Watson - Crick automata (WK automata)

$M = (V, \rho, K, s_0, F, \delta)$, ahol

V egy ábécé

K állapotok halmaza, véges nem-üres halmaz

$V \cap K = \emptyset$

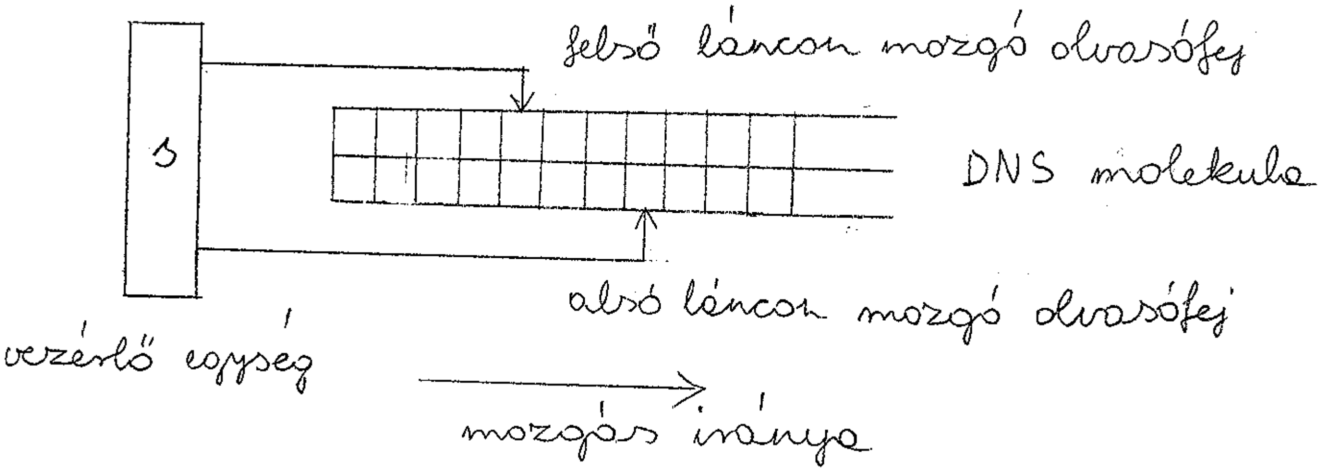
$s_0 \in K$ kezdőállapot

$F \subseteq K$ végállapotok halmaza

$\rho \subseteq V \times V$ szimmetrikus reláció

$\delta : K \times \begin{pmatrix} V^* \\ V^* \end{pmatrix} \to \mathcal{P}(K)$, $\delta(s, \begin{pmatrix} x \\ y \end{pmatrix}) \neq \emptyset$ csak véges sok $(s, x, y) \in K \times V^* \times V^*$

Ha $s' \in \delta(s, \begin{pmatrix} x_1 \\ x_2 \end{pmatrix})$, akkor $s \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \to \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} s' \in P$ jelölést is alkalmazzuk.



Legyen

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}, \begin{pmatrix} u_1 \\ u_2 \end{pmatrix}, \begin{pmatrix} w_1 \\ w_2 \end{pmatrix} \in \begin{pmatrix} V^* \\ V^* \end{pmatrix}, \quad \begin{bmatrix} x_1 u_1 w_1 \\ x_2 u_2 w_2 \end{bmatrix} \in WK_\rho(V),$$

$$s, s' \in K, \quad s \begin{pmatrix} u_1 \\ u_2 \end{pmatrix} \to \begin{pmatrix} u_1 \\ u_2 \end{pmatrix} s' \in P, \text{ akkor}$$

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} s \begin{pmatrix} u_1 \\ u_2 \end{pmatrix} \begin{pmatrix} w_1 \\ w_2 \end{pmatrix} \Rightarrow \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \begin{pmatrix} u_1 \\ u_2 \end{pmatrix} s' \begin{pmatrix} w_1 \\ w_2 \end{pmatrix} \quad \text{közvetlen átmenet}$$

$\overset{*}{\Rightarrow}$: a $\Rightarrow$ refl. és tranz. lezártja

$$L_u(M) = \{ w_1 \in V^* \mid \exists\, w_2 \in V^*, \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} \in WK_\rho(V), \quad s_0 \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} \overset{*}{\Rightarrow} \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} s_f, \ s_f \in F \}$$

(u - <u>u</u>pper strand utal)

WK automata speciális esetei :

- egy-állapotú, ha $K = F = \{s_0\}$
- végállapotos, ha $K = F$
- egyszerű, ha $\forall\ s \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \to \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} s' \in P$ esetén

  vagy $x_1 = \lambda$ vagy $x_2 = \lambda$
- 1-korlátos, ha $\forall\ s \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \to \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} s' \in P$ esetén

  $|x_1 x_2| = 1$

Ha M egy-állapotú, akkor K, s_0, F elhagyható, vagyis $M = (V, \rho, P)$

$AWK(u) = \{ L_u(M) \mid M$ tetszőleges WK automata$\}$

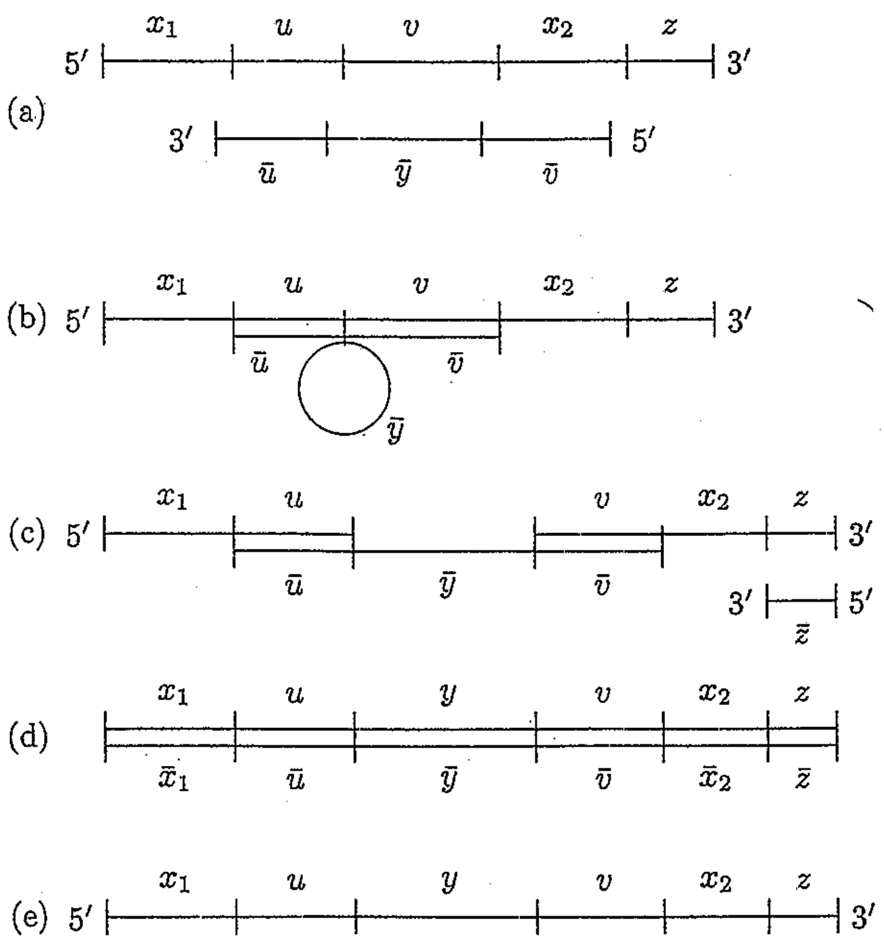
$1WK(u) = \{ L_u(M) \mid M$ tetszőleges 1-korlátos WK automata$\}$

$AWK(u) = 1WK(u)$

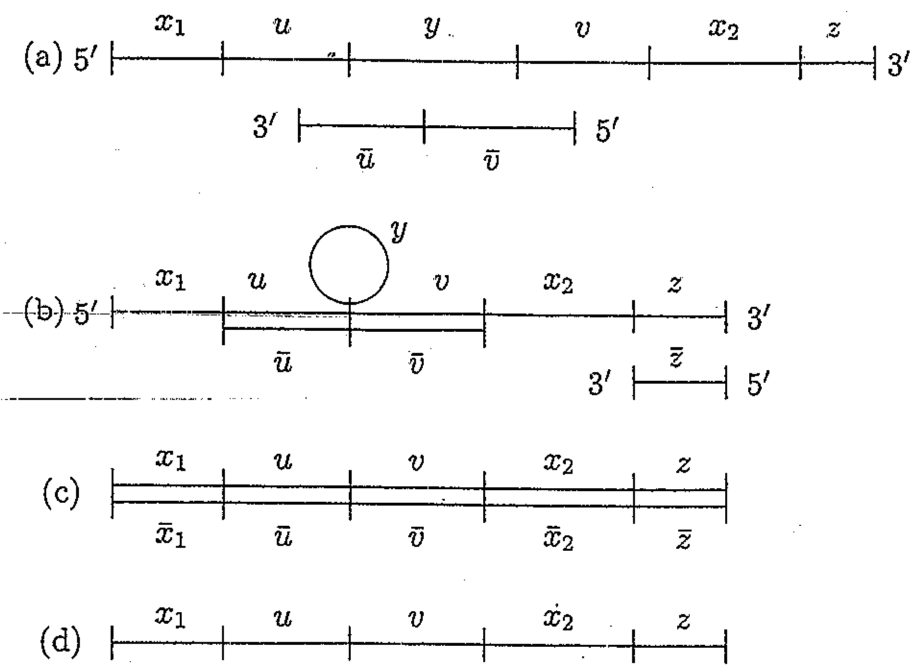
$AWK(u) \subset CS$

(pl: $L = \{ a^{2^n} \mid n \geq 1 \}$, $L \in CS - CF$,
$AWK(u)$ minden egy-betűs nyelve reguláris, így $L \notin AWK(u)$

Beszúró - törlő rendszer



y beszúrása az $5'-x_1 u v x_2 z-3'$ láncba az u és v közé.



y törlése az $5'-x_1 u y v x_2 z-3'$ láncból az u és v közül

Insdel rendszer

$\gamma = (V, T, A, R)$

V ábécé

$T \subseteq V$ terminális szimbólumok halmaza

$A \subseteq V^*$ egy véges nyelv V felett, elemei axiómák

R egy véges halmaz,
elemei $(u, \alpha/\beta, v)$ alakú, ahol $u, v \in V^*$ és
$(\alpha, \beta) \in (V^+ \times \{\lambda\}) \cup (\{\lambda\} \times V^+)$

$(u, \lambda/\beta, v)$ jelentése: u és v közé β beszúrása

$(u, \alpha/\lambda, v)$ jelentése: $u\alpha v$ részből α törlése

Levezetés:

$x, y \in V^*$, $x \Rightarrow y$, ha

1, $x = x_1 u v x_2$, $y = x_1 u \beta v x_2$, $x_1, x_2 \in V^*$, $(u, \lambda/\beta, v) \in R$

2. $x = x_1 u \alpha v x_2$, $y = x_1 u v x_2$, $x_1, x_2 \in V^*$, $(u, \alpha/\lambda, v) \in R$

$\overset{*}{\Rightarrow}$, $\overset{+}{\Rightarrow}$

γ által generált nyelv:

$L(\gamma) = \{ w \in T^* \mid x \overset{*}{\Rightarrow} w$, valamely $x \in A$ esetén $\}$

γ insdel rendszer súlya $(n, m; p, q)$, ahol

$n = \max\{ |\beta| \mid (u, \lambda/\beta, v) \in R \}$,

$m = \max\{ |u| \mid (u, \lambda/\beta, v) \in R$ vagy $(v, \lambda/\beta, u) \in R \}$

$p = \max\{ |\alpha| \mid (u, \alpha/\lambda, v) \in R \}$,

$q = \max\{ |u| \mid (u, \alpha/\lambda, v) \in R$ vagy $(v, \alpha/\lambda, u) \in R \}$

$INS_n^m DEL_p^q = \{ L \subseteq T^* \mid \exists \gamma, (n', m'; p', q')$ súlyú insdel rendszer, $L(\gamma) = L$, $n' \leq n$, $m' \leq m$, $p' \leq p$, $q' \leq q \}$

súlyparaméterben $*$: nincs korlát

$INS_*^* DEL_*^*$: az insdel rendszerek által generált nyelvek családja

INS_0^0 : nincs beszúrás

DEL_0^0 : nincs törlés

<u>Nyelvcsaládok közötti kapcsolatok:</u>

$INS_*^* DEL_*^* = RE$

$INS_3^2 DEL_3^0 = RE$

$INS_1^2 DEL_1^1 = RE$

$INS_1^1 DEL_2^0 = RE$

$REG \subset INS_*^* DEL_0^0$

$FIN \subset INS_*^0 DEL_0^0 \subset INS_*^1 DEL_0^0 \subset \ldots \subset INS_*^* DEL_0^0 \subset CS$

$LIN - INS_*^* DEL_0^0 \neq \emptyset$

$INS_*^1 DEL_0^0 \subset CF$

/ REG nem összehasonlítható $INS_*^m DEL_0^0$, $m \geq 0$,

LIN és CF nem összehasonlítható $INS_*^m DEL_0^0$, $m \geq 2$,

$INS_2^2 DEL_0^0$ tartalmaz nem-féllineáris nyelvet

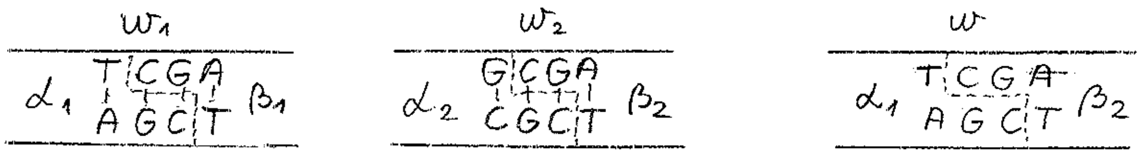
Splicing művelet

- a molekula meghatározott helyen való szétvágása (szétvágó enzimek)

  pl. (T, CG, A) , (G, CG, A)

- a szétvágott molekulák összekapcsolódása az illeszkedő végződések szerint, rekombináció (bázisok közötti H-kötés, ligase enzim- foszfodiészter kötés)

pl.



$r: ((T, CG, A), (G, CG, A))$ másként $(T, CGA; G, CGA)$

$(w_1, w_2) \vdash_r w$

WK komplementaritás → művelet egyszerű láncon

Splicing szabályok alakja egy V ábécé felett:

$u_1 \# u_2 \$ u_3 \# u_4$, ahol $u_1, u_2, u_3, u_4 \in V^*$ és $\#, \$ \notin V$

($u_1 u_2 = \lambda$ vagy $u_3 u_4 = \lambda$ eset kizárható, nem valóságos)

Legyen $x, y, z \in V^*$, $r = u_1 \# u_2 \$ u_3 \# u_4$ egy szabály,

$(x, y) \vdash_r z \iff x = x_1 u_1 u_2 x_2$, $y = y_1 u_3 u_4 y_2$, $z = x_1 u_1 u_4 y_2$, $x_1, x_2, y_1, y_2 \in V^*$

$(x_1 u_1 | u_2 x_2 , y_1 u_3 | u_4 y_2) \vdash_r (x_1 u_1 u_4 y_2)$

(1-es splicing)

H - séma

$\sigma = (V, R)$, ahol

V egy ábécé ,

$R \subseteq V^{*} \# V^{*} \$ V^{*} \# V^{*}$ a splicing szabályok halmaza

Legyen $x, y \in V^{*}$, (x, y) rendezett pár, akkor

$\sigma_1(x,y) = \{ z \in V^{*} \mid (x,y) \vdash_{r} z ,\ r \in R \}$

Legyen $L \subseteq V^{*}$ akkor

$$\sigma_1(L) = \bigcup_{x,y \in L} \sigma_1(x,y)$$

$R \subseteq (V \cup \{\#, \$\})^{*}$ vagyis egy nyelv .

Legyen FL egy nyelvcsalád

$\sigma = (V, R)$ H - séma FL-típusú , ha $R \in FL$

Legyen FL_1 és FL_2 egy - egy nyelvcsalád, akkor

$S_1(FL_1, FL_2) = \{ \sigma_1(L) \mid L \in FL_1 ,\ \sigma = (V, R) ,\ R \in FL_2 \}$

Ha $S_1(FL_1, FL_2) \subseteq FL_1$, akkor az FL_1 nyelvcsalád zárt az FL_2 - típusú H - sémákra .

/ Valóság :
A molekulák és a splicing szabályok halmaza véges. A molekulák tetszőlegesek lehetnek.
Az összeilleszkedő végződések rendszerint 6 - nál nem hosszabbak, vagyis a splicing szabályokban szereplő sztringek korlátos hosszúságúak. /

	FIN	REG	LIN	CF	CS	RE
FIN	FIN	FIN	FIN	FIN	FIN	FIN
REG	REG	REG	REG, LIN	REG, CF	REG, RE	REG, RE
LIN	LIN, CF	LIN, CF	RE	RE	RE	RE
CF	CF	CF	RE	RE	RE	RE
CS	RE	RE	RE	RE	RE	RE
RE	RE	RE	RE	RE	RE	RE

$S_1(FL_1, FL_2)$ nyelvcsaládok mérete

ahol FL_1 a sorok, FL_2 az oszlopok címkéje és a metszéspontokban vagy $S_1(FL_1, FL_2)$ van, vagy FL_3, FL_4 nyelvcsaládok, amelyekre $FL_3 \subset S_1(FL_1, FL_2) \subset FL_4$, és FL_3 és FL_4 a leg-jobb becslések a vizsgált 6 nyelvcsaládban.

Legyen $\sigma = (V, R)$ H-séma, $R \subseteq FIN$, akkor

$rad(\sigma) = \max\{|x| \mid x = u_i, 1 \le i \le 4, u_1 \# u_2 \$ u_3 \# u_4 \in R\}$

(σ H-séma sugara $rad(\sigma)$)

Definiálható $S_1(FL, [p])$ nyelvcsalád $p \ge 1$-re

$S_1(FL, [p]) = \{\sigma_1(L) \mid L \in FL, \sigma \text{ H-séma}, rad(\sigma) \le p\}$

Iterált splicing művelet a nyelveken

$\sigma = (V, R)$ egy H-séma, $L \subseteq V^*$, akkor

$\sigma_1^0(L) = L$,

$\sigma_1^{i+1}(L) = \sigma_1^i(L) \cup \sigma_1(\sigma_1^i(L))$, $i \geq 0$,

$\sigma_1^*(L) = \bigcup_{i \geq 0} \sigma_1^i(L)$.

Legyen FL_1, FL_2 egy-egy nyelvcsalád,

$H_1(FL_1, FL_2) = \{\sigma_1^*(L) \mid L \in FL_1,\ \sigma = (V, R) \text{ H-séma},\ R \in FL_2\}$

$H_1(FL, [p]) = \{\sigma_1^*(L) \mid L \in FL,\ \sigma = (V, R) \text{ H-séma},\ rad(\sigma) \leq p\}$

	FIN	REG	LIN	CF	CS	RE
FIN	FIN, REG	FIN, RE	FIN, RE	FIN, RE	FIN, RE	FIN, RE
REG	REG	REG, RE	REG, RE	REG, RE	REG, RE	REG, RE
LIN	LIN, CF	LIN, RE	LIN, RE	LIN, RE	LIN, RE	LIN, RE
CF	CF	CF, RE	CF, RE	CF, RE	CF, RE	CF, RE
CS	CS, RE	CS, RE	CS, RE	CS, RE	CS, RE	CS, RE
RE	RE	RE	RE	RE	RE	RE

$H_1(FL_1, FL_2)$ nyelvcsaládok mérete

ahol a sorok címkéi FL_1, az oszlopok címkéi FL_2, a sor és oszlop metszéspontjában vagy $H_1(FL_1, FL_2)$, vagy FL_3, FL_4 nyelvcsaládok vannak, amelyekre $FL_3 \subset H_1(FL_1, FL_2) \subset FL_4$ és FL_3, FL_4 a legjobb becslések a 6 nyelvcsaládra vonatkozóan.

$FIN \subset H_1(FIN, [1]) \subset H_1(FIN, [2]) \subset \ldots \subset H_1(FIN, FIN) \subset REG$

Kiterjesztett H rendszer

$\gamma = (V, T, A, R)$, ahol

V egy ábécé

$T \subseteq V$ a terminálisok halmaza

$A \subseteq V^*$ az axiómák halmaza

$R \subseteq V^* \# V^* \$ V^* \# V^*$, $\#, \$ \notin V$, a splicing szabályok halmaza

γ által generált nyelv:

$L(\gamma) = \sigma_1^*(A) \cap T^*$, ahol $\sigma = (V, R)$ H-séma.

Legyenek FL_1, FL_2 nyelvcsaládok, akkor

$EH_1(FL_1, FL_2) = \{ L(\gamma) \mid \gamma = (V, T, A, R),\ A \in FL_1,\ R \in FL_2 \}$

	FIN	REG	LIN	CF	CS	RE
FIN	REG	RE	RE	RE	RE	RE
REG	REG	RE	RE	RE	RE	RE
LIN	LIN, CF	RE	RE	RE	RE	RE
CF	CF	RE	RE	RE	RE	RE
CS	RE	RE	RE	RE	RE	RE
RE	RE	RE	RE	RE	RE	RE

$EH_1(FL_1, FL_2)$ nyelvcsaládok mérete

ahol FL_1 címkézi a sorokat, FL_2 címkézi az oszlopokat, a metszéspontokban vagy $EH_1(FL_1, FL_2)$, vagy ~~két olyan FL_3, FL_4, amelyek~~ a 6 nyelvcsaládot tekintve a legjobb alsó és felső becslések.

Egyszerű H-rendszer

$\gamma = (V, A, Q)$, ahol

V egy ábécé,

$A \subseteq V^*$ az axiómák véges halmaza

$Q \subseteq V$ a jelölő halmaz

Legyen $x, y, z \in V^*$, $a \in Q$, akkor

$(x, y) \vdash_a z \iff x = x_1 a x_2$, $y = y_1 a y_2$, $z = x_1 a y_2$, ahol $x_1, x_2, y_1, y_2 \in V^*$

Legyen

$R_Q = \{ a\#\$a\# \mid a \in Q \}$ és

$\sigma_Q = (V, R_Q)$ H-séma

γ egyszerű H-rendszer által generált nyelv:

$L(\gamma) = \sigma_Q^*(A)$.

$SH = \{ L(\gamma) \mid \gamma$ egyszerű H-rendszer $\}$

$SH \subset REG$

Univerzális véges irányított H-rendszer

2-es splicing alkalmazás:

$(x, y) \models_r (z, w) \iff x = x_1 u_1 u_2 x_2,\ y = y_1 u_3 u_4 y_2,$
$z = x_1 u_1 u_4 y_2,\ w = y_1 u_3 u_2 x_2,$
$x_1, x_2, y_1, y_2 \in V^*,$
$r = u_1 \# u_2 \$ u_3 \# u_4$ splicing szabály.

$\sigma = (V, R)$ H-rendszer, $L \subseteq V^*$ esetén

$\sigma_2(L) = \{ z \in V^* \mid (x, y) \models_r (z, w)$ vagy $(x, y) \models_r (w, z),$
$x, y \in L,\ r \in R \}$

1-es splicingre épülő fogalmak definiálhatók a 2-es splicingre is az előzőhöz hasonlóan.

Definiálhatók $\sigma_2^*(L)$, $S_2(FL_1, FL_2)$, $H_2(FL_1, FL_2)$, $EH_2(FL_1, FL_2)$.

$EH_1(FL_1, FL_2)$ nyelvcsaládok méreteire vonatkozó összefüggések érvényesek $EH_2(FL_1, FL_2)$-re is.

Kiterjesztett H-rendszer megengedett szakaszokkal:

$\gamma = (V, T, A, R)$, ahol

V egy ábécé, $T \subseteq V$,

$A \subseteq V^*$ axiómák véges halmaza,

R egy véges halmaza az $(r; C_1, C_2)$ alakú hármasoknak, ahol $r = u_1 \# u_2 \$ u_3 \# u_4$ splicing szabály V felett, $C_1, C_2 \subseteq V^*$ véges halmazok.

γ így definiálva egy véges rendszer

Legyen $x, y, z, w \in V^*$, $p \in R$, $p = (r; C_1, C_2)$.

$(x, y) \models_p (z, w) \iff (x, y) \models_r (z, w)$, x-ben C_1 minden eleme rész-sztringként megjelenik, y-ban C_2 minden eleme rész-sztringként megjelenik.

Jelölje $\sigma = (V, R)$ a γ-hoz tartozó H-sémát.

A γ által generált nyelv:

$$L(\gamma) = \sigma_2^*(A) \cap T^*$$

$EH_2([n], p[m]) = \{ L(\gamma) \mid \gamma = (V, T, A, R)$ kiterjesztett H-rendszer megengedett szakaszokkal, $|A| \leq n$, $rad(R) \leq m \}$, $n, m \geq 1$

$rad(R) = \max \{ |x| \mid x = u_i, 1 \leq i \leq 4, r = u_1 \# u_2 \$ u_3 \# u_4, (r; C_1, C_2) \in R \}$

$EH_2(FIN, pFIN) = RE$

Szigorúbb állítások is teljesülnek:

$EH_2(FIN, p[2]) = RE$

<u>Kiterjesztett véges H-rendszer tiltó szakaszokkal</u>

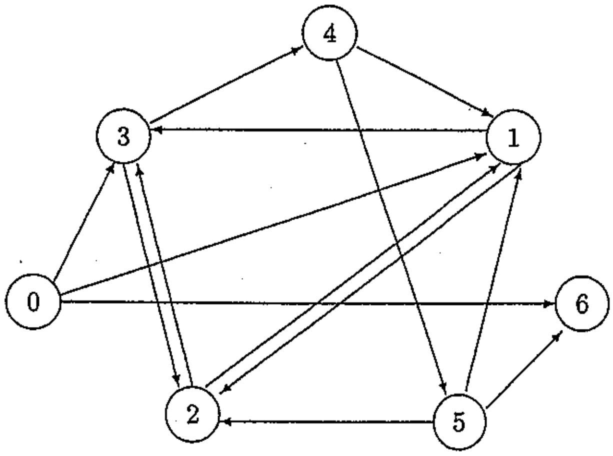
Hasonló az előzőhöz, kivéve R elemei $p = (r; D_1, D_2)$, ahol $D_1, D_2 \subseteq V^*$ véges halmazok,

$(x, y) \models_p (z, w) \iff (x, y) \models_r (z, w)$, rész-sztringként x-ben D_1, y-ban D_2 egyetlen eleme sem fordul elő.

Definiálható $L(\gamma) = \sigma_2^*(A) \cap T^*$, $EH_2([n], f[m])$, $n, m \geq 1$

$EH_2([1], f[2]) = RE$

Adleman kísérlete



1994-ben a fenti irányított gráfra a Hamilton-út problémát a gyakorlatban megoldotta DNS molekulák segítségével.

Kérdés: van-e Hamilton-út s_0-ból s_6-ba

NP-teljes probléma.

Megvalósítás:

- minden csúcshoz egy 20 nukleotid hosszú egyszerű láncot rendelt, 5'→3' irányút

  Ezek a 20 hosszú láncok olyan 10 hosszú láncokból épülnek fel, amelyek egyediek és komplementer láncaik is különböznek egymástól.

  pl.: $s_2 = \text{TATCGGATCG} \vdots \text{GTATATCCGA} = s_2' s_2''$

  $s_3 = \text{GCTATTCGAG} \vdots \text{CTTAAAGCTA} = s_3' s_3''$

  $s_4 = \text{GGCTAGGTAC} \vdots \text{CAGCATGCTT} = s_4' s_4''$

legyen $h : \{A, T, C, G\} \longrightarrow \{A, T, C, G\}$ a WK komplem szerint, legyen h értelmezve egyszerű DNS láncokra betűnként, megfordítva a lánc irányítását.

pl.: $h(s_2) = \text{ATAGCCTAGC} \vdots \text{CATATAGGCT}$

$h(s_3) = \text{CGATAAGCTC} \vdots \text{GAATTTCGAT}$

- minden $e_{i \to j}$ élhez hozzárendeli a $h(s_i'', s_j')$ egyszerű 20 nukleotid hosszúságú $3' \to 5$ irányú láncot.
  pl.:
  $e_{2\to 3}$ = CATATAGGCTICGATAAGCTC
  $e_{3\to 2}$ = GAATTTCGATIATAGCCTAGC
- a csúcsokhoz és élekhez tartozó DNS láncokból nagy mennyiséget vesz
- láncok közötti H-kötések, és a DNS ligase enzim hatására olyan kettős láncok alakulhatnak ki, amely egyik láncán csúcsokat a másikon éleket tartalmaz
  pl.

s_0	s_1	s_2	s_3	s_4	s_5	s_6
$e_{0\to 1}$	$e_{1\to 2}$	$e_{2\to 3}$	$e_{3\to 4}$	$e_{4\to 5}$	$e_{5\to 6}$	

s_0	s_6
$e_{0\to 6}$	

Kialakulnak a lehetséges utak a csúcsok között
- kiválasztja azokat a láncokat, amelyek s_0-al kezdődnek
- kiválasztja azokat a láncokat, amelyek s_6-al végződnek
- kiválasztja azokat a láncokat, amelyek hossza ≤ 140
- for $i = s_1$ to s_5 do csak azokat tartja meg, amelyekben i szerepel
- ellenőrzi, hogy maradt-e lánc.