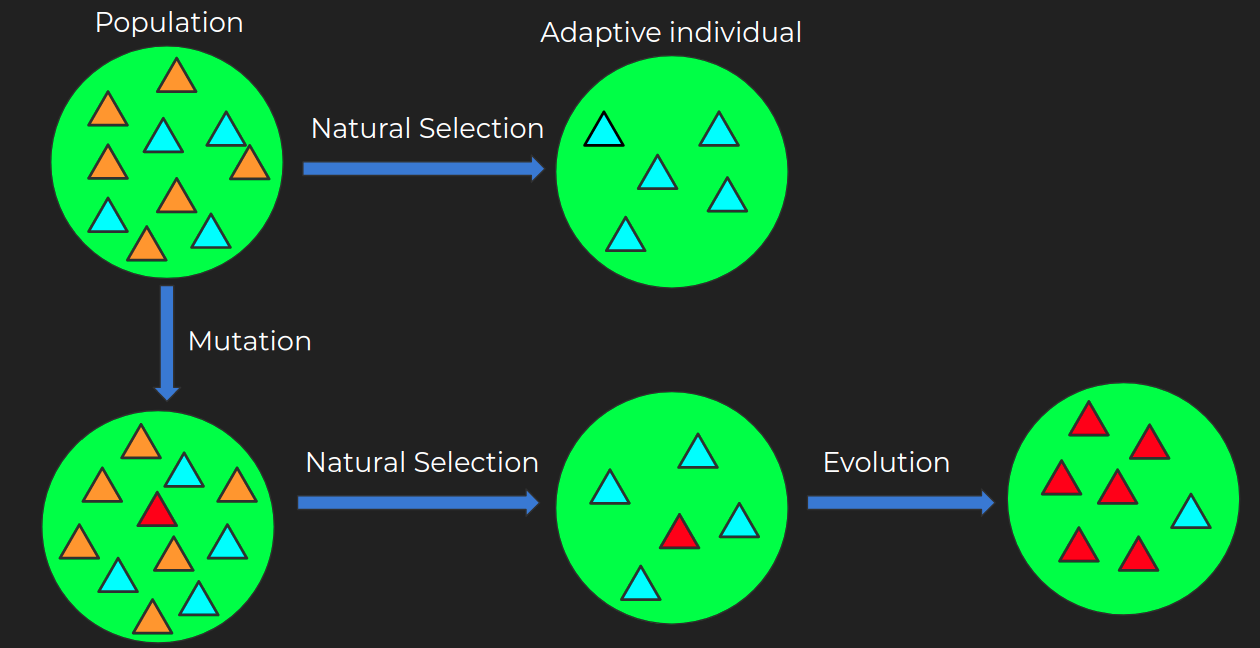
# Genetic Algorithm - Giải thuật di truyền

I. Thành phần (4 thành phần):



- Quần thể (Population):

* Ban đầu sẽ có những cá thể nhất định với những đặc tính khác nhau.
* Những đặc tính này sẽ quy định khả năng sinh sản, sinh tồn, khả năng đáp ứng điều kiện môi trường của từng cá thể.

- Chọn lọc tự nhiên (natural selection):

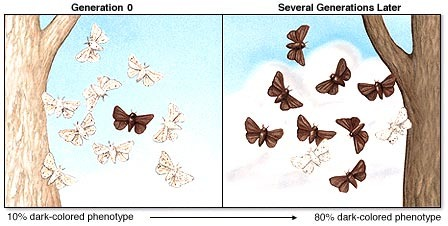
* Theo thời gian những cá thể yếu hơn, không có khả năng sinh tồn sẽ bị loại bỏ bởi những tác nhân như tranh chấp chuỗi thức ăn, môi trường tác độc, bị loài khác tiêu diệt, …
* Cuối cùng sẽ còn lại những cá thể có đặc tính ưu việt hơn sẽ được giữ lại - Adaptive individual (Cá thể thích nghi).

- Đột biến (mutation):

* Mỗi cá thể con được sinh ra sẽ được kế thừa lại những đặc tính của cả cha và mẹ.
* Sau một thời gian sinh sống, một quần thể sẽ đặt tới giới hạn của các cặp gen của con được tạo nên từ gen của bố mẹ.
* Để đạt được tới sự tiến hóa, Đột Biến chính là một trong những nguyên nhân chính, có vai trò đóng góp nguyên liệu cho quá trình Chọn lọc tự nhiên.

- Tiến hóa (evolution):

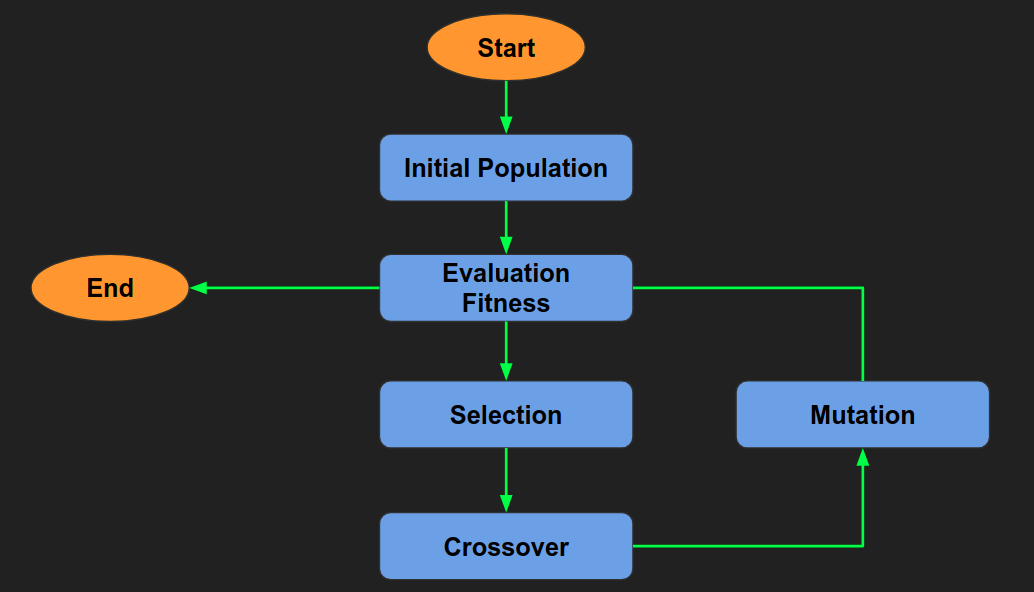
* Những cá thể đột biến không phải luôn là những cá thể mạnh mẽ và có đủ khả năng sinh tồn.
* Chọn lọc tự nhiên sẽ chọn ra những cá thể đột biến nhưng có thể thích nghi với môi trường sống tốt hơn những cá thể khác trong quần thể.
* Sau một thời gian sinh sản, những gen đột biến sẽ chiếm ưu thế và chiếm đa số trong quần thể.



II. Thuật toán di truyền:

- Có rất nhiều cách giải thích và biểu đồ khác nhau để diễn giải thuật toán.

- Nhưng hầu hết các thành phần chính của thuật toán sẽ không thay đổi.



**Xét bài toán tìm mật khẩu**

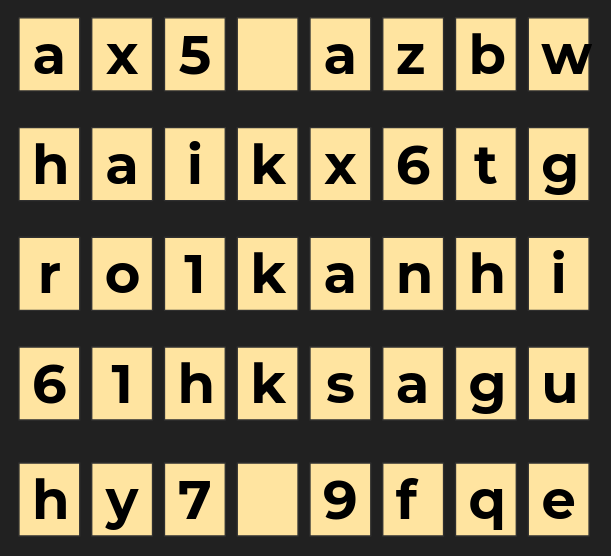
- Yêu cầu bài toán:

* + - Mật khẩu gồm 8 kí tự (chữ, số, khoảng cách) – ví dụ: holiamgi.
  + Mỗi lần thử, hệ thống sẽ báo về số lượng kí tự đúng với mật khẩu.
  + Yêu cầu tìm ra chuỗi mật khẩu cho trước.

- Thuật toán sẽ dừng lại khi tìm được cá thể đáp ứng được nhu cầu đề ra sau mỗi thế hệ mới.

- Quá trình sản sinh thế hệ tiếp theo sẽ là một vòng lặp (Evaluation Fitness -> Selection -> Crossover -> Mutation):

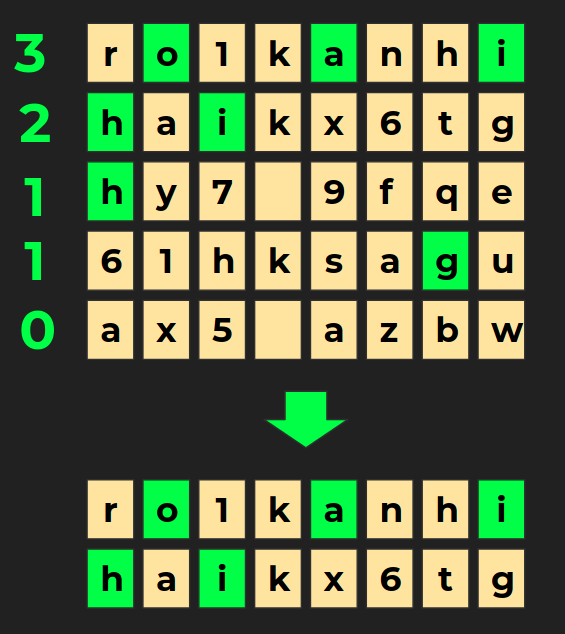
* Khởi tạo quần thể (initial population):
  + Quần thể sẽ bao gồm những chuỗi 8 kí tự, được sinh ra ngẫu nhiên.



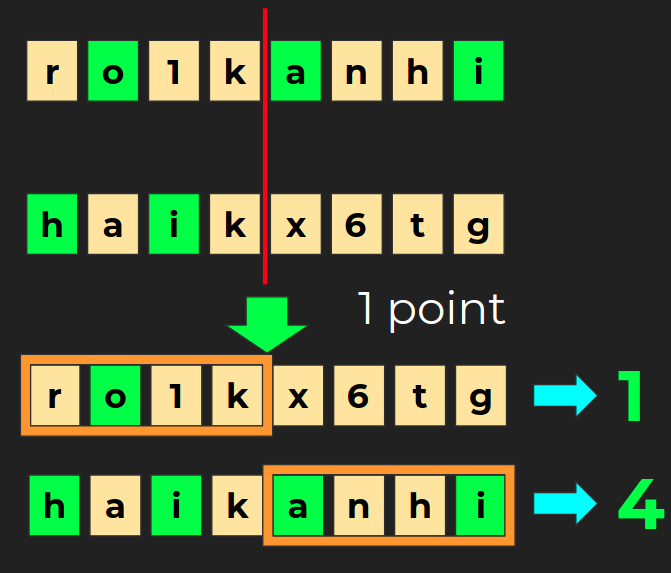
* Đánh giá năng (evaluation fitness):
  + Mỗi chuỗi mật khẩu sẽ được đánh giá sự chính xác so với mật khẩu cho trước, với mỗi kí tự giống với mật khẩu cho trước tại đúng vị trí sẽ được 1 point.



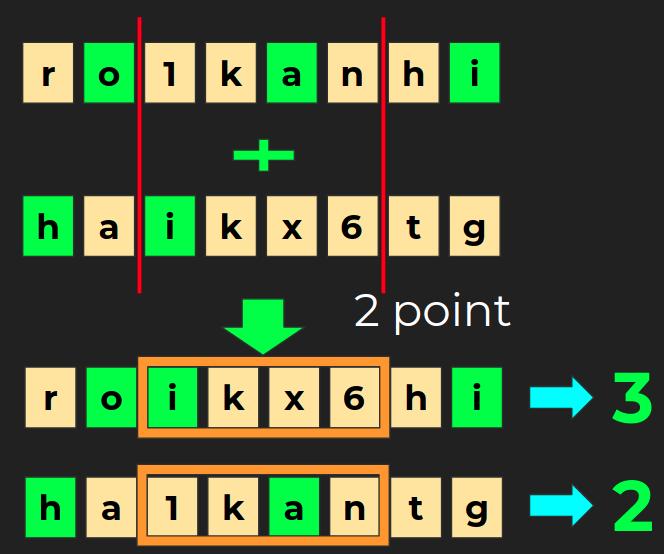
* + Point ở đây sẽ đại diện cho khả năng sinh tồn của cá thể trong quần thể, càng lớn tức cá thể đó càng thích nghi với môi trường tốt.
* Chọn lọc (selection):
  + Sau khi đã đánh giá được quần thể, các cá thể có khả năng sinh tồn tốt hơn sẽ có cơ hội được sinh sản nhiều hơn các cá thể còn lại.
  + Các chuỗi kí tự mật khẩu sẽ được lựa chọn theo số Point đang có.



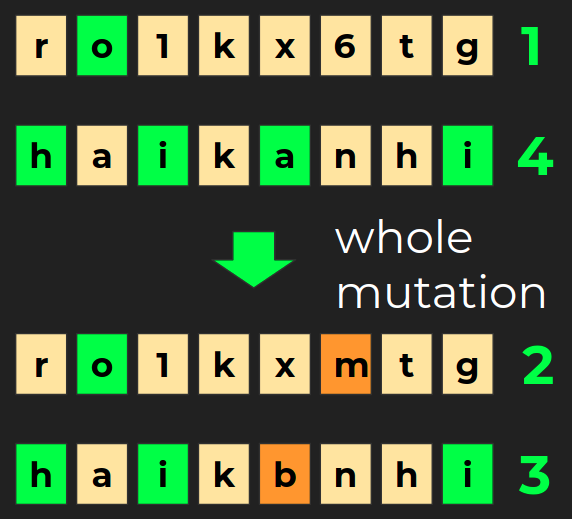
* Sinh sản (crossover):
  + Các cá thể con sẽ được kế thừa các đặc tính từ cả bố và mẹ.
  + Thông thường, cá thể con sẽ nhận một nửa gen từ mỗi bố, mẹ.



* + Có những kiểu lai tạo khác nhau như 2 Point, Uniform Selection.



* Đột biến (mutation):
  + Dễ nhận thấy rằng, nếu chỉ bằng việc sinh ngẫu nhiên và lai tạo, sẽ rất khó để tìm được nghiệm.
  + Trừ khi cá thể khởi tạo phù hợp luôn với yêu cầu đề bài, tức là có đáp án luôn từ đầu - Ăn May.
  + Như đã nói từ đầu, Đột biến chính là nguyên liệu của Chọn lọc tự nhiên, bằng việc lựa chọn ngẫu nhiên các vị trí và thay thế bằng một kí tự ngẫu nhiên nào đó, chúng ta có thể mô phỏng lại hiện tượng đột biến - [Đột biến điểm](https://vi.wikipedia.org/wiki/Đột_biến_điểm).



* + Các cá thể đột biến có thể sẽ có khả năng thích nghi tốt hơn (1 -> 2), hoặc cũng có thể ngược lại (4 -> 3). Quá trình này sẽ lặp lại cho đến khi tìm được đáp án phù hợp.

- Những vấn đề thường gặp:

* Thuật toán Di Truyền sẽ đạt được lợi thế trong những bài toán có không gian tìm kiếm quá lớn mà những giải thuật vét cạn không thể xử lí được.
* Tuy nhiên khi sử dụng, chúng ta sẽ cần cân nhắc những vấn đề ảnh hưởng tới việc lựa chọn thuật toán.

- Evaluation Fitness:

* Vấn đề đầu tiên gặp phải đó chính là việc đánh giá các cá thể trong quần thể, chúng ta cần một phương thức để có thể đánh giá sự thích nghi hay ưu thế của từng cá thể.

- Tuy nhiên khi thực hiện việc này, chúng ta sẽ cần cân nhắc 2 vấn đề:

* Tính khả thi: Việc tìm phương thức đánh giá không phải luôn khả thi, ví dụ trong bài toàn tìm một giai điệu mới, việc đánh giá giai điệu đó có “dễ nghe”, “hay” hay không thuộc về cảm quan của mỗi người, nên việc tìm một hàm đánh giá chính xác sẽ rất khó khăn.
* Chi phí: Đây cũng là một vấn đề cần cân nhắc, nếu chi phí tính toán của phương thức đánh giá quá lớn, việc sử dụng để tìm kiếm trong không gian sẽ mất nhiều thời gian, thậm chí lâu hơn vét cạn.

- Đánh giá:

* Trong bài toán Tìm mật khẩu thì chuỗi kí tự đại diện cho các Nhiễm sắc thể hay Chuỗi ADN.
* Việc số hóa những đặc tính của bài toán thành các bits, bytes để có thể sử dụng trong khâu Crossover, Mutation cũng gặp nhiều khó khăn trong những bài toán khác nhau.
* Bởi với mỗi sự thay đổi bit sẽ phải tương ứng với việc tạo ra một cá thể với đặc tính khác nhau.

### Ứng dụng thực tiễn

* Như đã đề cập trước đó, Giải thuật Di truyền hay Genetic Algorithm sẽ có ưu thế trong những bài toán mà có sẵn lời giải trong một không gian tìm kiếm lớn như những bài toán cổ điển Người du lịch, Cái Túi, … Ngoài ra, hiện nay với việc bùng nổ của AI, việc sử dụng mạng Neural cùng với giải thuật này cũng giúp giải quyết một số bài tóan về tự động, hành vi, …