**BỘ GIÁO DỤC VÀ ĐÀO TẠO**

**TRƯỜNG CÔNG NGHIỆP THỰC PHẨM**

**2020-2021**

**🙞🙜**

****

ĐỀ TÀI BÁO CÁO NHÓM 10:

GIẢI THUẬT TIẾN HOÁ

( EVOLUTION ALGORITHM)

Giáo Viên Hướng Dẫn: Trần Đình Toàn

Sinh Viên Thực Hiện:

2001181383 Vũ Minh Trung

2001181020 Nguyễn Chánh Bảo

2001181393 Huỳnh Tăng Anh Tú

EVOLUTIONARY ALGORITHM

[**I.** **Giải thuật tiến hoá** 3](#_Toc57096883)

[**1.** **Tiến hoá**: 3](#_Toc57096884)

[**2.Heuristic Là gì ?:** 4](#_Toc57096885)

[*2.1Thuật giải Heuristic* 4](#_Toc57096886)

[*2.2 Để xây dựng một thuật giải Heuristic* 4](#_Toc57096887)

[**3.** **Giải thuật di truyền (GA-Genetic Algorithm)** 5](#_Toc57096888)

[**II.** **Thuật toán tiến hoá :** 8](#_Toc57096889)

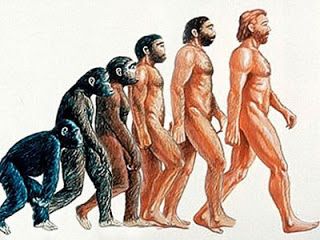
[**1)** **Mối quan hệ :** 8](#_Toc57096890)

[**2)** **Thuật toán di truyền** 10](#_Toc57096891)

[**3)** **Demo** 16](#_Toc57096892)

1. **Giải thuật tiến hoá**
2. **Tiến hoá**:

Tiến hóa nói đến quá trình hoàn thiện, biến đổi dần để hoàn thiện hơn các bộ phận, chức năng của các sinh vật để phù hợp hơn với điều kiện sinh tốn cũng đang dần thay đổi.



Trong sinh học, tiến hóa là sự thay đổi đặc tính di truyền của một quần thể sinh học qua những thế hệ nối tiếp nhau. Các quá trình tiến hóa làm nảy sinh sự đa dạng ở mọi mức độ tổ chức sinh học bao gồm loài, các cá thể sinh vật và cả các phân tử như ADN và protein.

**Một giải thuật tiến hóa (evolutionary algorithm - EA)** là một thuật toán tối ưu hóa **heuristic** sử dụng các kỹ thuật bắt nguồn từ các cơ chế tiến hóa hữu cơ (organic evolution) chẳng hạn như **biến dị, tái tổ hợp** và **chọn lọc tự nhiên** để tìm một cấu hình tối ưu cho một hệ thống với các ràng buộc cụ thể.

* **Biến dị sinh học** là những biến đổi mới mà cơ thể sinh vật thu được do tác động của các yếu tố môi trường và do quá trình tái tổ hợp di truyền.
* **Tái tổ hợp gen** là hiện tượng sinh vật bố mẹ sinh ra con có những tính trạng khác với bố mẹ do sự tổ hợp lại những gen vốn có của bố mẹ, phát sinh ra bởi sự chuyển đổi lô-cut của các gen trong hệ gen.
* **Chọn lọc tự nhiên** là quá trình phân hoá khả năng sống sót và mức thành đạt sinh sản của các kiểu gen khác nhau trong quần thể, từ đó dẫn đến đào thải các kiểu hình kém thích nghi, đồng thời tăng cường khả năng sống sót của các dạng thích nghi, tạo cơ hội cho các kiểu gen thích nghi này đóng góp vào vốn gen của quần thể ở thế hệ sau.

**2.Heuristic Là gì ?:**

Là các kỹ thuật dựa trên kinh nghiệm để giải quyết vấn đề, học hỏi hay khám phá nhằm đưa ra một giải pháp mà không được đảm bảo là tối ưu. Với việc nghiên cứu khảo sát không có tính thực tế, các phương pháp heuristic được dùng nhằm tăng nhanh quá trình tìm kiếm với các giải pháp hợp lý thông qua các suy nghĩ rút gọn để giảm bớt việc nhận thức vấn đề khi đưa ra quyết định. Ví dụ của phương pháp này bao gồm sử dụng ***một luật ngón tay cái,giả thuyết, phán đoán trực giác, khuôn mẫu hay nhận thức thông thường*.**

*2.1Thuật giải Heuristic*

*Thuật giải Heuristic là một sự mở rộng khái niệm thuật toán. Nó thể hiện cách giải bài toán với các đặc tính sau :*

* Heuristic tìm được lời giải tốt (không là lời giải tốt nhất)
* Heuristic là phương pháp tiếp cận bằng cảm tính, gần gũi với cách suy nghĩ và hành động của con người.
* Heuristic thường dễ dàng và nhanh chóng đưa ra kết quả hơn so với giải thuật tối ưu, vì vậy chi phí thấp hơn.

*2.2 Để xây dựng một thuật giải Heuristic*

* ***Nguyên lý vét cạn thông minh :***

Trong một bài toán tìm kiếm, khi không gian tìm kiếm lớn, tìm cách giới hạn lại không gian tìm kiếm hoặc thực hiện một kiểu dò tìm đặc biệt dựa vào đặc thù của bài toán để nhanh chóng tìm ra mục tiêu.

* ***Nguyên lý tham lam (Greedy):***

Lấy tiêu chuẩn tối ưu (trên phạm vi toàn cục) của bài toán để làm tiêu chuẩn chọn lựa hành động cho phạm vi cục bộ của từng bước (hay từng giai đoạn) trong quá trình tìm kiếm lời giải.

* ***Nguyên lý thứ tự :***

Thực hiện hành động dựa trên một cấu trúc thứ tự hợp lý của không gian khảo sát nhằm nhanh chóng đạt được một lời giải tốt.

* ***Hàm Heuristic:***

Trong việc xây dựng các thuật giải Heuristic, người ta thường dùng các *hàm Heuristic*. Ðó là các hàm đánh giá thô, giá trị của hàm phụ thuộc vào trạng thái hiện tại của bài toán tại mỗi bước giải. Nhờ giá trị này, ta có thể chọn được cách hành động tương đối hợp lý trong từng bước của thuật giải.

1. **Giải thuật di truyền (GA-Genetic Algorithm)**

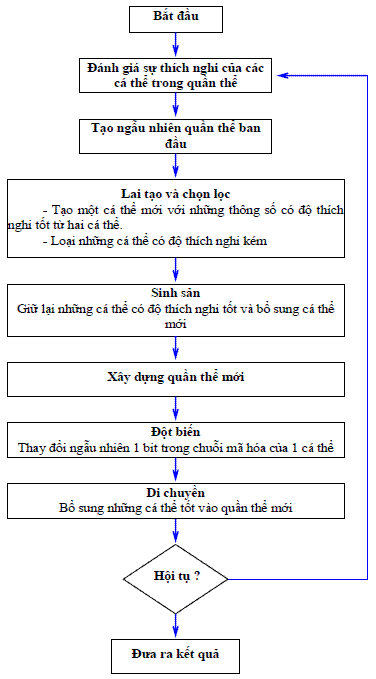
Là kỹ thuật phỏng theo quá trình thích nghi tiến hóa của các quần thể sinh học dựa trên học thuyết Darwin. GA là phương pháp tìm kiếm tối ưu ngẫu nhiên bằng cách mô phỏng theo sự tiến hóa của con người hay của sinh vật. Tư tưởng của thuật toán di truyền là mô phỏng các hiện tượng tự nhiên, là kế thừa và đấu tranh sinh tồn.

GA thuộc lớp các giải thuật xuất sắc nhưng lại rất khác các giải thuật ngẫu nhiên vì chúng kết hợp các phần tử tìm kiếm trực tiếp và ngẫu nhiên. Khác biệt quan trọng giữa tìm kiếm của GA và các phương pháp tìm kiếm khác là GA duy trì và xử lý một tập các lời giải, gọi là một quần thể (population).

Trong GA, việc tìm kiếm giả thuyết thích hợp được bắt đầu với một quần thể, hay một tập hợp có chọn lọc ban đầu của các giả thuyết. Các cá thể của quần thể hiện tại khởi nguồn cho quần thể thế hệ kế tiếp bằng các hoạt động lai ghép và đột biến ngẫu nhiên – được lấy mẫu sau các quá trình tiến hóa sinh học. Ở mỗi bước, các giả thuyết trong quần thể hiện tại được ước lượng liên hệ với đại lượng thích nghi, với các giả thuyết phù hợp nhất được chọn theo xác suất là các hạt giống cho việc sản sinh thế hệ kế tiếp, gọi là cá thể (individual). Cá thể nào phát triển hơn, thích ứng hơn với môi trường sẽ tồn tại và ngược lại sẽ bị đào thải. GA có thể dò tìm thế hệ mới có độ thích nghi tốt hơn.

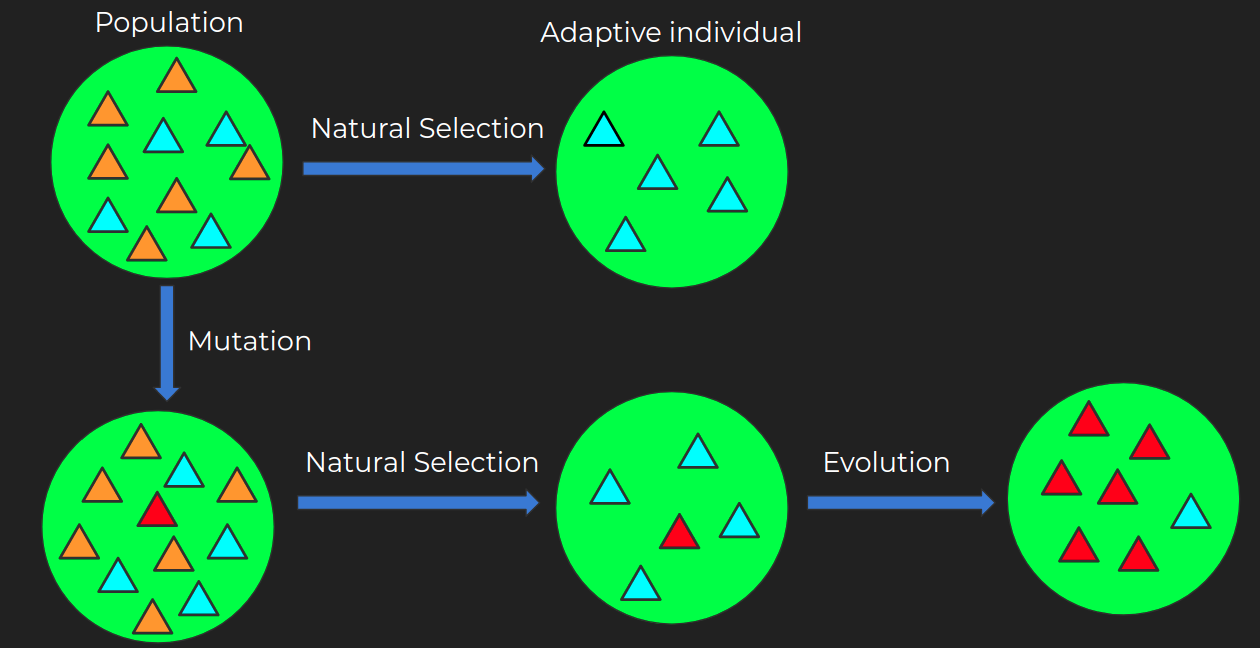
GA giải quyết các bài toán quy hoạch toán học thông qua các quá trình cơ bản***: lai tạo (crossover), đột biến (mutation) và chọn lọc (selection)*** cho các cá thể trong quần thể. Dùng GA đòi hỏi phải xác định được: khởi tạo quần thể ban đầu, hàm đánh giá các lời giải theo mức độ thích nghi – hàm mục tiêu, các toán tử di truyền tạo hàm sinh sản.

Sơ đồ thuật toán của GA:

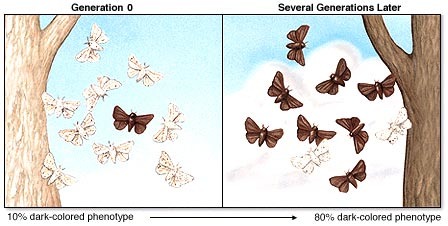


1. **Thuật toán tiến hoá :**
2. **Mối quan hệ :**

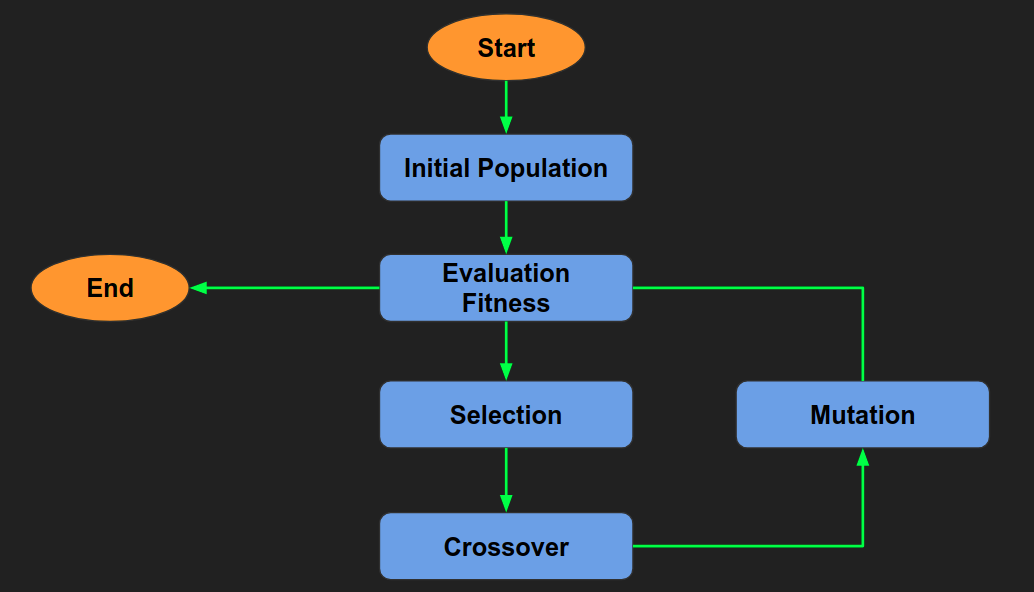
Để nắm rõ ý tưởng chính của thuật toán, biểu đồ dưới đây sẽ mổ tả cụ thể những thành phần và mối liên hệ của chúng.



1. **Population - Quần thể** : Một quần thể 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.
2. **Natural Selection - Chọn lọc tự nhiên** : 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**.
3. **Mutation - Đột biến** : Như chúng ta đã biết thì 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**.
4. **Evolution - Tiến hóa** : 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.



## **Thuật toán di truyền**



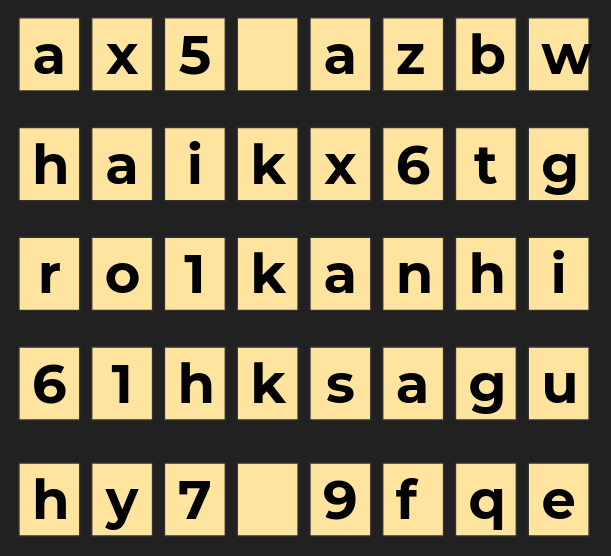
Xét bài toán **Tìm mật khẩu**, yêu cầu của bài toán như sau:

* Mật khẩu gồm 8 kí tự ( bao gồm chữ cái, chữ số và khoảng trắng) - Ví dụ: **hoilamgi**.
* 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**). Chúng ta sẽ cùng xây dựng những thành phần chính trong thuật toán để giải quyết bài toán này.

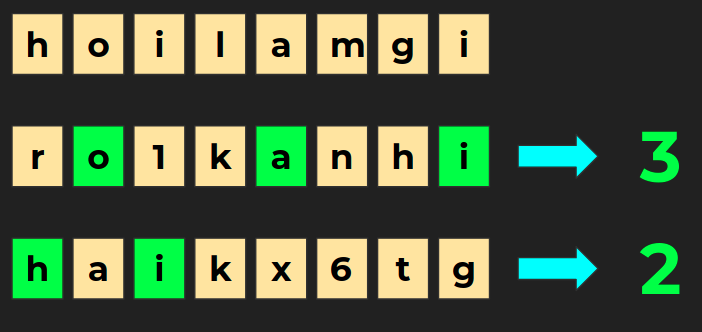
#### **Initial Population- Khởi tạo quần thể**

Nhìn chung thuật toán sẽ mô phỏng lại hầu hết những hiện tượng xảy ra trong quá trình tiến hóa của động vật. Vì vậy để thuật toán có thể vận hành được, thì điều đầu tiên cần có chính là **Quần thể**. Xét bài toán tìm mật khẩu, quần thể sẽ bao gồm những chuỗi 8 kí tự, được sinh ra ngẫu nhiên.



#### **Evaluation Fitness - Đánh giá năng lực**

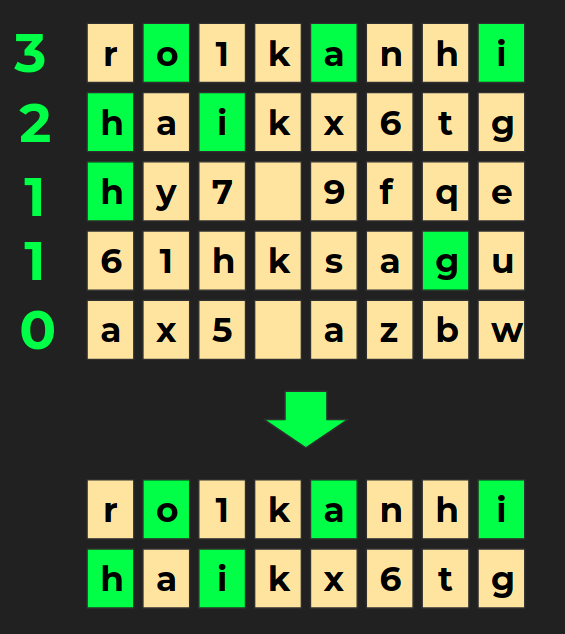
Tiếp theo, mỗi chuỗi mật khẩu sẽ được đánh giá sự chính xác so với mật khẩu cho trước, với mỗi kí tự giống với mật khẩu cho trước tại đúng vị trí sẽ được **1 point**.



Thành phần **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.

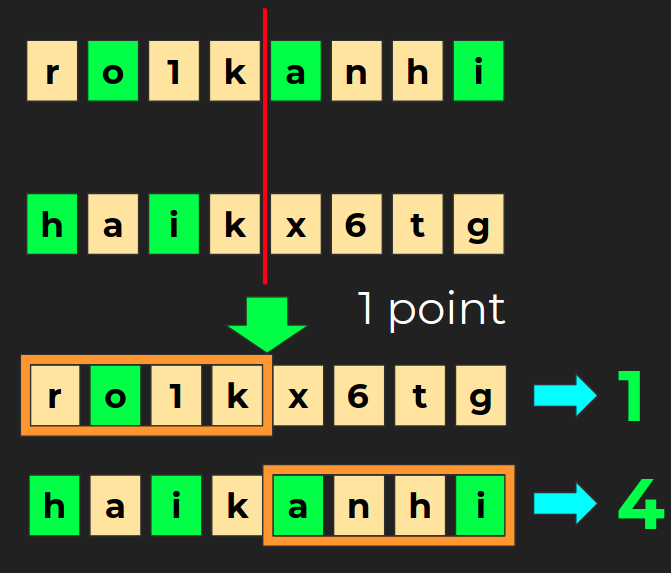
#### **Selection - Chọn lọc**

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.

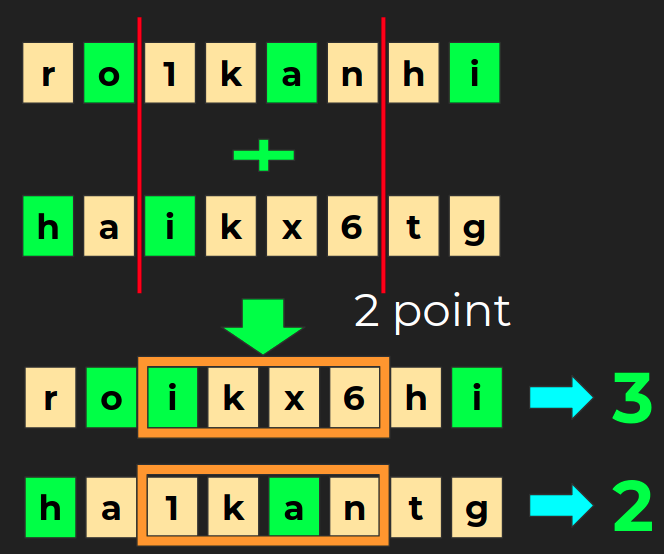


#### **Crossover - Sinh sản**

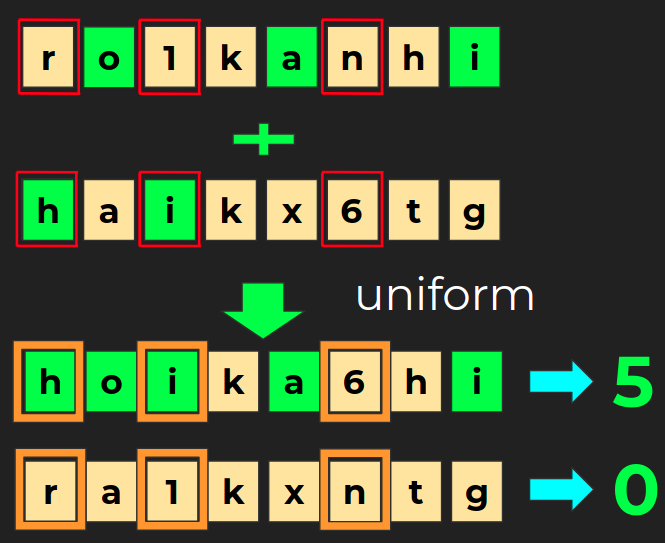
Vào giai đoạn sinh sản, các cá thể con sẽ được kế thừa các đặc tính từ cả bố và mẹ. Thông thường, cá thể con sẽ nhận một nửa gen từ mỗi bố, mẹ.



Cá thể con có thể sẽ thích nghi tốt hơn, hoặc kém hơn. Ngoài ra, có những kiểu lai tạo khác nhau như **2 Point**, **Uniform Selection**.



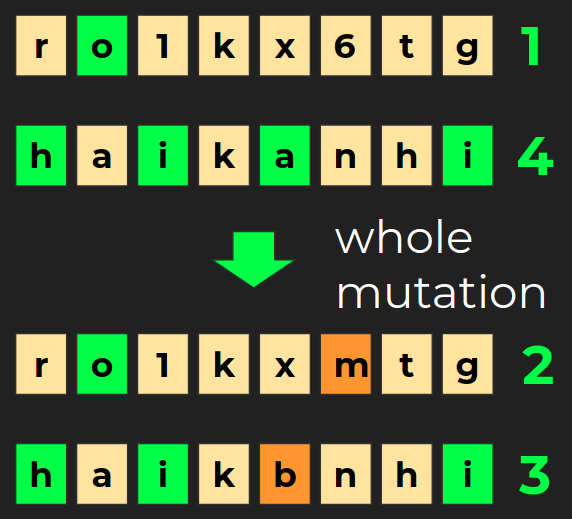
2 Point



Uniform

#### **Mutation - Đột biến**

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/%C4%90%E1%BB%99t_bi%E1%BA%BFn_%C4%91i%E1%BB%83m)



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.

🙑🙐 ***Nhận Xét :***

* 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.

1. **Demo**

Video :

Source Code:

1. **Đánh Giá :**

* 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.