



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2019-0096238
(43) 공개일자 2019년08월19일

(51) 국제특허분류(Int. Cl.)
G16B 40/00 (2019.01) G16B 20/00 (2019.01)
G16B 25/00 (2019.01)
(52) CPC특허분류
G16B 40/00 (2019.02)
G16B 20/00 (2019.02)
(21) 출원번호 10-2018-0015922
(22) 출원일자 2018년02월08일
심사청구일자 2018년02월08일

(71) 출원인
연세대학교 산학협력단
서울특별시 서대문구 연세로 50 (신촌동, 연세대학교)
(72) 발명자
한요섭
서울특별시 은평구 진관1로 77-8, 403동 204호 (진관동, 은평뉴타운포동아파트)
김휘
서울특별시 관악구 낙성대역18길 8, 203호(봉천동, 명진빌라)
마르코 코그네타
서울특별시 마포구 월드컵로3길 14, 202동 1411호(합정동)
(74) 대리인
민영준

전체 청구항 수 : 총 15 항

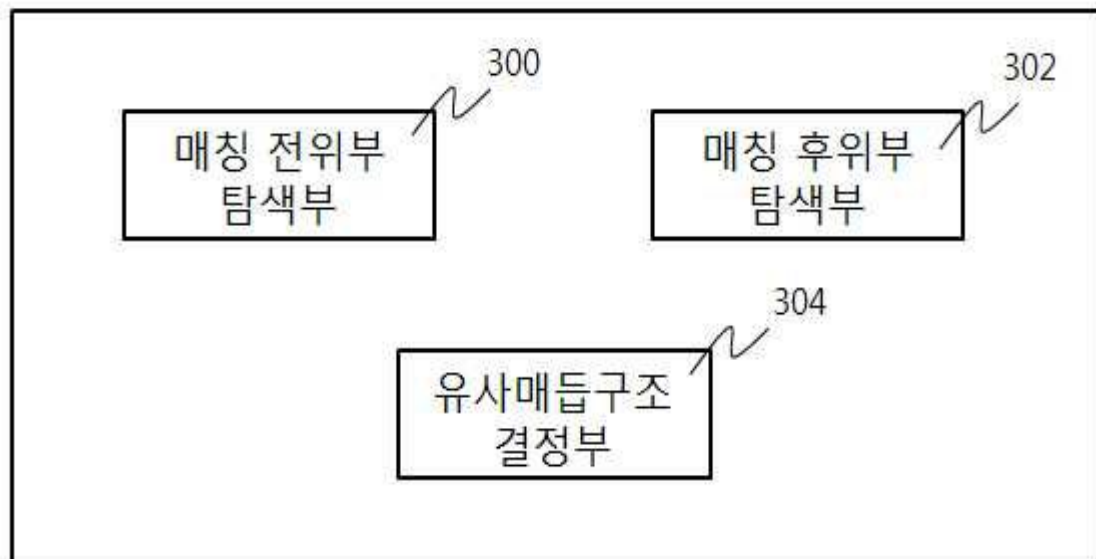
(54) 발명의 명칭 유사매듭구조 탐색 장치 및 방법

(57) 요약

유사매듭구조 탐색 장치 및 방법이 개시된다. 개시된 장치는, 주어진 서열로부터 전위부(u)와 상보적으로 결합되는 매칭 전위부의 길이가 후위부(x)와 상보적으로 결합되는 매칭 후위부의 길이보다 짧다고 가정한 상태에서 상기 서열의 위치 i에서 상보적 결합 조건 및 길이 조건을 만족하는 가장 긴 u의 길이인 c[i]와 상보적 결합 조건

(뒷면에 계속)

대표도 - 도3



은 만족하나 길이 조건을 만족하지 못하는 가장 짧은 u 의 길이인 $c' [i]$ 를 출력하는 매칭 전위부 탐색부; 주어진 서열로부터 후위부(x)와 상보적으로 결합되는 매칭 후위부의 길이가 전위부(u)와 상보적으로 결합되는 매칭 전위부의 길이보다 짧다고 가정한 상태에서 상기 서열의 위치 i 에서 상보적 결합 조건 및 길이 조건을 만족하는 가장 긴 x 의 길이인 $d[i]$ 와 상보적 결합 조건은 만족하나 길이 조건을 만족하지 못하는 가장 짧은 x 의 길이인 $d' [i]$ 를 출력하는 매칭 후위부 탐색부; 및 상기 $c[i]$, $c' [i]$, $d[i]$ 및 $d' [i]$ 를 이용하여 상기 전위부 및 후위부의 길이를 결정하는 유사매듭구조 결정부를 포함한다. 개시된 장치 및 방법에 의하면, 간소화된 연산으로 유사매듭구조를 탐색할 수 있는 장점이 있다.

(52) CPC특허분류

G16B 25/00 (2019.02)

이 발명을 지원한 국가연구개발사업

과제고유번호 2017080560

부처명 교육부

연구관리전문기관 한국연구재단

연구사업명 이공학개인지초연구지원사업

연구과제명 정형언어 기반의 효율적 분자 화합물 인코딩 기법 및 검색 기법

연구(3/3)(2015.11.1~2018.10.31)

기 여 율 1/1

주관기관 연세대학교 산학협력단

연구기간 2017.11.01 ~ 2018.10.31

명세서

청구범위

청구항 1

주어진 서열로부터 전위부(u)와 상보적으로 결합되는 매칭 전위부의 길이가 후위부(x)와 상보적으로 결합되는 매칭 후위부의 길이보다 짧다고 가정한 상태에서 상기 서열의 위치 i에서 상보적 결합 조건 및 길이 조건을 만족하는 가장 긴 u의 길이인 $c[i]$ 와 상보적 결합 조건은 만족하나 길이 조건을 만족하지 못하는 가장 짧은 u의 길이인 $c'[i]$ 를 출력하는 매칭 전위부 탐색부;

주어진 서열로부터 후위부(x)와 상보적으로 결합되는 매칭 후위부의 길이가 전위부(u)와 상보적으로 결합되는 매칭 전위부의 길이보다 짧다고 가정한 상태에서 상기 서열의 위치 i에서 상보적 결합 조건 및 길이 조건을 만족하는 가장 긴 x의 길이인 $d[i]$ 와 상보적 결합 조건은 만족하나 길이 조건을 만족하지 못하는 가장 짧은 x의 길이인 $d'[i]$ 를 출력하는 매칭 후위부 탐색부; 및

상기 $c[i]$, $c'[i]$, $d[i]$ 및 $d'[i]$ 를 이용하여 상기 전위부 및 후위부의 길이를 결정하는 유사매듭구조 결정부를 포함하는 것을 특징으로 하는 유사매듭구조 탐색 장치.

청구항 2

제1항에 있어서,

상기 매칭 전위부 탐색부는 $i=(n-1)$ 부터 감소시키면서 상기 $c[i]$ 및 $c'[i]$ 를 연산하며, $i+1$ 위치에서 상보적 결합이 있다고 판단되는 경우 상기 $c[i]$ 및 상기 $c'[i]$ 는 $s=c[i+1]$ 및 $p=c'[i+1]$ 에 기초하여 연산되는 것을 특징으로 하는 유사매듭구조 탐색 장치.

청구항 3

제2항에 있어서,

상기 길이 조건은 주어진 서열의 길이가 n일 경우 $i-s \geq s+1$ 및 $n-i-s-1 \geq s+1$ 인 것을 특징으로 하는 유사매듭구조 탐색 장치.

청구항 4

제2항에 있어서,

상기 길이 조건은 주어진 서열의 길이가 n일 경우 $i-s \geq s+1$ 및 $n-i-s-1 \geq s+1$ 인 것을 특징으로 하는 유사매듭구조 탐색 장치.

청구항 5

제4항에 있어서,

i 위치에서 상기 연속적인 상보적 결합 조건은 만족하나 상기 길이 조건을 만족하지 않을 경우, s-1 위치에서의 넥스트 함수값을 이용하여 상기 s의 사이클을 축소시킨 후 상기 연속적인 상보적 결합 조건과 상기 길이 조건을 판단하며, 상기 넥스트 함수는 i에서 끝나는 서브스트링 중 주어진 서열 w의 전위부가 되는 가장 긴 서브스트링의 길이인 것을 특징으로 하는 유사매듭구조 탐색 장치.

청구항 6

제5항에 있어서,

i 위치에서 상기 연속적인 상보적 결합 조건은 만족하나 상기 길이 조건을 만족하지 않을 경우, 상보적 결합 조건을 마지막으로 만족했던 p 를 찾아내 $c'[i]=p+1$ 로 설정하고 $p=p+1$ 로 업데이트하는 것을 특징으로 하는 유사매듭구조 탐색 장치.

청구항 7

제1항에 있어서,

상기 유사매듭구조 결정부는, $(d[i], c[i+1])$, $(d'[i], c[i+1])$ 및 $(d[i], c'[i+1])$ 의 조합 중 길이 조건을 만족하면서 u 의 길이 및 x 의 길이의 최소값 중 가장 긴 길이를 가지는 조합을 주어진 서열의 유사매듭구조로 결정하는 것을 특징으로 하는 유사매듭구조 탐색 장치.

청구항 8

주어진 서열로부터 전위부(u)와 상보적으로 결합되는 매칭 전위부의 길이가 후위부(x)와 상보적으로 결합되는 매칭 후위부의 길이보다 짧다고 가정한 상태에서 상기 서열의 위치 i 에서 상보적 결합 조건 및 길이 조건을 만족하는 가장 긴 u 의 길이인 $c[i]$ 와 상보적 결합 조건은 만족하나 길이 조건을 만족하지 못하는 가장 짧은 u 의 길이인 $c'[i]$ 를 출력하는 매칭 전위부 탐색 단계(a);

주어진 서열로부터 후위부(x)와 상보적으로 결합되는 매칭 후위부의 길이가 전위부(u)와 상보적으로 결합되는 매칭 전위부의 길이보다 짧다고 가정한 상태에서 상기 서열의 위치 i 에서 상보적 결합 조건 및 길이 조건을 만족하는 가장 긴 x 의 길이인 $d[i]$ 와 상보적 결합 조건은 만족하나 길이 조건을 만족하지 못하는 가장 짧은 x 의 길이인 $d'[i]$ 를 출력하는 매칭 후위부 탐색 단계(b); 및

상기 $c[i]$, $c'[i]$, $d[i]$ 및 $d'[i]$ 를 이용하여 상기 전위부 및 후위부의 길이를 결정하는 유사매듭구조 결정 단계(c)를 포함하는 것을 특징으로 하는 유사매듭구조 탐색 방법.

청구항 9

제8항에 있어서,

상기 단계(a)는 $i=(n-1)$ 부터 감소시키면서 상기 $c[i]$ 및 $c'[i]$ 를 연산하며, $i+1$ 위치에서 상보적 결합이 있다고 판단되는 경우 상기 $c[i]$ 및 상기 $c'[i]$ 는 $s=c[i+1]$ 및 $p=c'[i+1]$ 에 기초하여 연산되는 것을 특징으로 하는 유사매듭구조 탐색 방법.

청구항 10

제9항에 있어서,

상기 길이 조건은 주어진 서열의 길이가 n 일 경우 $i-s \geq s+1$ 및 $n-i-s-1 \geq s+1$ 인 것을 특징으로 하는 유사매듭구조 탐색 방법.

청구항 11

제10항에 있어서,

i 위치에서 연속적인 상보적 결합 조건 및 상기 길이 조건을 만족할 경우 $c[i]=s+1$ 로 설정되고, $s=s+1$ 로 업데이트되는 것을 특징으로 하는 유사매듭구조 탐색 방법.

청구항 12

제11항에 있어서,

i 위치에서 상기 연속적인 상보적 결합 조건은 만족하나 상기 길이 조건을 만족하지 않을 경우, s-1 위치에서의 넥스트 함수값을 이용하여 상기 s의 사이즈를 축소시킨 후 상기 연속적인 상보적 결합 조건과 상기 길이 조건을 판단하며, 상기 넥스트 함수는 i에서 끝나는 서브스트링 중 주어진 서열 w의 전위부가 되는 가장 긴 서브스트링의 길이인 것을 특징으로 하는 유사매듭구조 탐색 방법.

청구항 13

제12항에 있어서,

i 위치에서 상기 연속적인 상보적 결합 조건은 만족하나 상기 길이 조건을 만족하지 않을 경우, 상보적 결합 조건을 마지막으로 만족했던 p를 찾아내 c'[i]=p+1로 설정하고 p=p+1로 업데이트하는 것을 특징으로 하는 유사매듭구조 탐색 방법.

청구항 14

제8항에 있어서,

상기 단계(c)는 (d[i], c[i+1]), (d'[i], c[i+1]) 및 (d[i], c'[i+1])의 조합 중 길이 조건을 만족하면서 u의 길이 및 x의 길이의 최소값 중 가장 긴 길이를 가지는 조합을 주어진 서열의 유사매듭구조로 결정하는 것을 특징으로 하는 유사매듭구조 탐색 방법.

청구항 15

제8항 내지 제14항 중 어느 한 항의 방법을 실행하기 위한 프로그램이 기록되어 있으며 컴퓨터에 의한 판독 가능한 프로그램이 기록된 기록매체.

발명의 설명

기술 분야

[0001] 본 발명은 유사매듭구조 탐색 장치 및 방법에 관한 것이다.

배경 기술

[0003] 생물학 분야에서 DNA 및 RNA와 같은 생물학적 서열이 유사매듭 구조에 해당되는지 여부는 매우 중요한 연구 중 하나로 생물학 전문가뿐만 아니라 컴퓨터 분야에서도 많은 전문가들이 활발히 연구하고 있는 분야이다.

[0004] 실제로 유사매듭 구조는 RNA 서열에서 빈번히 존재하고 바이러스를 표현하는 프레임쉬프트 돌연변이와도 연관이 있다. 이러한 이유로 생물학 및 컴퓨터공학 연구자들이 이론적, 실험적 접근을 통해 유사매듭 구조를 탐색하는 기법을 활발히 연구하고 있다

[0005] 유사매듭구조는 생물학적 서열에서 상보적으로 결합하는 두 개의 hairpin 루프가 서로 교차하는 형태를 의미한다.

[0006] 도 1은 일반적인 유사매듭 구조의 일례를 나타낸 도면이다.

[0007] 도 1을 참조하면, 유사매듭구조는 생물학적 서열에서 상보적으로 결합하는 두 개의 hairpin 루프가 서로 교차하

는 형태를 의미한다는 것을 알 수 있으며, 도 1의 (a)는 서로 결합한 형태를 나타낸 것이고 도 1의 (b)는 유사 매듭구조를 명확히 확인할 수 있도록 1차원으로 배열된 서열에서 유사매듭 구조를 나타낸 것이다.

[0008] 유사매듭 구조에서 전위부(Prefix)와 상보적으로 결합되는 매칭 전위부가 존재하고, 후위부(Suffix)와 상보적으로 결합되는 매칭 후위부가 존재한다. 상보적 결합은 전위부의 서열이 매칭 전위부의 순서를 역전시킨 상태에서 순차적으로 결합 관계에 있는 것을 의미한다. 예를 들어, 상보적 변환을 나타내는 연산자를 θ 라고 할 때 $\theta(\text{ACTG})=\text{CAGT}$ 로 표현할 수 있으며, $\theta(\text{ATT})=\text{AAT}$ 로 표현될 수 있다.

[0009] 이러한 유사매듭구조는 $w = uv\theta(x)\theta(u)yx$ 로 정의될 수 있다. u 는 전위부이고, $\theta(u)$ 는 전위부와 상보적으로 결합되는 매칭 전위부이며, x 는 후위부이고, $\theta(x)$ 는 후위부와 상보적으로 결합되는 매칭 후위부이다. 매칭 전위부와 매칭 후위부는 서로 인접하고 매칭 후위부가 매칭 전위부에 비해 전단에 위치하게 된다.

[0010] 주어진 서열이 유사매듭구조에 해당되는지 여부 및 유사매듭구조의 전위부 및 후위부의 길이를 결정하는 것은 매우 복잡한 연산이고, 주어진 서열의 길이가 늘어날수록 유사매듭구조인지 판단하기 위한 연산량은 기하급수적으로 증가한다.

발명의 내용

해결하려는 과제

[0012] 본 발명은 간소화된 연산으로 유사매듭구조를 탐색할 수 있는 방법 및 장치를 제안한다.

과제의 해결 수단

[0014] 상기한 목적을 달성하기 위해 본 발명의 일 실시예에 따르면, 주어진 서열로부터 전위부(u)와 상보적으로 결합되는 매칭 전위부의 길이가 후위부(x)와 상보적으로 결합되는 매칭 후위부의 길이보다 짧다고 가정한 상태에서 상기 서열의 위치 i 에서 상보적 결합 조건 및 길이 조건을 만족하는 가장 긴 u 의 길이인 $c[i]$ 와 상보적 결합 조건은 만족하나 길이 조건을 만족하지 못하는 가장 짧은 u 의 길이인 $c'[i]$ 를 출력하는 매칭 전위부 탐색부; 주어진 서열로부터 후위부(x)와 상보적으로 결합되는 매칭 후위부의 길이가 전위부(u)와 상보적으로 결합되는 매칭 전위부의 길이보다 짧다고 가정한 상태에서 상기 서열의 위치 i 에서 상보적 결합 조건 및 길이 조건을 만족하는 가장 긴 x 의 길이인 $d[i]$ 와 상보적 결합 조건은 만족하나 길이 조건을 만족하지 못하는 가장 짧은 x 의 길이인 $d'[i]$ 를 출력하는 매칭 후위부 탐색부; 및 상기 $c[i]$, $c'[i]$, $d[i]$ 및 $d'[i]$ 를 이용하여 상기 전위부 및 후위부의 길이를 결정하는 유사매듭구조 결정부를 포함하는 유사매듭구조 탐색 장치가 제공된다.

[0015] 상기 매칭 전위부 탐색부는 $i=(n-1)$ 부터 감소시키면서 상기 $c[i]$ 및 $c'[i]$ 를 연산하며, $i+1$ 위치에서 상보적 결합이 있다고 판단되는 경우 상기 $c[i]$ 및 상기 $c'[i]$ 는 $s=c[i+1]$ 및 $p=c'[i+1]$ 에 기초하여 연산된다.

[0016] 상기 길이 조건은 주어진 서열의 길이가 n 일 경우 $i-s \geq s+1$ 및 $n-i-s-1 \geq s+1$ 이다.

[0017] i 위치에서 연속적인 상보적 결합 조건 및 상기 길이 조건을 만족할 경우 $c[i]=s+1$ 로 설정되고, $s=s+1$ 로 업데이트된다.

[0018] i 위치에서 상기 연속적인 상보적 결합 조건은 만족하나 상기 길이 조건을 만족하지 않을 경우, $s-1$ 위치에서의 넥스트 함수값을 이용하여 상기 s 의 사이즈를 축소시킨 후 상기 연속적인 상보적 결합 조건과 상기 길이 조건을 판단하며, 상기 넥스트 함수는 i 에서 끝나는 서브스트링 중 주어진 서열 w 의 전위부가 되는 가장 긴 서브스트링의 길이이다.

[0019] i 위치에서 상기 연속적인 상보적 결합 조건은 만족하나 상기 길이 조건을 만족하지 않을 경우, 상보적 결합 조건을 마지막으로 만족했던 p 를 찾아내 $c'[i]=p+1$ 로 설정하고 $p=p+1$ 로 업데이트된다.

[0020] 상기 유사매듭구조 결정부는, $(d[i], c[i+1])$, $(d'[i], c[i+1])$ 및 $(d[i], c'[i+1])$ 의 조합 중 길이 조건을 만족하면서 u 의 길이 및 x 의 길이의 최소값 중 가장 긴 길이를 가지는 조합을 주어진 서열의 유사매듭구조로 결정한다.

[0021] 본 발명의 다른 측면에 따르면, 주어진 서열로부터 전위부(u)와 상보적으로 결합되는 매칭 전위부의 길이가 후

위부(x)와 상보적으로 결합되는 매칭 후위부의 길이보다 짧다고 가정한 상태에서 상기 서열의 위치 i에서 상보적 결합 조건 및 길이 조건을 만족하는 가장 긴 u의 길이인 c[i]와 상보적 결합 조건은 만족하나 길이 조건을 만족하지 못하는 가장 짧은 u의 길이인 c'[i]를 출력하는 매칭 전위부 탐색 단계(a); 주어진 서열로부터 후위부(x)와 상보적으로 결합되는 매칭 후위부의 길이가 전위부(u)와 상보적으로 결합되는 매칭 전위부의 길이보다 짧다고 가정한 상태에서 상기 서열의 위치 i에서 상보적 결합 조건 및 길이 조건을 만족하는 가장 긴 x의 길이인 d[i]와 상보적 결합 조건은 만족하나 길이 조건을 만족하지 못하는 가장 짧은 x의 길이인 d'[i]를 출력하는 매칭 후위부 탐색 단계(b); 및 상기 c[i], c'[i], d[i] 및 d'[i]를 이용하여 전위부 및 후위부의 길이를 결정하는 유사매듭구조 결정 단계(c)를 포함하는 유사매듭구조 탐색 방법이 제공된다.

발명의 효과

[0023] 본 발명에 따르면, 간소화된 연산으로 유사매듭구조를 탐색할 수 있는 장점이 있다.

도면의 간단한 설명

[0025] 도 1은 일반적인 유사매듭 구조의 일례를 나타낸 도면.

도 2는 본 발명의 유사매듭구조 탐색을 위한 예시적 유사매듭 구조를 나타낸 도면.

도 3은 본 발명의 일 실시예에 따른 유사매듭 구조 탐색 장치의 전체적인 구조를 도시한 도면.

도 4는 c[i+1]에 기초하여 c[i]를 결정하는 일례를 나타낸 도면.

도 5는 본 발명에서 s의 길이를 축소시키기 위해 활용하는 넥스트 함수를 나타낸 도면.

도 6은 넥스트 함수를 이용하여 s의 길이를 축소시키는 일례를 나타낸 도면.

도 7은 본 발명의 일 실시예에 따른 유사매듭구조 판단 방법의 전체적인 흐름을 도시한 순서도,

발명을 실시하기 위한 구체적인 내용

[0026] 본 발명은 다양한 변경을 가할 수 있고 여러 가지 실시예를 가질 수 있는 바, 특정 실시예들을 도면에 예시하고 상세한 설명에 상세하게 설명하고자 한다. 그러나, 이는 본 발명을 특정한 실시 형태에 대해 한정하려는 것이 아니며, 본 발명의 사상 및 기술 범위에 포함되는 모든 변경, 균등물 내지 대체물을 포함하는 것으로 이해되어야 한다. 각 도면을 설명하면서 유사한 참조부호를 유사한 구성요소에 대해 사용하였다.

[0027] 이하에서, 본 발명에 따른 실시예들을 첨부된 도면을 참조하여 상세하게 설명한다.

[0028] 도 2는 본 발명의 유사매듭구조 탐색을 위한 예시적 유사매듭 구조를 나타낸 도면이다.

[0029] 앞서 설명한 바와 같이, 주어진 서열을 w라고 할 때 유사매듭구조는 $w = uv\theta(x)\theta(u)yx$ 로 정의되며, u는 전위부(Prefix)이고 $\theta(u)$ 는 전위부와 상보적 결합부(이하, "매칭 전위부")이며, x는 후위부(Suffix)이고 $\theta(x)$ 는 후위부와 상보적 결합부(이하, "매칭 후위부")이다. v와 y는 유사매듭구조와 관련없는 문자열들이다.

[0030] 본 발명은 전위부와 상보적으로 결합되는 매칭 전위부를 탐색하여 매칭 전위부의 길이를 출력하고, 후위부와 상보적으로 결합되는 매칭 후위부를 탐색하여 매칭 후위부의 길이를 출력한 후 최종적으로 유사매듭 구조를 결정한다.

[0031] 이때, 매칭 전위부를 탐색하는 방법 및 매칭 후위부를 탐색하는 방법은 동일하므로 본 실시예에서는 매칭 전위부를 탐색하는 방법에 대해 설명하기로 한다.

[0032] 이를 위해 도 2에는 전위부인 "AGACAGA"와 이와 상보적으로 결합되는 "TCTGTCT"가 도시되어 있다.

[0033] 도 3은 본 발명의 일 실시예에 따른 유사매듭 구조 탐색 장치의 전체적인 구조를 도시한 도면이다.

[0034] 도 3을 참조하면, 본 발명의 일 실시예에 따른 유사매듭 구조 탐색 장치는 매칭 전위부 탐색부(300), 매칭 후위부 탐색부(302) 및 유사매듭구조 결정부(304)를 포함한다.

- [0035] 매칭 전위부 탐색부(300)는 후보가 될 수 있는 매칭 전위부를 탐색하고 매칭 전위부의 후보 길이 정보를 출력한다. 구체적으로, 매칭 전위부 탐색부(300)는 후보가 될 수 있는 두 개의 전위부 길이($c[i]$ 및 $c'[i]$)를 출력한다.
- [0036] 탐색 시점에는 매칭 전위부와 매칭 후위부 중 무엇이 더 긴 길이를 가지는지 알 수 없다. 본 실시예에서는 매칭 전위부를 탐색할 경우에는 매칭 전위부의 최대 길이가 매칭 후위부의 최대 길이보다 짧다는 가정 하에서 탐색을 수행한다. 물론, 이러한 가정은 틀릴 수도 있는 것이기에 매칭 전위부 탐색부(300)는 후보가 될 수 있는 두 개의 길이를 출력하게 된다.
- [0037] $t = \max(\min(|u|, |x|))$ 로 정의한다. 여기서, $|u|$ 는 전위부의 길이를 의미하고, $|x|$ 는 후위부의 길이를 의미한다. 매칭 전위부 탐색부(300)는, $t = |u|$ 로 가정한 상태에서 전위부와 상보적 결합부를 탐색한다. $t = |u|$ 로 가정할 경우 $t \geq |u|$ 이므로, u 와 $\Theta(u)$ 사이 그리고, $\Theta(u)$ 의 뒤쪽으로는 최소한 $|u|$ 의 공간이 확보되어야 한다는 점을 확인할 수 있다.
- [0038] 매칭 전위부 탐색부(300)는 주어진 서열의 첫번째 문자의 위치를 0이라고 하고 마지막 문자의 위치를 $n-1$ 이라고 할 때 각각이 위치 i 에서 상보적 결합 조건 및 길이 조건을 만족하는 가장 긴 u 의 길이인 $c[i]$ 와 상보적 결합 조건은 만족하나 길이 조건을 만족하지 못하는 가장 짧은 u 의 길이인 $c'[i]$ 를 각각 출력한다.
- [0039] 매칭 전위부 탐색부는 i 를 $(n-1)$ 부터 점차 감소시키면서 $c[i]$ 및 $c'[i]$ 를 연산한다. 만일 임의의 위치 i 에서 전위부의 첫번째 문자와 상보적으로 결합되지 않는다면 $c[i]=0$ 이며, $c'[i]$ 역시 0이다.
- [0040] 만일 $i+1$ 위치에서 상보적 결합이 있다고 판단될 경우, $c[i]$ 및 $c'[i]$ 는 $c[i+1]$ 및 $c'[i+1]$ 에 기초하여 정해진다. $c[i+1]=s$ 이고 $c'[i+1]=p$ 라고 정의한다.
- [0041] 도 2에 도시된 예에서, 매칭 전위부의 마지막 문자인 T 의 위치가 $i+1$ 이라고 하고 매칭 전위부의 마지막에서 두번째 문자인 C 의 위치가 i 라고 가정한다. 이때, $c[i+1]=1$ 이고, $c'[i+1]=1$ 이다. 이는 T 가 전위부의 첫번째 문자인 A 와 상보적으로 결합되기 때문이다. 결국, $s=1$ 이고 $p=1$ 이다.
- [0042] $(i+1)$ 위치에서 상보적 결합이 있는데 i 에서 연속적으로 상보적 결합이 있고 길이 조건을 만족한다면 $c[i]=s+1$ 로 설정되고, $s=s+1$ 로 업데이트된다. 따라서, $c[i]=2$ 이다. 길이조건에 문제가 없는 경우 $c'[i]=2$ 이다. 도 2의 경우 마지막에서 두번째 문자인 "C"역시 상보적으로 결합하므로 $c[i]=2$ 이고, $c'[i]=2$ 이다.
- [0043] 도 4는 $c[i+1]$ 에 기초하여 $c[i]$ 를 결정하는 일례를 나타낸 도면이고, 앞서 설명한 방법이 도시되어 있다.
- [0044] 도 4를 참조하면, 연속적인 상보적 결합이 만족하는 상태가 도시되어 있으며, 노란색 박스가 서로 상보적으로 결합된다는 것을 의미한다. 이는 $w[i]=\Theta(w[s])$ 의 관계를 만족한다는 것을 의미한다. $c[i]$ 값이 $s+1$ 이 되려면, 길이 조건 역시 만족하여야 한다.
- [0045] 도 4를 참조하면, 길이 조건은 두 개이며, $i-s \geq s+1$ 그리고 $n-i-s-1 \geq s+1$ 이다.
- [0046] 매칭 전위부 탐색부(300)는 i 에서 상보적 결합 조건은 만족하나 위의 두 개의 길이 조건을 만족하지 않을 경우, $c[i+1]$ 인 s 를 축소시킨 후 다시 상보적 결합 여부를 판단한다.
- [0047] 이때, s 의 축소는 넥스트 함수를 통해 이루어진다. 넥스트 함수를 $a[i]$ 라고 정의할 때 $a[i]$ 는 i 에서 끝나는 서브스트링(substring) 중 주어진 서열 w 의 전위부가 되는 가장 긴 서브스트링의 길이로 정의된다. 만일 $abab$ 가 서브 스트링일 경우 마지막 문자인 b 의 위치에서의 $a[3]=2$ 로 설정된다. 이는 첫번째와 두번째 문자 ab 와 세번째 및 네번째 문자 ab 가 두개씩 일치하기 때문이다.
- [0048] 도 5는 본 발명에서 s 의 길이를 축소시키기 위해 활용하는 넥스트 함수를 나타낸 도면이다.
- [0049] 도 5를 참조하면, 서로 일치하는 두 개의 문자열이 전단부 및 후단부에 위치하고 일치하는 문자열의 최대 수가 $a[i]$ 라는 것을 확인할 수 있다.
- [0050] 도 2의 주어진 예제를 다시 참조하면, 매칭 전위부의 마지막 문자인 T 의 $s=1$ 이고, 마지막에서 두번째 문자인 C 의 $s=2$ 이며, 상보적 결합 관계가 계속적으로 만족하므로 매칭 전위부의 마지막에서 5번째 문자인 T 에 대한 $s=5$ 로 설정된다.
- [0051] 그런데, 매칭 전위부의 마지막에서 6번째 문자인 C 에서 연속적인 상보적 결합은 만족하나 길이 조건을 만족하지 않는다고 가정한다. 이 경우에 넥스트 함수를 통한 s 의 사이즈 축소를 수행하며, 또한 이 경우에 $c[i]$ 와 $c'[i]$

가 불일치하는 경우가 발생한다.

- [0052] "C"가 마지막에서 6번째 문자이므로 $i=6$ 에서 넥스트 함수를 적용하며, 이는 서브 스트링 "AGACAG"에서 마지막 문자 G 위치에서의 넥스트 함수의 값을 획득하게 되며, 이는 $a[s-1]$ 을 획득하는 것을 의미한다. 이때 전단부의 AG와 후단부의 AG가 서로 일치하므로 $a[s-1]$ 은 2며, s 를 5에서 2로 축소시킨다. s 가 2로 축소되었으므로 마지막에서 6번째 문자인 "C"의 $c[i]=3$ 이 된다.
- [0053] 도 6은 넥스트 함수를 이용하여 s 의 길이를 축소시키는 일례를 나타낸 도면이다.
- [0054] 도 6에 도시된 바와 같이, $a[s-1]$ 의 값에 기초하여 s 의 사이즈를 축소시킨 후 i 번째 위치에서의 $c[i]$ 를 연산하게 된다.
- [0055] 이와 같이 길이 조건은 만족하지 않으나 상보적 결합 조건이 만족할 경우 $c[i]$ 와 $c'[i]$ 값이 달라지게 된다. 이 경우 $w[i]=\Theta(w[p])$ 를 만족했던 p 를 찾아내 $c'[i]=p+1$ 로 설정하고 $p=p+1$ 로 업데이트한다. 결국, 매칭 전위부의 마지막에서 6번째 문자인 "C"에서의 $c'[i]=6$ 이 된다. 동일한 방법으로 매칭 전위부의 첫번째 문자인 "T"의 $c[i]=4$ 이고, $c'[i]=7$ 이다.
- [0056] 만일, 넥스트 함수에 의해 s 의 사이즈를 줄인 경우에도 여전히 길이 조건을 만족하지 않을 경우 다시 한번 넥스트 함수를 적용하여 길이를 계속적으로 s 의 사이즈를 계속적으로 줄일 수 있을 것이다.
- [0057] 매칭 전위부 탐색부(300)는 가장 큰 값을 가지는 $c[i]$ 및 $c'[i]$ 를 최종 결과로 출력한다.
- [0058] 한편, 매칭 후위부 탐색부(302)는 매칭 전위부 탐색부(300)와 동일한 방법으로 매칭 후위부를 탐색하며, 이때 $c[i]$ 와 동일한 $d[i]$ 및 $c'[i]$ 와 동일한 $d'[i]$ 를 동일한 방법으로 연산한다.
- [0059] 매칭 후위부 탐색부(302)는 $t=|x|$ 로 가정한 상태에서 $d[i]$ 및 $d'[i]$ 를 연산하며, 가장 큰 값을 가지는 $d[i]$ 및 $d'[i]$ 를 출력한다. .
- [0060] 매칭 후위부 및 매칭 전위부로부터 가장 큰 값을 가지는 $c[i]$, $c'[i]$, $d[i]$ 및 $d'[i]$ 가 각각 출력되면, 유사매듭구조 결정부(304)는 최종적인 유사매듭 구조를 판단한다. 상정 가능한 유사 매듭 구조는 $(d[i], c[i+1])$, $(d'[i], c[i+1])$ 및 $(d[i], c'[i+1])$ 이다. 이와 같은 구조 중 길이 조건을 만족하면서 u 의 길이 및 x 의 길이의 최소값 중 가장 긴 길이를 가지는 조합을 주어진 서열의 유사매듭구조로 결정한다. 결국, t 값이 가장 큰 길이를 가지는 구조를 주어진 서열에 대한 유사매듭 구조로 결정하는 것이다.
- [0061] 도 7은 본 발명의 일 실시예에 따른 유사매듭구조 판단 방법의 전체적인 흐름을 도시한 순서도이다.
- [0062] 도 7을 참조하면, $t=|u|$ 로 t 의 길이를 가정한다(단계 700).
- [0063] $t=|u|$ 로 가정한 상태에서 상보적 결합 조건 및 길이 조건을 만족하는 가장 긴 u 의 길이인 $c[i]$ 와 상보적 결합 조건은 만족하나 길이 조건을 만족하지 못하는 가장 짧은 u 의 길이인 $c'[i]$ 를 연산한다(단계 702).
- [0064] 또한, $t=|x|$ 로 t 의 길이를 가정한다(단계 704).
- [0065] $t=|x|$ 가정한 상태에서 상보적 결합 조건 및 길이 조건을 만족하는 가장 긴 x 의 길이인 $d[i]$ 와 상보적 결합 조건은 만족하나 길이 조건을 만족하지 못하는 가장 짧은 u 의 길이인 $d'[i]$ 를 연산한다(단계 706).
- [0066] 단계 702 및 단계 704에서 연산되는 $c[i]$, $c'[i]$, $d[i]$ 및 $d'[i]$ 를 이용하여 유사매듭 구조를 결정한다(단계 708).
- [0067] 앞서 설명한 바와 같이, $(d[i], c[i+1])$, $(d'[i], c[i+1])$ 및 $(d[i], c'[i+1])$ 의 조합 중 길이 조건을 만족하면서 u 의 길이 및 x 의 길이의 최소값 중 가장 긴 길이를 가지는 조합을 주어진 서열의 유사매듭구조로 결정한다.
- [0068] 이상에서 설명한 바와 같은 본 발명의 유사매듭구조 탐색 장치 및 방법은 기존의 연구들과 비교할 때 단순한 연산으로 유사매듭구조를 결정할 수 있는 장점이 있다.
- [0069] 앞서 설명한 기술적 내용들은 다양한 컴퓨터 수단을 통하여 수행될 수 있는 프로그램 명령 형태로 구현되어 컴퓨터 판독 가능 매체에 기록될 수 있다. 상기 컴퓨터 판독 가능 매체는 프로그램 명령, 데이터 파일, 데이터 구조 등을 단독으로 또는 조합하여 포함할 수 있다. 상기 매체에 기록되는 프로그램 명령은 실시예들을 위하여 특별히 설계되고 구성된 것들이거나 컴퓨터 소프트웨어 당업자에게 공지되어 사용 가능한 것일 수도 있다. 컴퓨터 판독 가능 기록 매체의 예에는 하드 디스크, 플로피 디스크 및 자기 테이프와 같은 자기 매체(magnetic

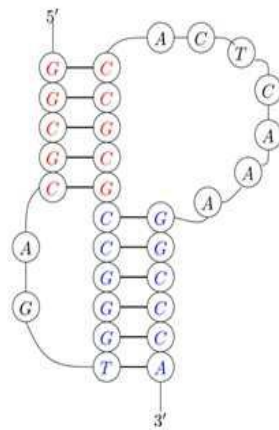
media), CD-ROM, DVD와 같은 광기록 매체(optical media), 플로피컬 디스크(floptical disk)와 같은 자기-광 매체(magneto-optical media), 및 롬(ROM), 램(RAM), 플래시 메모리 등과 같은 프로그램 명령을 저장하고 수행하도록 특별히 구성된 하드웨어 장치가 포함된다. 프로그램 명령의 예에는 컴파일러에 의해 만들어지는 것과 같은 기계어 코드뿐만 아니라 인터프리터 등을 사용해서 컴퓨터에 의해서 실행될 수 있는 고급 언어 코드를 포함한다. 하드웨어 장치는 실시예들의 동작을 수행하기 위해 하나 이상의 소프트웨어 모듈로서 작동하도록 구성될 수 있으며, 그 역도 마찬가지이다.

[0071]

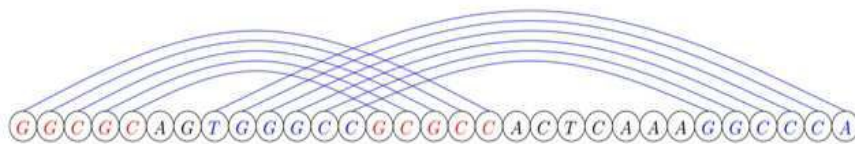
이상과 같이 본 발명에서는 구체적인 구성 요소 등과 같은 특정 사항들과 한정된 실시예 및 도면에 의해 설명되었으나 이는 본 발명의 보다 전반적인 이해를 돕기 위해서 제공된 것일 뿐, 본 발명은 상기의 실시예에 한정되는 것은 아니며, 본 발명이 속하는 분야에서 통상적인 지식을 가진 자라면 이러한 기재로부터 다양한 수정 및 변형이 가능하다. 따라서, 본 발명의 사상은 설명된 실시예에 국한되어 정해져서는 아니되며, 후술하는 특허청구범위뿐 아니라 이 특허청구범위와 균등하거나 등가적 변형이 있는 모든 것들은 본 발명 사상의 범주에 속한다고 할 것이다.

도면

도면1

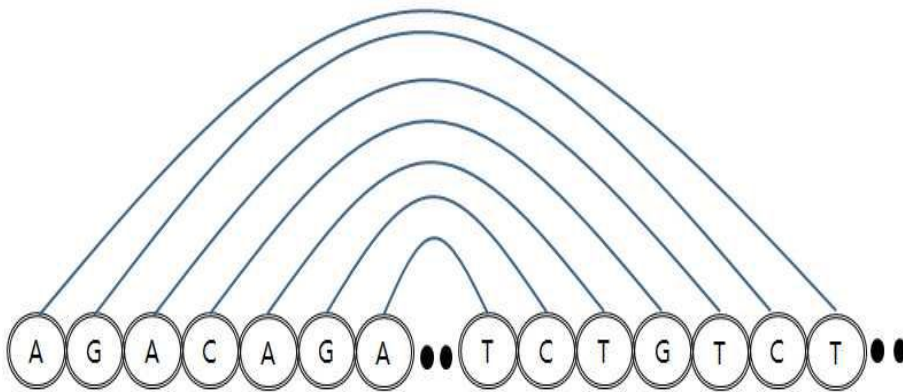


(a)

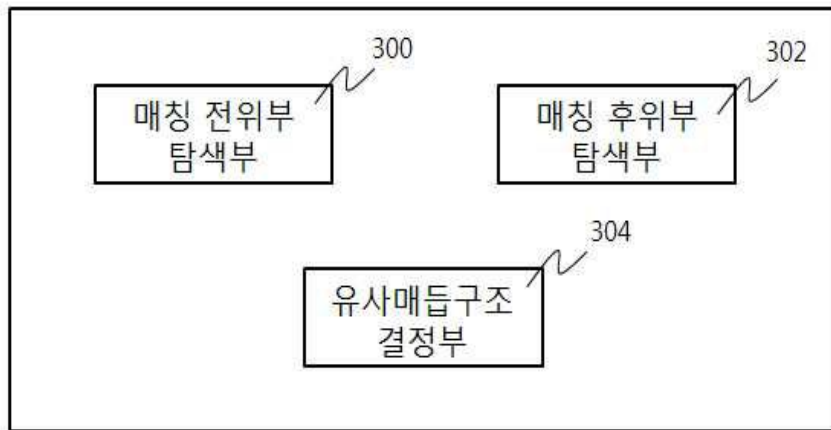


(b)

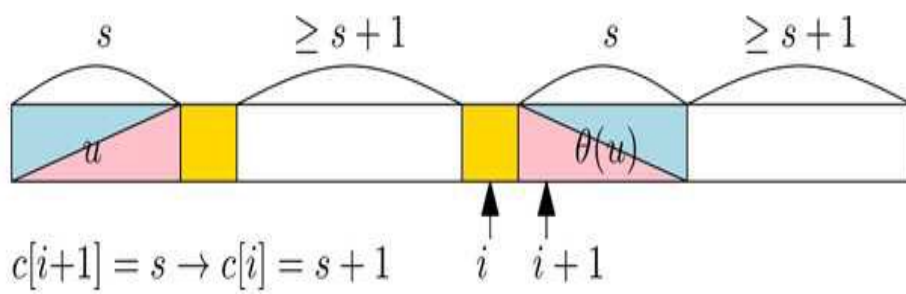
도면2



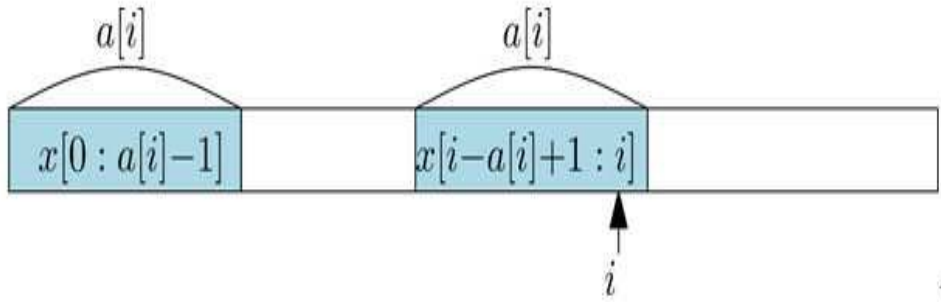
도면3



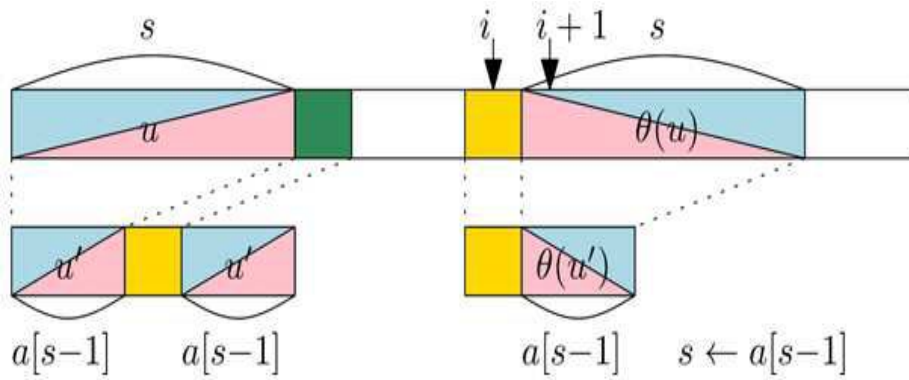
도면4



도면5



도면6



도면7

