1. **Selection Sort (Sắp xếp chọn):**
2. Ý tưởng: Ta thấy rằng khi một dãy ai, …, an đã được sắp xếp thì phần tử ai luôn là phần tử nhỏ nhất trong dãy ai, …, an. Nên ta có được ý tưởng để sắp xếp một dãy theo phương pháp chọn: tìm vị trí của phần tử nhỏ nhất trong n phần tử ban đầu và đưa giá trị của phần tử đó về đầu dãy. Tiếp tục với n - 1 phần tử còn lại. Và đến khi còn lại 1 phần tử thì phần dừng và phần tử đó chắc chắn là số lớn nhất. Cuối cùng, dãy đã được sắp xếp theo thứ tự tăng dần.
3. Thuật toán: Cho mảng gồm n phần tử a1, a2, …, an.

* Bước 1: Chọn i = 1.
* Bước 2: Tìm phần tử amin nhỏ nhất trong mảng ai, …, an.
* Bước 3: Hoán vị hai phần tử amin và an.
* Bước 4:
  + Nếu i ≤ n – 1 🡺 Lặp lại bước 2.
  + Ngược lại 🡺 Dừng thuật toán.

1. Đánh giá thuật toán:

* Độ phức tạp về thời gian: O(n2).
* Độ phức tạp về không gian: O(1).

1. **Insertion Sort (Sắp xếp chèn):**
2. Ý tưởng: Đối với một dãy đã được sắp xếp, khi ta chèn một phần tử vào dãy đó ở một vị trí thích hợp thì dãy vẫn giữ nguyên được thứ tự ban đầu.

* Cho một dãy gồm n phần tử a1, a2, …, an. Chia dãy ban đầu thành 2 dãy con a1, a2, …, ai (gọi là dãy B) và ai + 1, ai + 2, …, an (gọi là dãy C). Giả sử dãy a1, a2, …, ai đã có thứ tự (tăng dần). Chọn một phần tử từ dãy C, để tiện ta chọn ai + 1 (tương đương C0) và chèn vào dãy B sao cho nó vẫn giữ nguyên thứ tự (tăng dần).
* Tìm vị trí k trong đoạn [1, i] thoả ak - 1 ≤ ai + 1 < ak, có 2 trường hợp xảy ra:
  + Nếu tồn tại k: dời tất cả phần tử trong đoạn [k, i] về sau (sang phải) một vị trí. Và sau đó cập nhật lại ak bằng với ai + 1.
* Ngược lại: các phần tử trong đoạn từ [1, i + 1] đã có thứ tự.
* Tiếp tục thực hiện cho đến khi chèn hết phần tử từ dãy C sang dãy B. Kết quả là dãy B sẽ chứa tất cả các phần tử ban đầu theo thứ tự tăng dần.

1. Thuật toán: Cho mảng gồm n phần tử a1, a2, …, an.

* Bước 1: Chọn i = 2.
* Bước 2: Tìm vị trí k trong đoạn [1, i - 1] thoả ak > ai. Nếu k tồn tại thì dời các phần tử trong đoạn [k, i – 1] về sau một đơn vị (dịch sang phải một đơn vị) và chèn ai vào vị trí k (gán ak = ai).
* Bước 3: Tăng i lên thêm 1 (i = i + 1).
  + Nếu i ≤ n 🡺 Lặp lại bước 2.
  + Ngược lại 🡺 Dừng chương trình.

1. Đánh giá thuật toán:

* Độ phức tạp về thời gian:
  + Trường hợp tốt nhất: O(n).
  + Trường hợp xấu nhất: O(n2).
  + Trường hợp trung bình: O(n2).
* Độ phức tạp về không gian: O(1).

1. **Bubble Sort (Sắp xếp nổi bọt):**
2. Ý tưởng: Bất đầu duyệt từ cuối dãy và hoán vị 2 phần tử kề nhau để đưa được phần tử nhỏ hơn về đầu dãy. Và tiếp tục thực hiện như vậy sau khi đã bỏ ra phần tử đã được đưa về đầu dãy. Thực hiện cho đến khi dãy chỉ còn một phần tử, lúc này dãy đã được sắp xếp theo thứ tự.
3. Thuật toán: Cho mảng gồm n phần tử a1, a2, …, an.

* Bước 1: Cho i = 1 và j = n.
* Bước 2: Trong khi j > i (đưa phần tử nhỏ nhất của mảng ai, ai + 1, …, an về đầu mảng):
  + So sánh 2 phần tử kề nhau a­j – 1 và aj và hoán vị giá trị nếu chúng là hai phần tử nghịch thế.
  + Giảm j xuống 1 (j = j – 1).
* Bước 3: Tăng i lên thêm 1 (i = i + 1).
  + Nếu i < n (mảng còn nhiều hơn một phần tử) 🡺 Lặp lại bước 2.
  + Ngược lại 🡺 Dừng thuật toán.

1. Đánh giá thuật toán:

* Độ phức tạp về thời gian: O(n2).
* Độ phức tạp về không gian: O(1).

1. **Heap Sort (Sắp xếp vun đống):**
2. Ý tưởng:

* Khi thực hiện việc tìm vị trí phần tử nhỏ nhất ở bước i trong thuật toán sắp xếp chọn (selection sort), người dùng đã không tận dụng được kết quả của các phép so sánh ở bước i – 1. Để giải quyết vấn đề trên, mọi người đã tìm ra được một cấu trúc dữ liệu để lưu trữ các thông tin của phép so sánh ở các bước trước, có dạng cây như sau:

Ta có dãy số: 15 5 3 10 8 13 14.

A close up of a clock

Description automatically generated

Trong đó, phần tử ở mức i là phần tử lớn nhất ở mức i + 1. Do đó phần tử ở mức 0 (gốc của cây) sẽ là phần tử lớn nhất của dãy số. Vì vậy, khi loại bỏ phần tử ở gốc (tức là đưa phần tử lớn nhất về đầu dãy), ta chỉ cần quan tâm đến nhánh mà chứa phần tử đó và cập nhật lại nó, còn các nhánh khác đều được bảo toàn. Do đó, khi thực hiện bước kế tiếp ta đã tận dụng lại kết quả của phép so sánh ở bước trước đó.

Để cài đặt được thuật toán sắp xếp đó phải cần đến 2n – 1 ô nhớ để lưu trữ (mỗi ô là một 1 nút trên cây) giá trị các phần tử trên cây. Nên để tối ưu hơn J.Williams đã đề xuất ra một khái niệm mới là heap và thuật toán sắp xếp heap sort để giải quyết vấn đề trên mà chỉ cần n ô nhớ để lưu trữ.

* Định nghĩa Heap: Giả sử có một dãy số a1, a2, …, an và cần sắp xếp theo thứ tự tăng dần.
  + Các phần tử ai với i ∈ [1, n] luôn thoả: ai ≥ a2i và ai ≥ a2i + 1. Và (ai, a2i), (ai, a2i + 1) được gọi là các cặp phần tử liên đới.
  + Có 3 tính chất sau:
    - Tính chất 1: Nếu a1, a2, …, an là một heap thì khi loại bỏ một số phần tử ở hai đầu của heap, dãy còn lại vẫn là một heap.
    - Tính chất 2: Nếu a1, a2, …, an là một heap thì phần tử ở đầu dãy luôn là phần tử lớn nhất.
    - Tính chất 3: Mọi dãy aL, aL + 1, …, aR với 2L > R là một heap.
* Thuật toán xây dựng Heap: Cho một dãy số a1, a2, …, an.
  + Bước 1: Chọn L = n / 2. Vì dãy số an/2 +1, an/2 + 2, …, an đã là một heap theo tính chất 3.
  + Bước 2: Trong khi L ≥ 0 (thêm hết nửa dãy đầu vào dãy số đã là heap phía sau để dãy số ban đầu trở thành một heap):
    - Hiệu chỉnh dãy số aL, aL + 1 , …, an (dãy số aL + 1 , aL + 2, …, an đã là một heap nên khi thêm aL vào đầu mảng cần hiểu chỉnh lại dãy số để dãy số sau khi thêm aL vẫn là một heap).
    - Giảm L xuống thêm 1.
  + Bước 3: Dừng thuật toán.
* Thuật toán hiệu chỉnh dãy số: Có dãy số aL, aL + 1, …, aR.
  + Bước 1: Chọn i = L, j = 2i.
  + Bước 2: Nếu j ≤ R:
    - Tìm max(aj, aj + 1) (nếu có) để so sánh với phần tử cần hiệu chỉnh lại và gán lại j bằng với chỉ số phần tử lớn nhất.
    - Nếu ai > aj 🡺 Chuyển sang bước 3.
    - Ngược lại 🡺 Hoán vị giá trị ai và aj. Vì việc hoán vị này đã thay đổi giá trị aj nên có thể dãy số aj, aj + 1, …, aR đã không còn là heap nên lặp lại bước 1 với dãy số aj, aj + 1, …, aR.
  + Bước 3: Dừng thuật toán.

1. Thuật toán: Cho dãy số a1, a2, …, an.

* Bước 1: Xây dựng dãy số ban đầu thành heap.
* Bước 2: Hoán vị phần tử a1 và an (để đưa phần tử lớn nhất a1 về cuối dãy số).
* Bước 3: Hiệu chỉnh lại dãy số với n – 1 phần tử còn lại (bỏ phần tử cuối).
* Bước 4: Lặp lại bước 2 cho đến khi dãy chỉ còn 1 phần tử.

1. Đánh giá thuật toán:

* Độ phức tạp về thời gian: O(nlog(n)).
* Độ phức tạp về không gian: O(1).

1. **Merge Sort (Sắp xếp trộn):**
2. Ý tưởng: Dựa theo kỹ thuật chia để trị (divide and conquer). Lần lượt chia đôi dãy ban đầu thành 2 dãy con (kích thước dãy ban đầu giảm một nửa). Tiếp tục việc chia đôi dãy con như vậy cho đến khi được n dãy con (mỗi dãy con chứa một phần tử). Trộn theo thứ tự 2 dãy con được chia ra từ dãy con trước đó. Tiếp tục trộn như vậy cho đến khi nhận được một dãy có kích thước bằng với dãy ban đầu. Và dãy này cũng là dãy đã được sắp xếp theo thứ tự (tăng dần).

* Thuật toán trộn 2 mảng: mảng a gồm n phần tử a1, a2, …, an và mảng b gồm m phần tử b1, b2, …, bm.
* Bước 1: Khởi tạo mảng c với k = n + m phần tử và ic = ia = ib = 1.
* Bước 2: Trong khi ia ≤ n và ib ≤ m:
  + Nếu aia > bib (cặp phần tử nghịch thế) 🡺 Thêm phần tử aia vào mảng c và tăng ia lên thêm 1.
  + Ngược lại thêm phần tử bib vào mảng c và tăng ib lên thêm 1.
  + Tăng ic lên thêm 1.
* Bước 3:
  + Nếu ia ≤ n 🡺 Thêm tất cả phần tử còn lại của a vào c.
  + Ngược lại nếu ib ≤ m 🡺 Thêm tất cả phần tử còn lại của b vào c.
  + Ngược lại 🡺 Dừng thuật toán.

1. Thuật toán: Cho mảng gồm n phần tử a1, a2, ..., an.

* Bước 1: Chọn k = 1 (chiều dài của mảng con ở bước hiện tại).
* Bước 2: Tách mảng a1, a2, …, an thành 2 mảng b và c như sau:
  + Mảng b gồm: a1, …, ak, a2k + 1, …, a3k, …
  + Mảng c gồm: ak + 1, …, a2k, a3k + 1, …, a4k, …
* Bước 3: Tiến hành trộn từng mảng con gồm k phần tử của 2 mảng b và c vào a.
* Bước 4: Tăng k lên 2 lần (k = k \* 2).
  + Nếu k < n 🡺 Lặp lại bước 2.
  + Ngược lại 🡺 Dừng thuật toán.

1. Đánh giá thuật toán:

* Độ phức tạp về thời gian: O(nlog(n)).
* Độ phức tạp về không gian: O(n).

1. **Quick Sort (Sắp xếp nhanh):**
2. Ý tưởng: Áp dụng thuật toán phân hoạch dãy a1, a2, …, an thành 2 phần:

* Dãy con 1: gồm các phần tử nhỏ hơn x.
* Dãy con 2: gồm các phần tử lớn hơn x.

Với x là một phần tử bất kỳ trong dãy ban đầu. Sau khi phân hoạch dãy ban đầu ta có được 3 phần:

* Dãy 1: ak < x với k = 1…i.
* Dãy 2: ak = x với k = (i + 1)…(j – 1).
* Dãy 3: ak > x với k = j…n.

Trong đó dãy 2 đã có thứ tự, nếu dãy 1 hoặc dãy 3 chỉ chứa 1 phần tử thì dãy đó cũng có thứ tự. Ngược lại, nếu dãy 1 hoặc dãy 3 chứa nhiều hơn 1 phần tử thì tiếp tục thực hiện việc phân hoạch với các dãy đó như đã trình bày.

* Thuật toán phân hoạch: có dãy gồm n phần tử a1, a2, …, an.
* Bước 1: Chọn i = L = 1, j = R = n và x = ak với L ≤ k ≤ R.
* Bước 2: Thực hiện việc chỉnh lại vị trí cho 2 phần tử ai, aj đang bị đặt sai vị trí:
  + Trong khi ai < x thì tăng i lên thêm 1.
  + Trong khi aj > x thì giảm j xuống 1.
  + Nếu i ≤ j (có 2 phần tử đang bị đặt sai vị trí) thì hoán vị 2 phần tử ai, aj và tăng i thêm 1, giảm j xuống 1.
* Bước 3:
  + Nếu i ≤ j (chưa duyệt hết dãy số) thì lặp lại bước 2.
  + Ngược lại thì dừng thuật toán.

1. Thuật toán: Cho dãy a gồm n phần tử a1, a2, …, an.

* Bước 1: Chọn L = 1 và R = n.
* Bước 2: Áp dụng thuật toán phân hoạch dãy aL, …, aR thành 3 phần:
  + Dãy 1: ak < x với k = L…j.
  + Dãy 2: ak = x với k = (j + 1)…(i – 1).
  + Dãy 3: ak > x với k = i…R.
* Bước 3:
  + Nếu L < j (dãy 1 vẫn còn chứa nhiều hơn 1 phần tử) thì lặp lại bước 2 với R = j và giữ nguyên L.
  + Nếu i < R (dãy 3 vẫn còn chứa nhiều hơn 1 phần tử) thì lặp lại bước 2 với L = I và giữ nguyên R.
* Bước 4: Dừng chương trình.

1. Đánh giá thuật toán:

* Độ phức tạp về thời gian:
  + Trường hợp tốt nhất: O(nlog(n)).
  + Trường hợp tệ nhất: O(n2).
  + Trường hợp trung bình: O(nlog(n)).
* Độ phức tạp về không gian: O(log(n))

1. **Radix Sort (Sắp xếp theo cơ số):**
2. Ý tưởng: Phương pháp sắp xếp này không quá quan tâm đến việc so sánh giá trị của 2 phần tử mà dựa trên nguyên tắc phân loại thư của bưu điện. Cụ thể, để chuyển một khối lượng thư lớn đến tay người nhận. Bưu điện sẽ phân loại thư thuộc các tỉnh thành cụ thể và gửi đến tỉnh thành đó. Từ đó, các tỉnh thành sẽ phân loại thư của từng quận, huyện và gửi về quận, huyện đó. Thực hiện tương tự với phường xã. Do đó, mô phỏng lại cách chuyển thư của bưu điện, ta sẽ sắp xếp dãy số dựa vào việc phân loại các phần tử theo chữ số hàng đơn vị, hàng chục, hàng trăm,….
3. Thuật toán: Cho dãy số a1, a2, …, an. Qui ước ai[k] là chữ số thứ k của phần tử ai (từ trái sang phải).

* Bước 1: Tạo một mảng hai chiều b với n dòng và 10 cột (có thể thay bằng một mảng chứa 10 hàng đợi (queue) để tiết kiệm bộ nhớ). Xác định số lượng chữ số của phần tử lớn nhất (phần tử có nhiều chữ số nhất) và lưu vào k.
* Bước 2: Trong khi k > 0:
  + Duyệt qua n phần tử và thêm phần tử thứ i vào cột ai[k] – 1 của b.
  + Lấy các phần tử ra theo từng cột từ 1 đến 10 (theo đúng trình tự thêm vào của cột) và thêm vào a (theo thứ tự từ a1, a2, …, an).
  + Giảm k xuống 1 (k = k – 1).
* Bước 3: Dừng thuật toán.

1. Đánh giá thuật toán:

* Độ phức tạp về thời gian: O(k\*n) với k số chữ số của phần tử lớn nhất.
* Độ phức tạp về không gian: O(k + n).

**TÀI LIỆU THAM KHẢO**

1. **Đỗ Xuân Lôi**: *“Cấu trúc dữ liệu và giải thuật”*. Nhà xuất bản Đại học Quốc gia Hà Nội – 2010.
2. **Trần Hạnh Nhi – Dương Anh Đức**: *“NHẬP MÔN CẤU TRÚC DỮ LIỆU VÀ GIẢI THUẬT”*. Trường Đại học Khoa học Tự nhiên TP.HCM – 2003.
3. **Trần Đan Thư – Nguyễn Thanh Phương – Đinh Bá Tiến – Trần Minh Triết – Đặng Bình Phương**: *“KỸ THUẬT LẬP TRÌNH”*. Nhà xuất bản Khoa học và Kỹ thuật – 2019.
4. <https://codelearn.io/learning/cau-truc-du-lieu-va-giai-thuat/853804>