Közelitő illesztés

Lehotay-Kéry Péter
lepuaai@inf.elte.hu
Ben Langmead diasora alapján
(www.langmead-lab.org/teaching-materials)

Szükség van közelitő illesztésre

- Szekvencia eltérések miatt
  - Szekvenálási hibák
  - Természetes variációk

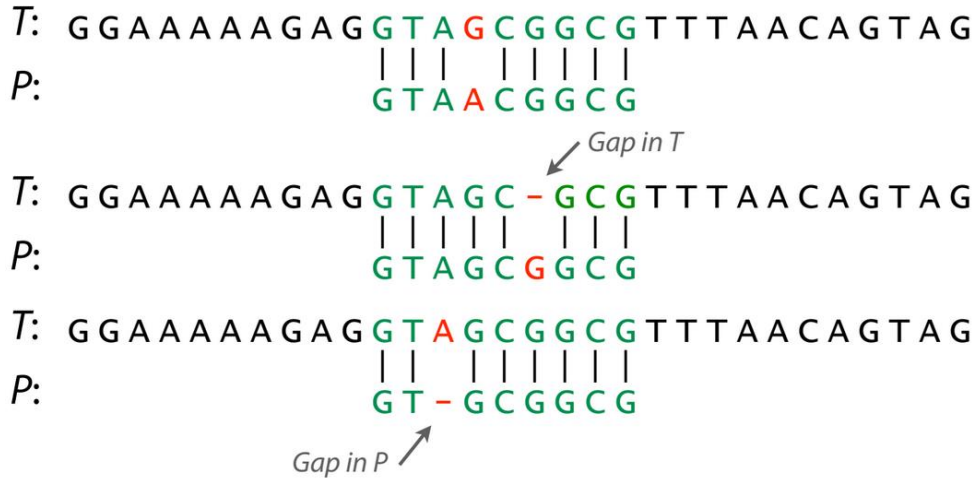
Eltérés, behelyettesités

Beillesztés

Törlés

Közelitő string illesztés

- Olyan helyeket keresünk, ahol P illeszkedik T-re, megengedve egy adott számú eltérést, vagy szerkesztést.
- Egy eltérés egy karakternyi helyettesités.
- Egy szerkesztés lehet egy karakternyi helyettesités, vagy hézag.



Hamming és szerkesztés távolság

- 2 azonos hosszú X és Y stringre, a hamming távolság a szükséges minimális számú egy-karakteres behelyettesitési, ahhoz hogy egyik stringet a másikká változtassuk.

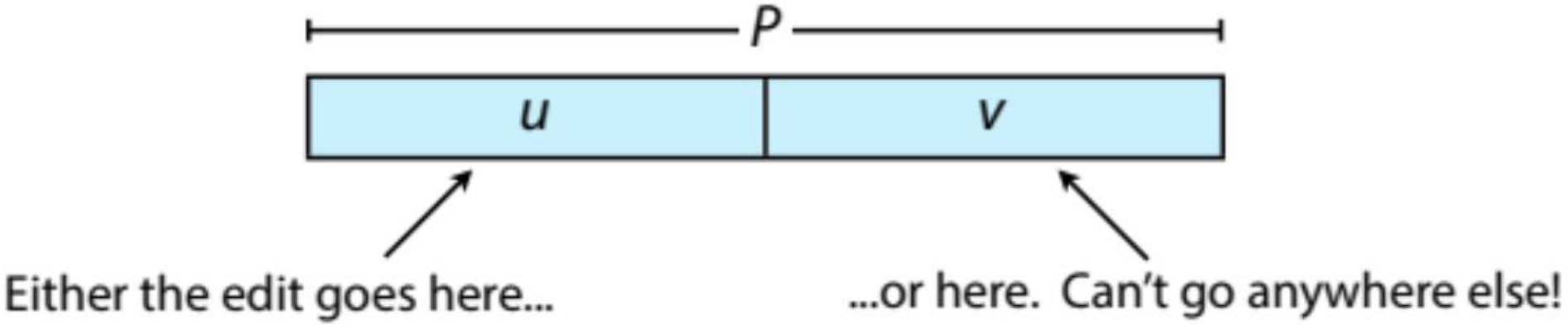
Hamming distance = 3

- Szerkesztés távolság a szükséges minimális számú szerkesztés, ahhoz, hogy egyik a másikká váljon.

Edit distance = 2

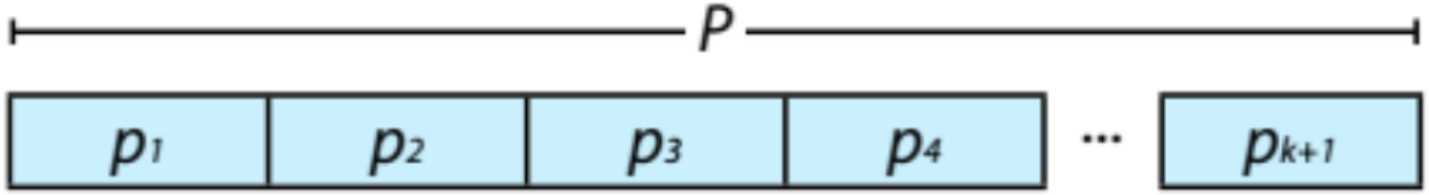
Közelitő string illesztés

- Hogyan tegyük Boyer-Moore és indexek által segitett illesztéseket közelitővé?
- Osszuk P-t nem-üres, nem-átfedő részstringekre: u és v. Ha P előfordul T-ben 1 szerkesztéssel, u-nak, vagy v-nek pontosan egyeznie kell.



Közelitő string illesztés

- Általánosabban: Legyen p1, p2, ..., pk+1 P particionálása k+1 nem-átfedő, nem-üres stringekbe. Ha P előfordul T-ben maximum k szerkesztéssel, akkor p1, p2, ..., pk+1 közül legalább egynek pontosan kell egyeznie.



- Használjuk a pontos illesztés algoritmust hogy pontos illeszkedést találjunk p1, p2, ..., pk+1-el. Keressünk hosszabb közelitő illeszkedést a pontos illeszkedés körül.