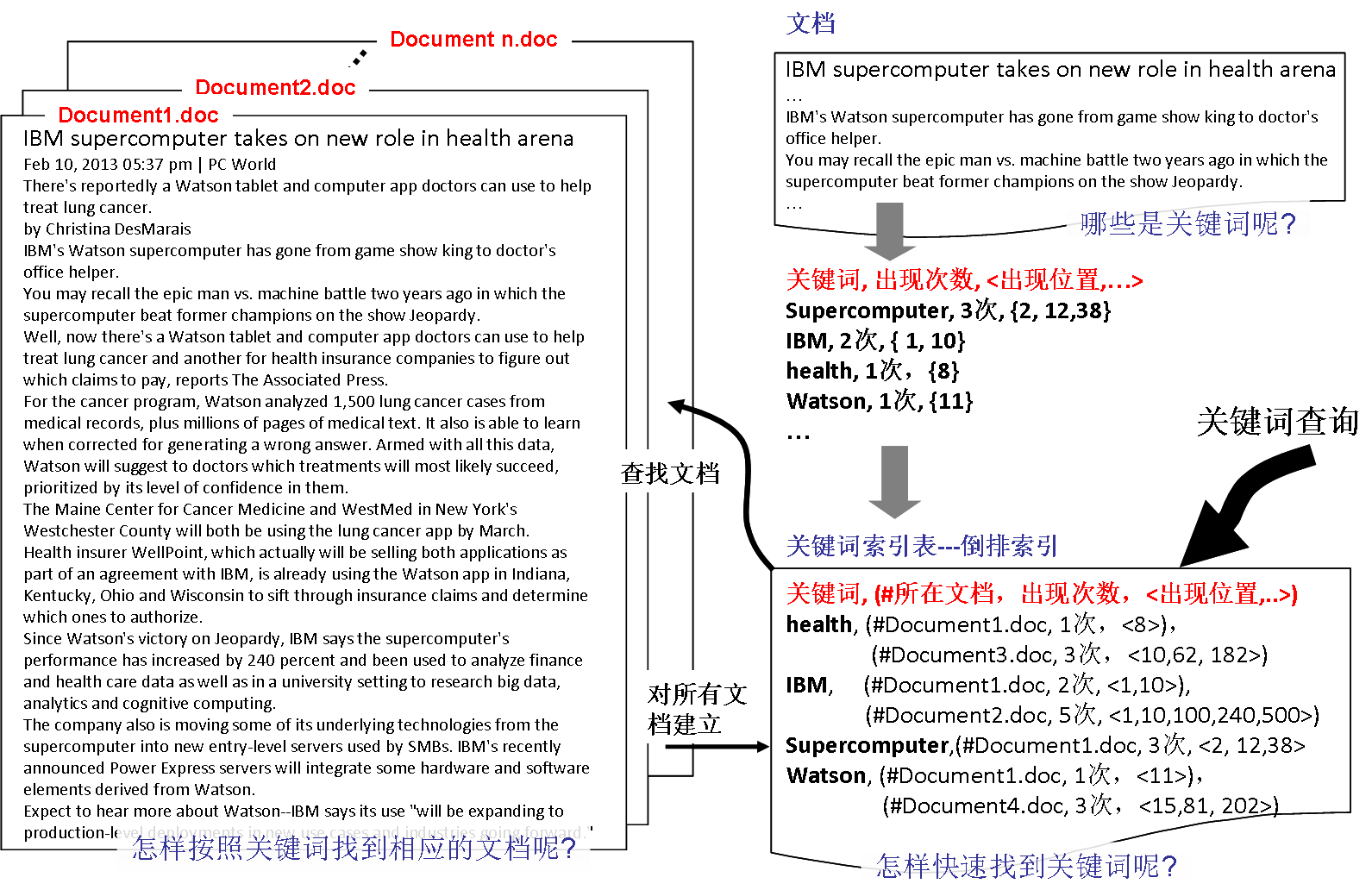
对于算法A2，最好情况：读取1条记录，刚好等于目标数，算法结束；或者读取1条记录，该记录小于目标数，因成绩降序排列，剩余数据都小于目标，所以记录中没有目标查询数，算法结束。

最坏情况：读取了n条记录，最后一条记录等于，则返回结果，不等于，则不在记录中。算法结束。

综上，选择（C）。

具体内容请参考排序算法以及第八章课件。

3、关于“非结构化数据(文档)的查找与搜索”问题，参考下图，回答下列问题。注意每份文档可能包含数千数万的词汇。



(1)若要在n个全文文档中(n可能很大)查找有无某个关键词的文档，为提高检索效率，最好的做法是\_\_\_\_\_。

(A)直接用给定关键词来匹配每一份文档中的每一个词汇。若该文档存在匹配成功的词汇，则输出该文档；否则，不输出该文档。

(B)对这n个文档，首先建立一个“关键词”索引表，该索引表记录着“关键词”及包含该关键词的“文档编号”。在此基础上，用给定关键词来匹配索引表中的关键词。如果匹配成功，则输出索引表中相对应的文档编号；否则，则输出信息“没有含该关键词的文档”。

(C)对这n个文档，首先建立一个“关键词”索引表，该索引表记录着“关键词”及包含该关键词的“文档编号”，并按关键词进行字母序的排序。在此基础上，用给定关键词来匹配索引表中的关键词。如果匹配成功，则输出索引表中相对应的文档编号，否则，则输出信息“没有含该关键词的文档”。

(D)选项(B)(C)比选项(A)的做法好，但选项(B)(C)没有效率上的差别。

答案：C

解释：

本题考核排序算法的研究

由题意知，n个全文文档数据量很大，若全文搜索，检索效率会很低，因此排除（A）。建立“关键字”索引表，并包含“文档编号”，可以有效的缩小查找范围。同时，对索引表进行排序，可以有效的提高查找效率。因此，选择（C）。

具体内容请参考查找，排序算法以及第八章课件。

(2)若要在n个全文文档中(n可能很大)查找与某个关键词最相关的文档，为提高检索效果和检索效率，最好的做法是\_\_\_\_\_。

(A)对这n个文档，首先建立一个“关键词”索引表，该索引表记录着“关键词”及包含该关键词的“文档编号”，并按关键词进行字母序的排序。在此基础上，用给定关键词来匹配索引表中的关键词。如果匹配成功，则输出索引表中相对应的文档编号，否则，则输出信息“没有含该关键词的文档”。

(B)对这n个文档，首先建立一个“关键词”索引表，该索引表记录着“关键词”，包含该关键词的“文档编号”，以及该关键词在该文档中出现的“次数”，并按关键词进行字母序的排序。在此基础上，用给定关键词来匹配索引表中的关键词。如果匹配成功，则进一步寻找同一关键词“次数”最多的m个索引项，输出相对应的文档编号；否则，则输出信息“没有含该关键词的文档”。

(C)对这n个文档，首先建立一个“关键词”索引表，该索引表记录着“关键词”，包含该关键词的“文档编号”，以及该关键词在该文档中出现的“次数”；对索引表，按关键词进行字母序的排序；如果关键词相同，则进一步按“次数”对同一关键词的若干文档进行降序排序。在此基础上，用给定关键词来匹配索引表中的关键词。如果匹配成功，则进一步寻找同一关键词“次数”最多的m个索引项，输出相对应的文档编号；否则，则输出信息“没有含该关键词的文档”。

(D)选项(B)(C)比选项(A)的做法好，但选项(B)(C)在执行效果和执行效率方面没有什么差别。

答案：C

解释：

本题考核排序算法的研究

由题意知，要寻找与关键词最相关的文档，做法是以关键词出现次数为评判标准。建立索引表，包含“关键词”，因要查找与关键词最相关的文档，所以需求相关文档中出现的“次数”，故而排除（A）选项，对于（B）（C），（C）选项中对关键词相同的文档进一步按“次数”进行降序排序，可以提高检索效率，因此选择（C）选项。

具体内容请参考查找，排序算法以及第八章课件。

(3)针对下列问题求解方法：对n个文档，首先建立一个“关键词”索引表，该索引表记录着“关键词”，包含该关键词的“文档编号”，以及该关键词在该文档中出现的“次数”；对索引表，按关键词进行字母序的排序；如果关键词相同，则进一步按“次数”对同一关键词的若干文档进行降序排序。在此基础上，用给定关键词来匹配索引表中的关键词。如果匹配成功，则进一步寻找次数最多的m个索引项，输出相对应的文档编号；否则，则输出信息“没有含该关键词的文档”。问该方法涉及到几类算法，说法正确的是\_\_\_\_\_。

(A)涉及字符串的字母序排序算法；

(B)涉及数值属性排序算法；

(C)涉及字符串匹配算法；

(D)涉及数值属性查找算法；

(E)涉及上述全部算法；

答案：E

解释：

本题考核排序算法的研究

查找关键词，设计字符串匹配算法。对索引表按关键字进行字母序的排序，设计字符串的字母序排序算法。关键词相同，则进一步按“次数”对同一关键词的若干文档进行降序排序，涉及数值属性排序算法。寻找次数最多的m个索引项，涉及数值属性查找算法。综上，选择（E）。

具体内容请参考查找算法以及第八章课件。

(4)上图给出了一种“自动获取文档关键词”的方法，关于该方法的表述，最好的是\_\_\_\_\_。

(A)文档中出现次数最多的词汇必定是关键词；

(B)文档中去掉标点符号后，出现次数最多的词汇必定是关键词；

(C)文档中去掉标点符号和一些辅助词汇， 出现次数最多的词汇必定是关键词；

(D)文档中去掉标点符号和一些辅助词汇， 出现次数最多且次数达到一定数值的词汇必定是关键词；

(E)上述说法都不正确；

答案：D

解释：

本题考核排序算法的研究

“自动获取文档关键词”的方法是一个充分不必要条件。

关键词应排除标点符号和辅助词汇，因此排除（A）（B）。出现次数最多，但不能对单个文档而言就凭次数判断关键词，必须达到一定的标准，故而选择（D）。

具体内容请参考排序算法以及第八章课件。

4、关于“内排序”算法和“外排序”算法，下列说法不正确的是\_\_\_\_\_。

(A)“内排序”算法通常是内存中数据排序常用的算法，而“外排序”算法通常是大规模数据排序常用的算法；

(B)“内排序”算法由于内存排序应用的频繁性，所以算法要考虑用尽可能少的步骤，而“外排序”算法由于要利用磁盘保存中间结果，所以算法主要考虑尽可能少的读写磁盘；

(C)无论是“内排序”算法，还是“外排序”算法，都需要考虑读写磁盘的代价问题；

(D)对一组需要排序的数据，能应用“内排序”算法时，尽量不用“外排序”算法；

答案：C

解释：

本题考核排序算法的研究

内排序：待排序的数据可一次性地装入内存中，即排序者可以完整地看到和操纵所有数据；

外排序：待排序的数据保存在磁盘上，不能一次性装入内存，即排序者不能一次完整地看到和操纵所有数据，需要将数据分批装入内存分批处理；

由上可知（A）（B）正确，（C）错误。

具体内容请参考排序算法以及第八章课件。

5、下列三种算法是经常应用的内排序算法：插入排序、选择排序和冒泡排序。阅读下列算法，回答下列问题。

**INSERTION-SORT(A)**

***1. for i=2 to N***

***2. { key = A[i] ;***

***3. j =i-1;***

***4. While (j>0 and A[j]>key) do***

***5. { A[j+1]=A[j];***

***6. j=j-1; }***

***7. A[j+1]=key;***

***8. }***

***SELECTION-SORT(A)***

***1. for i=1 to N-1***

***2. { k=i;***

***3． for j=i+1 to N***

***4. { if A[j]<A[k] then k=j; }***

***5. if k<>i then***

***6. {***

***7. temp =A[k];***

***8. A[k]=A[i];***

***9. A[i]=temp;***

***10. }***

***11. }***

***BUBBLE-SORT(A)***

***1. for i=1 to N-1***

***2. { haschange=false;***

***3. for j=1 to N-i***

***4. { if A[j]>A[j+1] then***

***5. { temp =A[j];***

***6. A[j]=A[j+1];***

***7. A[j+1]=temp;***

***8. haschange=true;***

***9. }***

***10. }***

***11. if (haschange ==false) then break;***

***12. }***

(1)关于INSERTION-SORT算法的基本思想，下列说法正确的是\_\_\_\_\_。

(A)一个元素一个元素的处理。每次处理一个元素，通过与当前已排序元素的比较，将该元素放入到当前正确排序的位置。直到最后一个元素则算法结束。

(B)一个轮次一个轮次的处理。将元素集合分成两个部分，已排序元素集合和未排序元素集合，开始时已排序元素集合为空。在每一轮次，从未排序元素集合中找出最小值的元素，将其移入已排序元素集合；直到未排序元素集合为空时则算法结束。

(C)一个轮次一个轮次的处理。在每一轮次中依次对待排序数组元素中相邻的两个元素进行比较：如不符合排序关系，则交换两个元素。直到某一轮次没有元素交换发生则结束。

(D)上述说法都不正确。

答案：A

解释：

本题考核排序算法的研究

（A）描述的是插入排序，（B）描述的是选择排序，（C）描述的是冒泡排序。

具体内容请参考内排序算法以及第八章课件。

(2)关于SELECTION-SORT算法的基本思想，下列说法正确的是\_\_\_\_\_。

(A)一个元素一个元素的处理。每次处理一个元素，通过与当前已排序元素的比较，将该元素放入到当前正确排序的位置。直到最后一个元素则算法结束。

(B)一个轮次一个轮次的处理。将元素集合分成两个部分，已排序元素集合和未排序元素集合，开始时已排序元素集合为空。在每一轮次，从未排序元素集合中找出最小值的元素，将其移入已排序元素集合；直到未排序元素集合为空时则算法结束。

(C)一个轮次一个轮次的处理。在每一轮次中依次对待排序数组元素中相邻的两个元素进行比较：如不符合排序关系，则交换两个元素。直到某一轮次没有元素交换发生则结束。

(D)上述说法都不正确。

答案：B

解释：

本题考核排序算法的研究

（A）描述的是插入排序，（B）描述的是选择排序，（C）描述的是冒泡排序。

具体内容请参考内排序算法以及第八章课件。

(3)关于BUBBLE-SORT算法的基本思想，下列说法正确的是\_\_\_\_\_。

(A)一个元素一个元素的处理。每次处理一个元素，通过与当前已排序元素的比较，将该元素放入到当前正确排序的位置。直到最后一个元素则算法结束。

(B)一个轮次一个轮次的处理。将元素集合分成两个部分，已排序元素集合和未排序元素集合，开始时已排序元素集合为空。在每一轮次，从未排序元素集合中找出最小值的元素，将其移入已排序元素集合；直到未排序元素集合为空时则算法结束。

(C)一个轮次一个轮次的处理。在每一轮次中依次对待排序数组元素中相邻的两个元素进行比较：如不符合排序关系，则交换两个元素。直到某一轮次没有元素交换发生则结束。

(D)上述说法都不正确。

答案：C

解释：

本题考核排序算法的研究

（A）描述的是插入排序，（B）描述的是选择排序，（C）描述的是冒泡排序。

具体内容请参考内排序算法以及第八章课件。

(4)关于排序的选择法和冒泡法，下列说法不正确的是\_\_\_\_\_。

(A)“选择法”和“冒泡法”都是每一轮次找出一个最小值元素，只是寻找最小值元素的方法不一样，在效率方面没有什么差别；

(B)“选择法”通过将所有未排序元素与当前轮次待寻找的最小值元素进行比较，获得当前轮次的最小值元素；而“冒泡法”通过相邻元素的两两比较，一个轮次完成也能获得一个最小值元素；

(C)虽然“选择法”和“冒泡法”都是每一轮次找出一个最小值元素，但选择法每轮次仅比较，没有交换，直至找到最小值后做一次交换；而冒泡法每一轮次是通过相邻元素比较来找最小值，如果不满足排序，则交换相邻两个元素，交换可能频繁发生。这样来看，选择法比冒泡法要快一些；

(D)上述说法有不正确的。

答案：A

解释：

本题考核排序算法的研究

“选择”与“冒泡”都是每轮查找出最小值，方法不同（B）选项是正确的，但效率是有不同的，（C）中描述正确，选择排序调用swap要比冒泡排序少，因此，选择（A）

具体内容请参考内排序算法以及第八章课件。

(5)关于三种排序算法，下列说法正确的是\_\_\_\_\_。

(A)三种算法的时间复杂度都为O(n2)，所以三种算法的执行效率是一样的；

(B)尽管三种算法的时间复杂度都为O(n2)，但细致比较还是有差别的，例如冒泡法排序比选择法排序要快一些；

(C)尽管细致比较三种算法的执行时间是有差别的，但这种差别对内排序问题而言是可以忽略不计的；

(D)尽管细致比较三种算法的执行时间是有差别的，这种差别对内排序问题而言是重要的，因为内排序算法可能要被频繁的执行。

答案：D

解释：

本题考核排序算法的研究

三种算法的时间复杂度都是O(n2)，但是细致比较是有差别的，（A）错误，如果被排序的记录很大，执行交换所需时间是客观的，将远远大于比较关键字和计算数组下标的时间。冒泡和插入排序调用次数最多，为n(n-1)/2，选择排序每遍只在选出最小记录后进行一次交换，需调用n-1次，因此（B）错误，有因为内排序算法要被频繁执行，因此这种差别对内排序问题而言是重要的，故而选择（D）。

具体内容请参考内排序算法以及第八章课件。

(6)阅读INSERTION-SORT算法，关于第4.行至第6.行间程序段的作用，下列说法正确的是\_\_\_\_\_。

(A)将当前待处理元素，依次与已经排序的第j个元素进行比较，j采取递减方式循环，以找到当前元素所应在的位置，并将该位置以前的元素依次向后移动一个位置；

(B)将当前待处理元素，依次与已经排序的第j个元素进行比较，j采取递增方式循环，以找到当前元素所应在的位置，并将该位置以前的元素依次向后移动一个位置。

(C)将当前待处理元素，依次与已经排序的第j个元素进行比较，j采取递增方式循环，以找到当前元素所应在的位置，并将该位置以后的元素依次向前移动一个位置。

(D)将当前待处理元素，依次与已经排序的第j个元素进行比较，j采取递减方式循环，以找到当前元素所应在的位置，并将该位置以后的元素依次向后移动一个位置。

答案：D

解释：

本题考核排序算法的研究

插入排序，算法中J从I-1到1，递减，排除（B）（C），将大于key的元素依次向后移动，因而选择（D）。

具体内容请参考内排序算法以及第八章课件。

(7)阅读SELECTION-SORT算法，关于第3.行至第4.行间程序段的作用，下列说法正确的是\_\_\_\_\_。

(A)循环地在未排序元素集合中找最小值元素的位置，该位置保存在变量k中；

(B)循环地在未排序元素集合中找最小值元素，该元素保存在变量k中；

(C)循环地在未排序元素集合中找最大值元素的位置，该位置保存在变量k中；

(D)循环地在未排序元素集合中找最大值元素，该元素保存在变量k中；

答案：A

解释：

本题考核排序算法的研究

选择排序，算法中k记录的是最小元素的位置，交换条件：A[j]<A[k]，因而是找出最小元素，所以选择（A）。

具体内容请参考内排序算法以及第八章课件。

(8)阅读BUBBLE-SORT算法，已知N=20，下列说法正确的是\_\_\_\_\_。

(A)第5轮次，是将第1个元素至第15个元素之间的元素，相邻者进行比较；

(B)第4轮次，是将第1个元素至第20个元素之间的元素，相邻者进行比较；

(C)第8轮次，是将第20个元素至第12个元素之间的元素，相邻者进行比较；

(D)第11轮次，是将第20个元素至第1个元素之间的元素，相邻者进行比较；

答案：A

解释：

本题考核排序算法的研究

冒泡排序，每遍找出最小元素，方法是依次将相邻元素两两比较，逐渐分成两个部分：已排和未排，N=20，第5次时，16至20号元素已排序完成，需要将是将第1个元素至第15个元素之间的元素，相邻者进行比较，选择（A）。

具体内容请参考内排序算法以及第八章课件。

(9)阅读BUBBLE-SORT算法，下列说法正确的是\_\_\_\_\_。

(A)该算法在N=20时，必定要执行20个轮次的内循环；

(B)该算法在N=20时，必定要执行19个轮次的内循环；

(C)该算法在N=20时，最多要执行20个轮次的内循环；

(D)该算法在N=20时，最多要执行19个轮次的内循环；

答案：D

解释：

本题考核排序算法的研究

冒泡排序，内循环是两两相邻元素依次比较，其中如果发生交换，即元素大小次序不满足要求，变量haschange为true，若haschange为false，则说明经过一次遍历之后，没有发生过相邻元素交换，即当前序列满足排序的要求，说明任务已完成，算法结束，不需要执行内循环。因此，当N=20时，i到N-1，最多需要执行19次内循环。因此选择（D）。

具体内容请参考内排序算法以及第八章课件。

(10)阅读BUBBLE-SORT算法，其中关于haschange变量的作用，下列说法不正确的是\_\_\_\_\_。

(A)haschange用于标记每个轮次的相邻元素比较中，是否有“交换”发生；

(B)haschange用于判断至某个轮次时是否已完成排序，以便提前结束算法；

(C)haschange需要在每轮次之前置初始值为假，表示没有“交换”发生；

(D)涉及haschange的语句全部去掉，算法仍旧能够正确执行；

(E)上述说法有不正确的。

答案：E

解释：

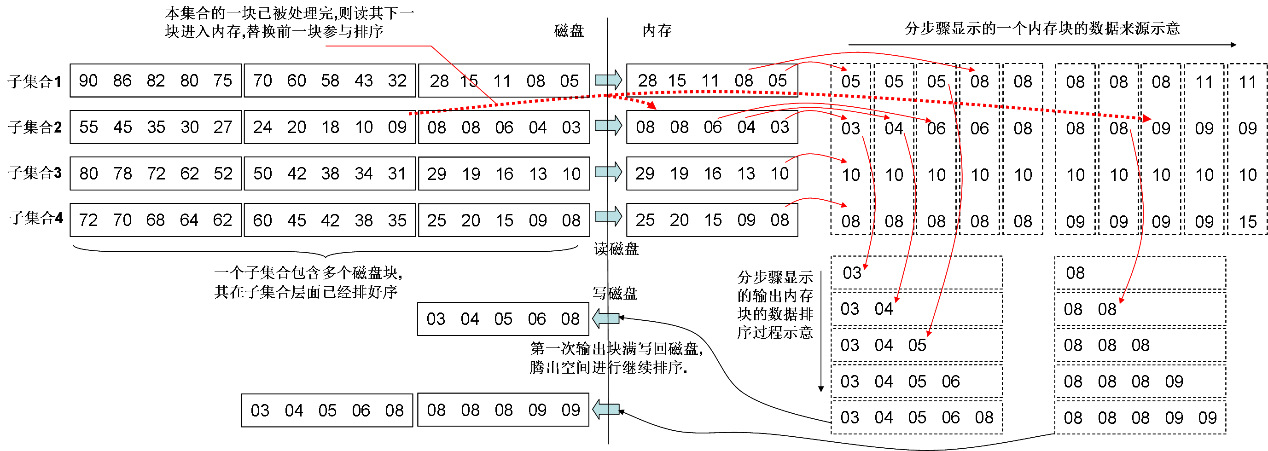
本题考核排序算法的研究

冒泡排序，内循环是两两相邻元素依次比较，其中如果发生交换，即元素大小次序不满足要求，变量haschange为true，若haschange为false，则说明经过一次遍历之后，没有发生过相邻元素交换，即当前序列满足排序的要求，说明任务已完成，算法可结束，如果将设计haschange的语句全部去掉，算法依旧能够执行，只是在已排序的记录上执行。

综上，选择（E）。

具体内容请参考内排序算法以及第八章课件。

6、外排序是**需要使用硬盘等外部存储设备进行大数据集合排序**的过程或算法，其中一种策略是“排序-归并”，如下图所示。仔细理解该图所表达的基本思想，回答下列问题。



(1)关于“排序-归并”算法，下列说法不正确的是\_\_\_\_\_。

(A)“排序-归并”算法是一个两阶段完成排序的算法，第一个阶段称为子集合排序，第二个阶段称为归并排序；

(B)“排序-归并”算法是在这样环境下应用的算法：待排序数据元素数目大于或远大于内存中可装入数据元素数目；

(C)“排序-归并”算法可以对任意大规模的数据集合进行排序；

(D)“排序-归并”算法是通过多次读写磁盘完成大规模数据集合的排序工作的；

(E)上述说法有不正确的。

答案：E

解释：

本题考核对“排序-归并”算法的理解。

(A)(B)(C) (D)的叙述都是正确的，所以选择(E)。

具体内容请参考第八章课件之“基本排序算法--外排序算法”。

(2)参见图示，内存块数为Bmemory=6，每块可装载Rblock=5个元素，如果经过一个轮次的归并操作便能完成排序，则关于待排序元素集合的大小，下列说法正确的是\_\_\_\_\_。

(A)待排序元素数目应 >= (Bmemory-2) \* Bmemory \* Rblock；

(B)待排序元素数目应 <= (Bmemory-2) \* Bmemory \* Rblock；

(C)待排序元素数目应 <= Bmemory \* Rblock；

(D)待排序元素数目应 >= Bmemory \* Rblock；

(E)上述说法都不正确。

答案：B

解释：

本题考核对排序-归并算法的理解。

首先，子集合排序时，将Bproblem块数据划分为N个子集合, 要求每个子集合的块数不超过内存可用块数（Bproblem）；其次，归并过程中，除了每个子集合中的一块要放到内存中，还要剩余2块，一块用于装载输出数据块，另一块用于存放待比较元素数据块，即子集合数不应该超过(Bmemory-2) ；所以待排序元素数目应 <= (Bmemory-2) \* Bmemory \* Rblock，(B)正确。

具体内容请参考第八章课件之“基本排序算法--外排序算法”。

(3)参见图示。如果：内存块数为Bmemory=6，每块可装载Rblock=5个元素，待排序元素集合所占用磁盘块数Bproblem=24，则关于此集合的排序问题，下列说法正确的是\_\_\_\_\_。

(A)首先将待排序元素集合划分为2个子集合，每个子集合为12块，将每个子集合从磁盘装入内存并采用任何内排序算法进行排序后再写回磁盘；然后再一个轮次对这2个已排序子集合进行归并操作，完成最终排序；

(B)首先将待排序元素集合划分为4个子集合，每个子集合为6块，将每个子集合从磁盘装入内存并采用任何内排序算法进行排序后再写回磁盘；然后再对这4个已排序子集合进行归并操作，完成最终排序；

(C)首先将待排序元素集合划分为6个子集合，每个子集合为4块，将每个子集合从磁盘装入内存并采用任何内排序算法进行排序后再写回磁盘；然后再对这6个已排序子集合进行一个轮次的归并操作，完成最终排序；

(D)前述(A)(B)(C)都正确。

答案：B

解释：

本题考核排序-归并算法的子集合划分问题。

首先，子集合排序时，将Bproblem块数据划分为N个子集合, 要求每个子集合的块数小于内存可用块数，即：Bproblem/N ≤Bmemory，(A)不符合这个要求；其次，归并过程中，除了每个子集合中的一块要放到内存中，还要剩余2块，一块用于装载输出数据块，另一块用于存放待比较元素数据块，(C)中把每个子集合中的元素放入内存中就没有剩余的内存了，所以错误，所以(B)正确。

具体内容请参考第八章课件之“基本排序算法--外排序算法”。

(4)参见图示。如果：内存块数为Bmemory=6，每块可装载Rblock=5个元素，待排序元素集合所占用磁盘块数Bproblem=24，进行升序排序，此集合已被划分为4个子集合并对每个子集合元素已进行升序排序并写回磁盘，则关于归并问题，下列说法不正确的是\_\_\_\_\_。

(A)内存共有6块，其使用分配如下：4块内存中的每一块分别用于装载4个子集合中的一块；剩余2块，一块用于装载输出数据块，另一块用于存放待比较元素数据块，该块中的元素分别来自于4个子集合中。

(B)待比较元素数据块中的最小者，被送到输出数据块中；同时，再从其对应的子集合数据块中依次补充进一个元素；

(C)当某子集合在内存的数据被处理完时，则再从磁盘上将该子集合的下一块读入到内存中，直到该子集合的所有块都已经被处理完为止；

(D)当输出数据块被装满时，则将输出数据块依次写回到磁盘上；

(E)上述说法有不正确的。

答案：E

解释：

本题考核排序-归并算法的归并过程。

(A)(B)(C) (D)的叙述都是正确的，所以选择(E)。

具体内容请参考第八章课件之“基本排序算法--外排序算法”。

(5)参见图示。如果：内存块数为Bmemory=6，每块可装载Rblock=5个元素，待排序元素集合所占用磁盘块数Bproblem=24，采用排序-归并算法进行升序排序，下列说法正确的是\_\_\_\_\_。

(A)算法以磁盘块读写次数衡量的时间复杂性为1Bproblem；

(B)算法以磁盘块读写次数衡量的时间复杂性为2Bproblem；

(C)算法以磁盘块读写次数衡量的时间复杂性为4Bproblem；

(D)算法以磁盘块读写次数衡量的时间复杂性为8Bproblem。

答案：C

解释：

本题考核排序-归并算法的时间复杂性。

4个子集合4次内排序---2个Bproblem； 1次4路归并最终完成---2个Bproblem，最终时间复杂性为4Bproblem。

具体内容请参考第八章课件之“基本排序算法--外排序算法”。

(\*6)参见图示。如果：内存块数为Bmemory=4，待排序元素集合所占用磁盘块数Bproblem=40，进行升序排序。如果：归并过程中，整体的数据集被从磁盘读入内存，再由内存写回磁盘，被称为一个轮次，则下列说法正确的是\_\_\_\_\_。

(A)该数据集可以经过1个轮次的2路归并完成最终排序；

(B)该数据集可以经过2个轮次的2路归并完成最终排序；

(C)该数据集可以经过3个轮次的2路归并完成最终排序；

(D)该数据集可以经过多于3个轮次的2路归并完成最终排序；

答案：D

解释：

本题考核排序-归并算法的归并过程。

10个子集合：5次2路归并形成5个集合--1个轮次；2次2路归并形成2个集合+1个集合—1个轮次；1次2路形成1个集合+1个集合--1个轮次；1次2路归并最终完成—1个轮次；所以该数据集可以经过多于3个轮次的2路归并完成最终排序，(D)正确。

具体内容请参考第八章课件之“基本排序算法--外排序算法”。

(\*7)参见图示。如果：内存块数为Bmemory=4，待排序元素集合所占用磁盘块数Bproblem=40，进行升序排序。如果：从磁盘装入内存，再从内存写回磁盘，被称为内存利用了一次，则下列说法正确的是\_\_\_\_\_。

(A)该数据集基于“排序-归并”策略完成最终排序，需要利用内存19次；

(B)该数据集基于“排序-归并”策略完成最终排序，需要利用内存9次；

(C)该数据集基于“排序-归并”策略完成最终排序，需要利用内存10次；

(D)该数据集基于“排序-归并”策略完成最终排序，需要利用内存5次；

答案：A

解释：

本题考核排序-归并算法对内存的利用。

10个子集合进行10次内排序—内存利用10次；5次2路归并形成5个集合—内存利用5次；2次2路归并形成2个集合+1个集合—内存利用2次；1次2路形成1个集合+1个集合—内存利用1次；1次2路归并最终完成—内存利用1次；所以总的内存利用次数是19次。

具体内容请参考第八章课件之“基本排序算法--外排序算法”。

(\*8)参见图示。如果：内存块数为Bmemory=4，待排序元素集合所占用磁盘块数Bproblem=40，采用排序-归并算法进行升序排序，下列说法正确的是\_\_\_\_\_。

(A)算法以磁盘块读写次数衡量的时间复杂性为4Bproblem；

(B)算法以磁盘块读写次数衡量的时间复杂性为8Bproblem；

(C)算法以磁盘块读写次数衡量的时间复杂性为10Bproblem；

(D)算法以磁盘块读写次数衡量的时间复杂性为19Bproblem。

答案：C

解释：

本题考核排序-归并算法的时间复杂性。

10个子集合10次内排序---2个Bproblem；5次2路归并形成5个集合---2个Bproblem；2次2路归并形成2个集合+1个集合---2个Bproblem；1次2路形成1个集合+1个集合---2个Bproblem；1次2路归并最终完成---2个Bproblem，最终时间复杂性为10Bproblem。

具体内容请参考第八章课件之“基本排序算法--外排序算法”。

(\*9)参见图示。如果：内存块数为Bmemory=8，待排序元素集合所占用磁盘块数Bproblem=80，首先，80个磁盘块的待排序元素集合被分成10个子集合，分别进行子集合排序；然后再进行归并处理完成最终排序。关于归并操作，几个子集合同时装入内存进行归并就被称为几路归并，则下列说法不正确的是\_\_\_\_\_。

(A)对10个已排序子集合可以先进行2个5路归并形成2个子集合，然后再进行1个2路归并便可完成最终的排序；

(B)对10个已排序子集合可以先进行3个3路归并形成3个子集合，外加剩余子集合共4个子集合，然后再进行1个4路归并便可完成最终的排序；

(C)对10个已排序子集合可以先进行1个5路归并形成1个子集合，外加剩余5个子集合共6个子集合，再进行1个6路归并便可完成最终的排序；

(D)前述(A)(B)(C)归并策略都可以，但性能有所不同，最好的是(A)策略。

答案：D

解释：

本题考核对归并策略性能的理解。

(A)(B)(C)的归并策略都可以，但是最好的是(C)策略，因为(C)仅有5个子集合进行了2个轮次的归并，5个子集合进行了一个轮次的归并；(A)所有10个子集合都进行了2个轮次的归并；所以(D)错误。

具体内容请参考第八章课件之“基本排序算法--外排序算法”。

(\*10)已知内存块数为Bmemory，待排序元素集合所占用磁盘块数Bproblem，设计一个“排序-归并”算法的基本思路，下列描述不正确的是\_\_\_\_\_。

(A)首先划分子集合，每个子集合最大可为Bmemory块，可以划分为Bproblem/Bmemory个子集合。这样划分的理由：一是子集合可以全部装载入内存执行内排序，二是最大限度地利用内存产生尽可能少数目的子集合；

(B)将Bmemory块内存留出两块，一块作为输出数据块，一块用于待比较元素数据块。其余Bmemory-2块用于装载尽可能多数目的子集合，即尽可能采用更多路的归并。这样做的理由：尽可能最大限度地利用内存，以便减少归并的次数；

(C)如果子集合参与归并一次被称为一个轮次，则整个数据集的轮次是指该数据集中参与归并次数最多的子集合的轮次。归并算法应考虑以尽可能少轮次的归并为目标来衡量各种不同归并策略的好坏。也可以定义一个参数“子集合轮次累积和”，即所有子集合参与归并轮次的总和，来衡量性能好坏，即“子集合轮次累积和”越小，算法性能越好；

(D)假设Bmemory=6，Bproblem=60，则按照上述(A)(B)(C)思想，可自动确定出：子集合数目=10，第一次将10个子集合分成3组(3个、3个和4个)并分别采用3路归并和4路归并将其归并成3个子集合；第二次对这3个集合再采用3路归并完成最终的排序。这样做的算法是最优的。

(E)假设Bmemory=6，Bproblem=60，则按照上述(A)(B)(C)思想，可自动确定出：子集合数目=10，第一次采用4路归并分别对8个子集合、4个一组归并成2个子集合；第二次对这2个集合与剩余的2个子集合一起采用4路归并完成最终的排序。这样做的算法是最优的；

答案：D

解释：

本题考核“排序-归并”算法的设计。

(A)(B)(C)的说法都是正确的，(E)相比(D)而言最优，因为E仅有8个子集合进行了2个轮次的归并，2个子集合进行了一个轮次的归并，而D所有10个子集合都进行了2个轮次的归并。从“子集合轮次累积和”角度(E)比(D)优，所以(D)错误。

具体内容请参考第八章课件之“基本排序算法--外排序算法”。

(11)关于内排序和外排序算法设计的关键点，下列说法不正确的是\_\_\_\_\_。

(A)外排序算法体现了受限资源环境下的算法构造，这里内存是一种受限资源；

(B)外排序算法强调尽可能少地读写磁盘，尽可能充分地利用内存来完成算法构造；

(C)外排序算法体现了与内排序算法设计不一样的关注点，前者更关注磁盘读写，后者更关注CPU执行操作的步数；

(D)外排序算法因内存环境的变化可以采用不同的策略，而不同策略算法的性能可能有所不同，这体现了问题求解算法的多样性，体现了算法需要“优化”；

(E)上述说法有不正确的。

答案：E

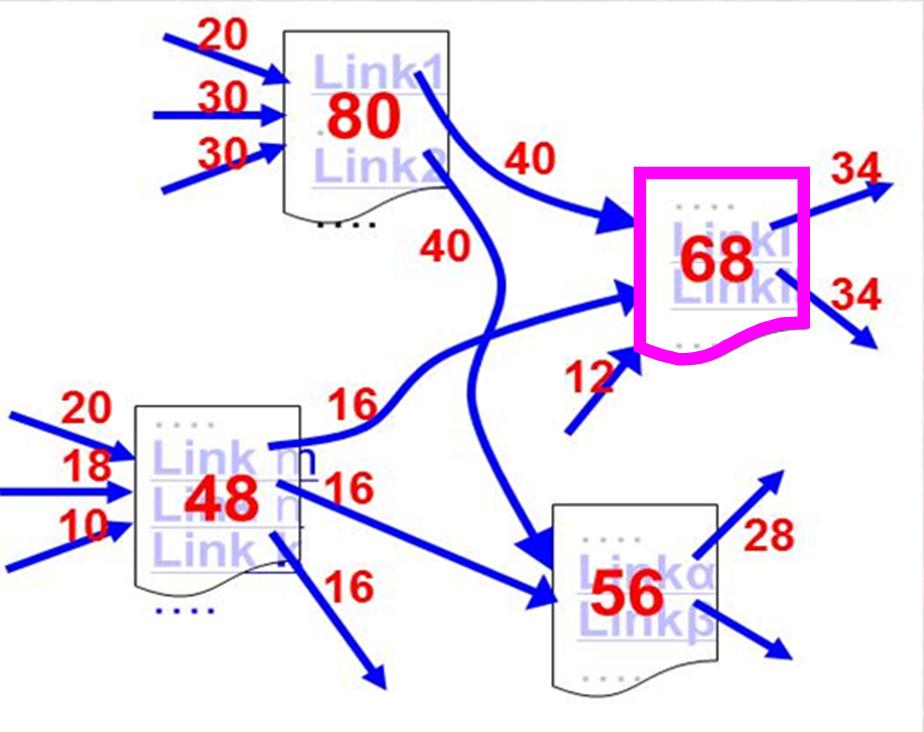
解释：

本题考核内排序和外排序的区别。

内排序是指待排序的数据可一次性地装入内存中，即排序者可以完整地看到和操纵所有数据，使用数组或其他数据结构便可进行统一的排序处理的排序问题；外排序是指待排序的数据保存在磁盘上，不能一次性装入内存，即排序者不能一次完整地看到和操纵所有数据，需要将数据分批装入内存分批处理的排序问题； (A)(B) (C)(D)的叙述都是正确的，所以(E)错误。

具体内容请参考第八章课件。

7、PageRank是Google公司提出的计算网页重要度的一种方法。参见下图，简单而言，网页是由“文本”和“链接”构成的，“链接”可使用户从一个网页跳转到另一个网页。因此，所谓“链接”即是某一个网页的地址，通过网页链接的读取，可以建立起各个网页之间的链接关系。对一个网页而言，其链接到其他网页的链接被称为“正向链接”，而所有链接到该网页的链接被称为“反向链接”。关于PageRank算法，回答下列问题。



图I.

(1)关于PageRank计算网页重要度的基本思想，下列说法正确的是\_\_\_\_\_。

(A)反向链接数越多的网页越重要----被链接次数越多越重要；

(B)反向链接加权和越高的网页越重要----被重要网页链接次数越多越重要；

(C)正向链接数越多的网页，其链接的权值越低----正向链接数越多的网页越不重要；

(D)上述全部。

答案：D

解释：

本题考查PageRank的基本思想--通俗语义。

一个网页的重要度等于其所有反向链接的加权和，所以反向链接加权和越高的网页越重要，反向链接数越多的网页越重要；一个正向链接的权值等于网页的重要度除以其正向链接数，所以正向链接数越多的网页，其链接的权值越低，所以(A)(B)(C)都正确，选(D)。

具体内容请参考第九章课件之“PageRank排序—排序问题的不同思考方法”。

(2-1)按照PageRank的思想，一个网页的重要度被定义为\_\_\_\_\_。

(A)其所拥有的所有反向链接的数目；

(B)其所拥有的所有正向链接的数目；

(C)其所拥有的所有链接的数目；

(D)上述都不正确。

答案：D

解释：

本题考查PageRank对网页重要度的定义。

一个网页的重要度等于其所有反向链接的加权和，所以(A)(B)(C)都不正确，选(D)。

具体内容请参考第九章课件之“PageRank排序—排序问题的不同思考方法”。

(2-2)按照PageRank的思想，一个网页的重要度被定义为\_\_\_\_\_。

(A)其所拥有的所有反向链接的数目；

(B)其所拥有的所有反向链接的加权和；

(C)其所拥有的所有正向链接的数目；

(D)其所拥有的所有正向链接的加权和；。

答案：B

解释：

本题考查PageRank对网页重要度的定义。

一个网页的重要度等于其所有反向链接的加权和，所以(B)正确。

具体内容请参考第九章课件之“PageRank排序—排序问题的不同思考方法”。

(3)按照PageRank的思想，一个网页链接的权值被定义为\_\_\_\_\_。

(A)网页重要度除以该网页所拥有的正向链接数；

(B)网页重要度除以该网页所拥有的反向链接数；

(C)网页重要度除以该网页所拥有的所有链接数；

(D)上述都不正确；。

答案：A

解释：

本题考查PageRank对链接权值的定义。

一个网页链接的权值等于网页的重要度除以其正向链接数，所以(A)正确。

具体内容请参考第九章课件之“PageRank排序—排序问题的不同思考方法”。

(4)PageRank将网页的链接关系，抽象为一个n × n的矩阵A：网页被从1到n进行编号；如果网页i有一个指向网页j的链接，则矩阵的aij元素(即第i行第j列元素)值为1，否则矩阵aij元素值为0。然后将A做一个转置处理(即矩阵的行列互换)，形成转置矩阵AT，为什么要转置，原因是\_\_\_\_\_。

(A)有利于体现反向链接的重要性；

(B)有利于更好地区分反向链接与正向链接；

(C)有利于计算权值矩阵(可被称为转移概率矩阵M)：将AT的一列中的各行除以该列中1的个数，即可形成权值矩阵M；

(D)有利于由AT计算的权值矩阵M与网页重要度矩阵R的乘积符合网页重要度的计算方法：反向链接的加权和。

答案：D

解释：

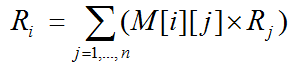
本题考查PageRank对网页链接关系的抽象--邻接矩阵及其转置的理解。

将A转置的原因是有利于由AT计算的权值矩阵M与网页重要度矩阵R的乘积符合网页重要度的计算方法：反向链接的加权和，选(D)。

具体内容请参考第九章课件之“PageRank排序—排序问题的不同思考方法”。

(5)PageRank算法中出现了一个“转移概率矩阵”，参见下图，其意义是\_\_\_\_\_。

(A)转移概率矩阵是基于网页链接关系矩阵AT计算得到的，按列来看是，网页j有多少个正向链接，其权值为多少分之一，反映了链接权值的计算方法；

(B)转移概率矩阵是基于网页链接关系矩阵AT计算得到的，按行来看是，网页i有多少个反向链接及其权值，可反映网页i的重要度计算方法即：，由其他网页的重要度及其权值计算该网页i的重要度；

(C)网页i的重要度Ri可以迭代地计算得到，设第m次得到的Ri记为Ri(m)，称为网页重要度Ri的一个状态，则状态转移概率为Ri由一个状态转变为另一个状态的概率；

(D)状态转移概率可广泛用于计算客观事物呈现状态序列S(0),...., S(m-1),S(m),...，而S(m)的计算仅由S(m-1)的值来确定的情况；

(E)上述说法都正确。

答案：E

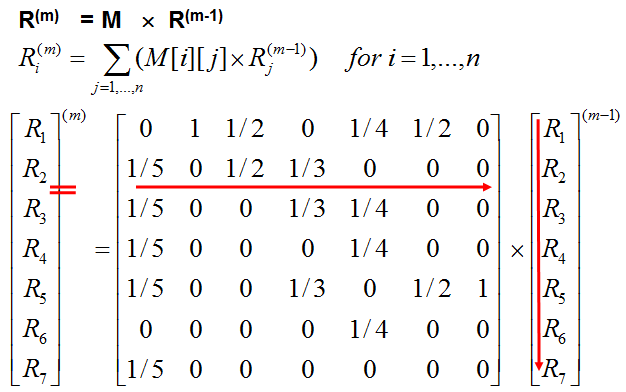
解释：

本题考查PageRank对转移概率矩阵的理解。

(A)(B)(C) (D)的叙述都正确，所以选(E)。

具体内容请参考第九章课件之“PageRank排序—排序问题的不同思考方法”。

(6)前述说过 PageRank网页i重要度Ri可以通过迭代地计算得到，即由m-1状态下各个网页的重要度R(m-1)，依转移概率矩阵计算m状态下网页重要度R(m)，参见下图。



关于网页重要度的计算过程，下列说法正确的是\_\_\_\_\_。

(A)在得到了转移概率矩阵M后，任意给出网页重要度的一组值，记为R(0)，是一向量，参见下图，继续进行(B)；

(B)不断地计算R(m)=M×R(m-1)，m从0开始，为迭代次数。当R(m) = R(m-1)时，迭代计算终止，此时的向量R即为所求的各个网页的重要度；

(C)选项(A)(B)是将状态序列R(0)，...，R(m-1)，R(m)，...不断迭代产生后趋于稳定的，或者说收敛的R(m)，作为最终的R，即是已知M情况下，求方程R = MR的解；

(D)上述说法都正确。

答案：D

解释：

本题考查PageRank对迭代法计算的理解。

网页重要度的计算过程正如(A)(B)所示，(C)是对(A)(B)的一个总结，都是正确的，所以选(D)。

具体内容请参考第九章课件之“PageRank排序—排序问题的不同思考方法”。

(7)前述说过 PageRank，通过不断地计算R(m)=M×R(m-1)来计算网页重要度，即由第(m-1)次的网页重要度来计算第(m)次的网页重要度，那么网页重要度的初始值R(0)应如何获得呢?

下列说法正确的是\_\_\_\_\_。

(A)随机产生各网页重要度的一组值，该组值对最终计算结果没有影响；

(B)由专家给出各网页重要度的一组值，该组值的质量好坏直接影响计算结果；

(C)设定各网页重要度都是1；

(D)随机产生各网页重要度的一组值，使网页重要度界于0和1之间，但该组值对最终结果没有影响。

答案：D

解释：

本题考查PageRank对初始值的处理方法的理解。

网页重要度的初始值R(0) 是随机产生各网页重要度的一组值，使网页重要度界于0和1之间，但该组值对最终结果没有影响，选(D)。

具体内容请参考第九章课件之“PageRank排序—排序问题的不同思考方法”。