Literatur – DNA Computing

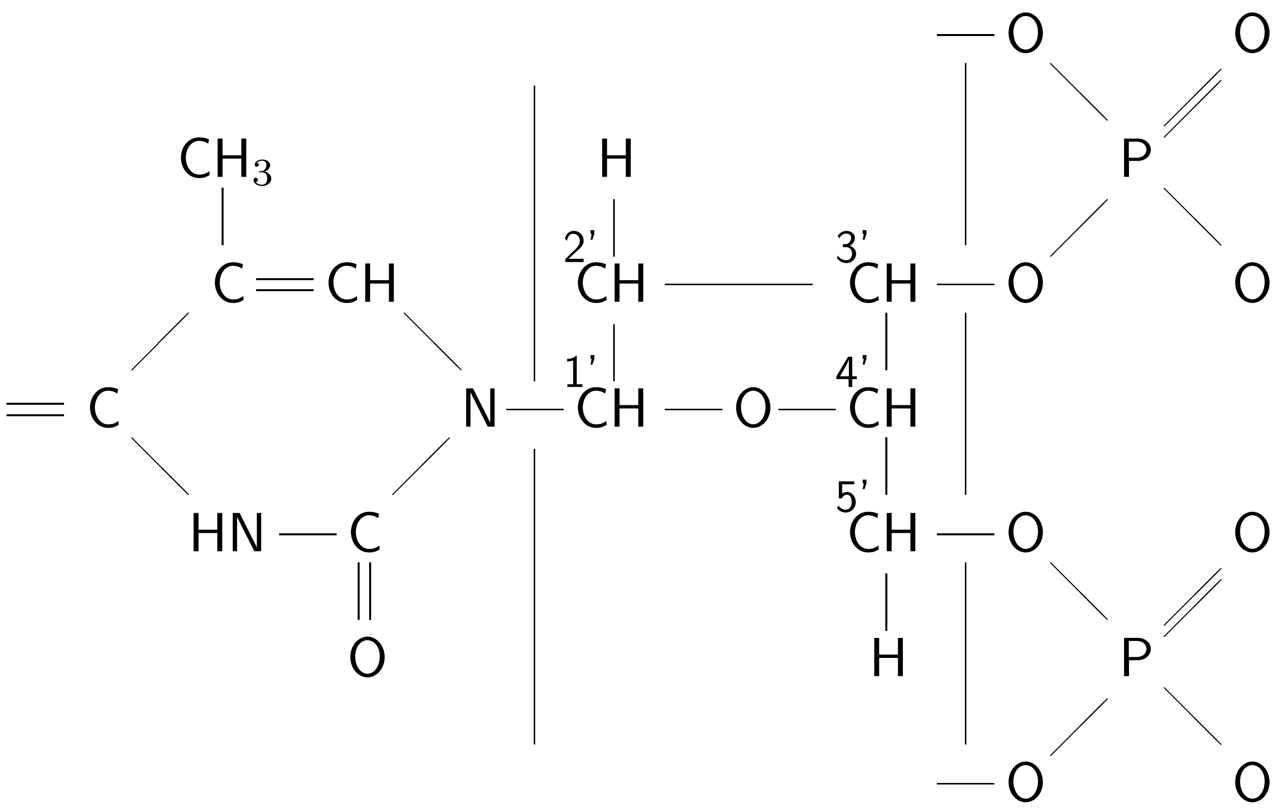
T. HEAD, Formal language theory and DNA: An analysis of the generative capacity of specific recombinant behaviors. *Bull. Math. Biology* **49** (1987) 737–759.

L. M. ADLEMAN, Molecular computation of solutions to combinatorial problems. *Science* **226** (1994) 1021–1024.

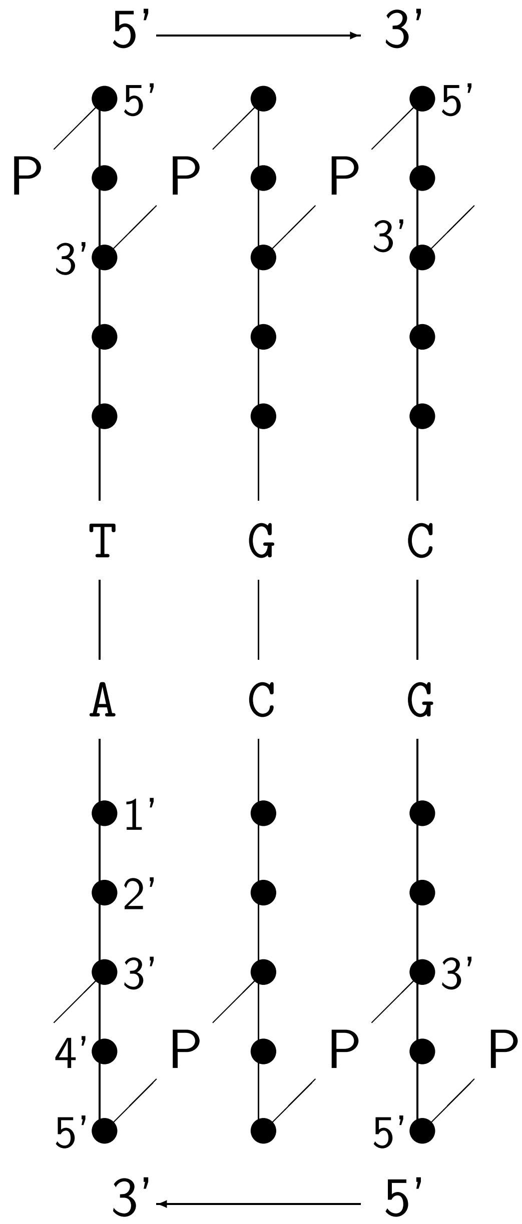
T. HEAD, GH. PĂUN and D. PIXTON, Language theory and molecular genetics. In: G. ROZENBERG and A. SALOMAA (eds.), *Handbook of Formal Languages*, Springer-Verlag, 1997, Vol. II, Chapter 7, 295–360.

GH. PĂUN, G. ROZENBERG and A.SALOMAA, *DNA Computing - New Computing Paradigms*. Springer-Verlag, Berlin, 1998.

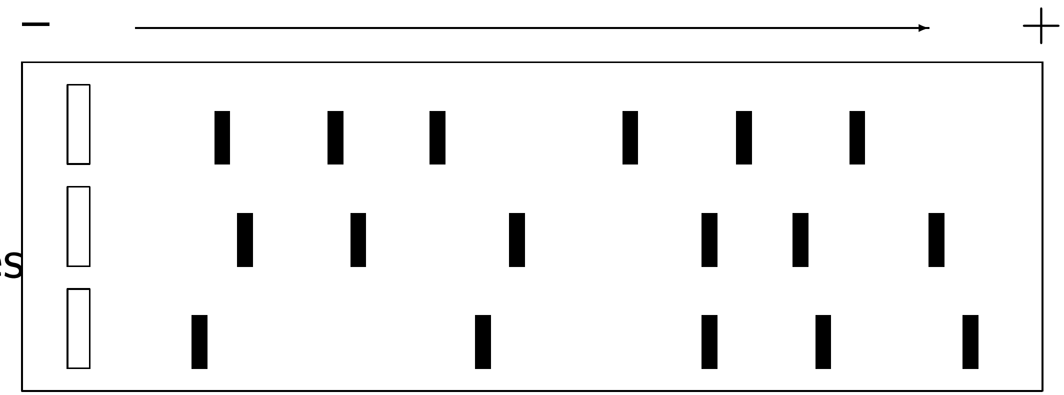
Molecule with Thymine Base



Double Stranded DNA Molecule



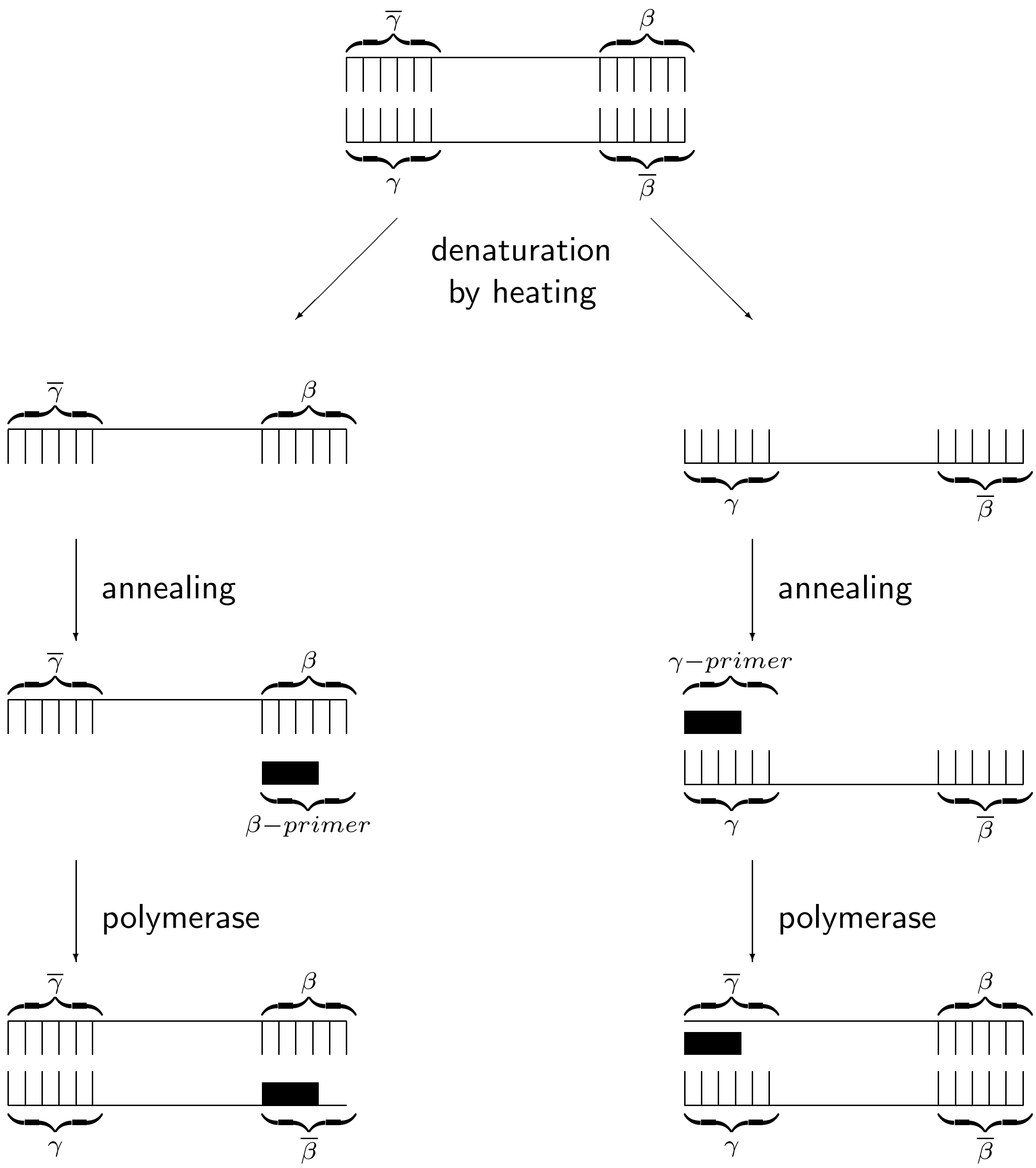
Measuring the Length of DNA Molecules by Gel Electrophoresis



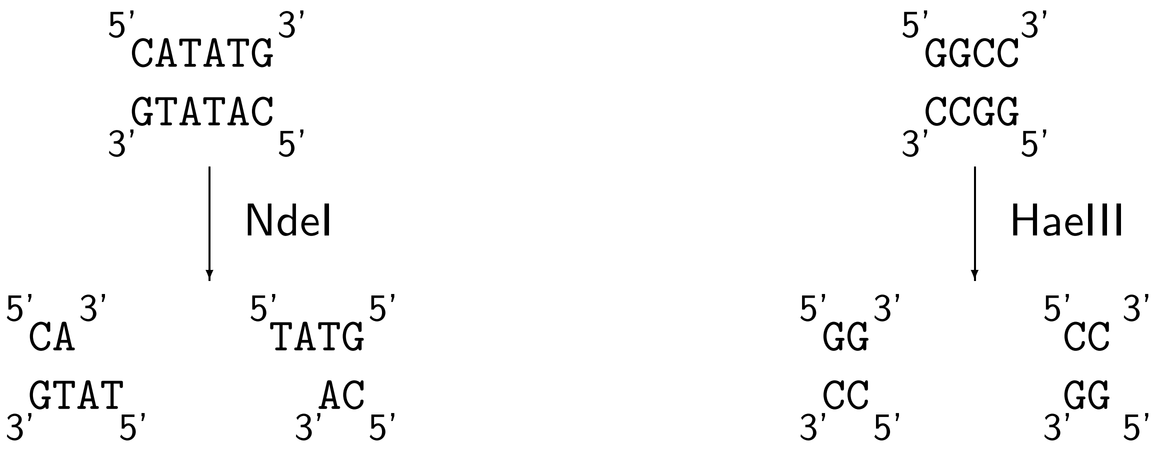
Polymerase

```
5'  3'            5'   3'            5'    3'                   5'        3'
 CGGA              CGGAG              CGGAGA                     CGGAGATGGA
 GCCTCTACCT  --->  GCCTCTACCT  --->   GCCTCTACCT  --->  ...  --->  GCCTCTACCT
3'        5'      3'        5'       3'        5'               3'        5'
```

Polymerase Chain Reaction



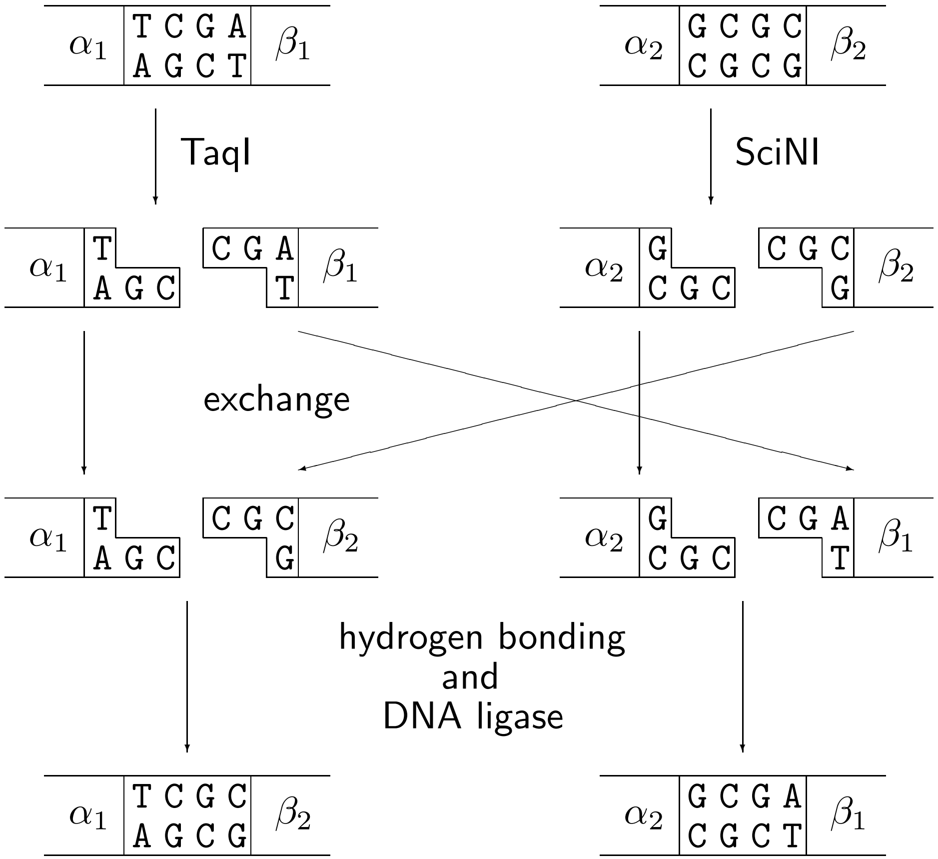
Endonuclease



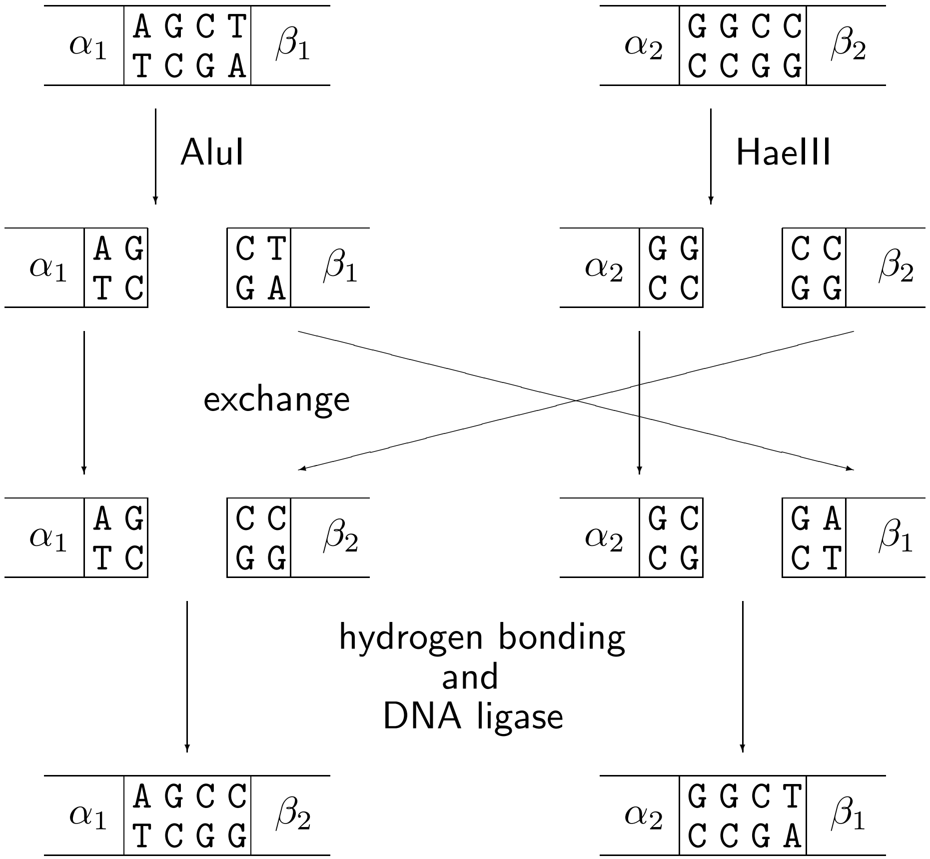
Hydrogen Bonding and DNA Ligase

```
C-A     T-A-T-G    hydrogen    C-A T-A-T-G    ligase    C-A-T-A-T-G
| |         | |   ---------->  | | | | | |  ---------->  | | | | | |
G-T-A-T     A-C    bonding     G-T-A-T A-C              G-T-A-T-A-C
```

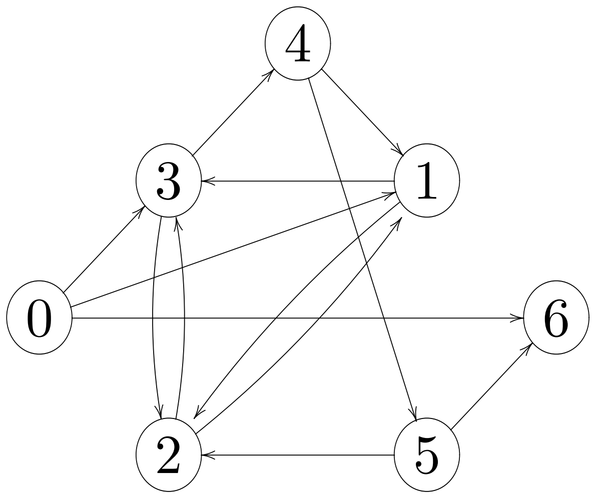
Splicing with Sticky Ends



Splicing with Blunt Ends



Adleman's Experiment



V(2) TATCGGATCGGTATATCCGA

E(2,3) CATATAGGCTCGATAAGCTC

V(3) GCTATTCGAGCTTAAAGCTA

E(3,4) GAATTTCGATCCGATCCATG

Splicing Scheme and Splicing Operation I

Definition:

A splicing scheme is a pair (V, R), where

– V is an alphabet and

– R is a subset of $V^*\#V^*\$V^*\#V^*$.

The elements of R are called splicing rules.

Definition:

We say that $w \in V^*$ and $z \in V^*$ are obtained from $u \in V^*$ and $v \in V^*$ by the splicing rule $r = r_1\#r_2\$r_3\#r_4$ and write $(u, v) \vdash_r w$ and $(u, v) \vdash_r z$, if the following conditions are satisfied:

– $u = u_1r_1r_2u_2$ and $v = v_1r_3r_4v_2$,

– $w = u_1r_1r_4v_2$ and $z = v_1r_3r_2u_2$.

Splicing Scheme and Splicing Operation II

For a language L over V and a splicing scheme (V, R) we set

$$spl(L, R) = \{w \mid (u, v) \vdash_r w,\ u \in L,\ v \in L,\ r \in R\}.$$

For two language families $\mathcal{L}_1$ and $\mathcal{L}_2$ we set

$$spl(\mathcal{L}_1, \mathcal{L}_2) = \{L \mid L = spl(L_1, L_2) \text{ for } L_1 \in \mathcal{L}_1 \text{ and a splicing scheme } (V, R) \text{ with } R \in \mathcal{L}_2\}.$$

Splicing Operation – Examples

$L = \{a^n b^n \mid n \geq 0\}$ and $R = \{a\#b\$a\#b\}$
$spl(V, R) = \{a^n b^m \mid n \geq 1, m \geq 1\}$

$L \subset V^*$ arbitrary, $L' \subset V^*$ arbitrary, $(V \cup \{c\}), R)$, $R = \{\#xc\$c\# \mid x \in L'\}$
$spl(L\{c\}, R) = \{w \mid wz \in L$ for some $z \in L'\}$

$\{a^n b^n\} \notin spl(\mathcal{L}(REG), \mathcal{L}(RE))$

Generative Power of the Splicing Operation

Theorem:
The following table holds where where at the intersection of the row marked by X and the column marked by Y we give Z if $\mathcal{L}(Z) = spl(\mathcal{L}(X), \mathcal{L}(Y))$ and Z_1/Z_2 if $\mathcal{L}(Z_1) \subset spl(\mathcal{L}(X), \mathcal{L}(Y)) \subset \mathcal{L}(Z_2)$.

	FIN	REG	CF	CS	RE
FIN	FIN	FIN	FIN	FIN	FIN
REG	REG	REG	REG/CF	REG/RE	REG/RE
CF	CF	CF	RE	RE	RE
CS	RE	RE	RE	RE	RE
RE	RE	RE	RE	RE	RE

Some Lemmas I

Lemma:
For any language families $\mathcal{L}_1, \mathcal{L}_2, \mathcal{L}'_1, \mathcal{L}'_2$ with $\mathcal{L}_1 \subseteq \mathcal{L}'_1$ and $\mathcal{L}_2 \subseteq \mathcal{L}'_2$, we have $spl(\mathcal{L}_1, \mathcal{L}_2) \subseteq spl(\mathcal{L}'_1, \mathcal{L}'_2)$.

Lemma:
If $\mathcal{L}_1$ is closed under concatenation with symbols, then $\mathcal{L}_1 \subseteq spl(\mathcal{L}_1, \mathcal{L}_2)$ for all language families $\mathcal{L}_2$.

Lemma:
If $\mathcal{L}$ is closed under concatenation, homomorphism, inverse homomorphisms and intersections with regular sets, then $spl(\mathcal{L}, \mathcal{L}(REG)) \subseteq \mathcal{L}$.

Einige Lemmata II

Lemma:
If $\mathcal{L}$ is closed under homomorphism, inverse homomorphisms and intersections with regular sets, then $spl(\mathcal{L}(REG), \mathcal{L}) \subseteq \mathcal{L}$.

Lemma:
For any recursively enumerable language L, there are context-free languages L_1 and L_2 such that $L = \{u \mid uv \in L_1 \text{ for some } v \in L_2\}$.

Lemma:
For any recursively enumerable language $L \subset V^*$, there are a context-sensitive language Sprache L' and letters c_1 and c_2, which are not in V, such that $L' \subseteq L\{c_1\}\{c_2\}^*$ holds, and for any $w \in L$ there is a number $i \geq 1$ such that $wc_1c_2^i \in L'$.

Splicing Systems

Definition:
A splicing system is a triple $G = (V, R, A)$, where
– V is an alphabet,
– R is a subset of $V^*\#V^*\$V^*\#V^*$, and
– A is a subset of V^*.

Definition:
The language $L(G)$ generated by a splicing system G is defined by the following settings:
– $spl^0(G) = A$ and $spl^{i+1}(G) = spl(spl^i(G), R)) \cup spl^i(G)$ for $i \geq 0$,
– $L(G) = \cup_{i\geq 0} spl^i(G)$.

Example:
$G = (\{a, b\}, \{a\#b\$a\#b\}, \{(a^n b^n)^m \mid n \geq 1, m \geq 1\})$
$L(G) = \{a^{r_1} b^{s_1} a^{r_2} b^{s_2} \ldots a^{r_m} b^{s_m} \mid m \geq 1,\ r_i \geq 1,\ s_i \geq 1,\ 1 \leq i \leq m\}$

Extended Splicing Systems

Definition:

i) An extended splicing system is a quadruple $G = (V, T, R, A)$ where

– $H = (V, R, A)$ is a splicing system and

– T is a subset of V.

ii) The language generated by an extended splicing system G is defined as $L(G) = L(H) \cap T^*$.

Example:

$G = (\{a, b, c\}, \{a, b\}, \{\#c\$c\#a\}, \{c^m a^n b^n \mid n \geq 1\}$

$L(G) = \{a^n b^n \mid n \geq 1\}$

Definition:

For two language families $\mathcal{L}_1$ and $\mathcal{L}_2$, we define $Spl(\mathcal{L}_1, \mathcal{L}_2)$ ($ESpl/\mathcal{L}_1, \mathcal{L}_2$) as the set of all languages $L(G)$ generated by some (extended) splicing system $G = (V, R, A)$ ($G = (V, T, R, A)$) with $A \in \mathcal{L}_1$ and $R \in \mathcal{L}_2$.

The Power of Splicing Systems

Theorem:
The following table holds, where at the intersection of the row marked by X and the coloumn marked by Y we give Z if $\mathcal{L}(Z) = Spl(\mathcal{L}(X), \mathcal{L}(Y))$ and Z_1/Z_2 if $\mathcal{L}(Z_1) \subset Spl(\mathcal{L}(X), \mathcal{L}(Y)) \subset \mathcal{L}(Z_2)$.

	FIN	REG	CF	CS	RE
FIN	FIN/REG	FIN/RE	FIN/RE	FIN/RE	FIN/RE
REG	REG	REG/RE	REG/RE	REG/RE	REG/RE
CF	CF	CF/RE	CF/RE	CF/RE	CF/RE
CS	CS/RE	CS/RE	CS/RE	CS/RE	CS/RE
RE	RE	RE	RE	RE	RE

The Power of Extended Splicing Systems

Theorem:
The following table holds, where at the intersection of the row marked by X and the coloumn marked by Y we give Z if $\mathcal{L}(Z) = ESpl(\mathcal{L}(X), \mathcal{L}(Y))$.

	FIN	REG	CF	CS	RE
FIN	REG	RE	RE	RE	RE
REG	REG	RE	RE	RE	RE
CF	CF	RE	RE	RE	RE
CS	RE	RE	RE	RE	RE
RE	RE	RE	RE	RE	RE

Some Lemmas III

Lemma:
For any language families $\mathcal{L}_1, \mathcal{L}_2, \mathcal{L}'_1, \mathcal{L}'_2$ with $\mathcal{L}_1 \subseteq \mathcal{L}'_1$ and $\mathcal{L}_2 \subseteq \mathcal{L}'_2$, we have $ESpl(\mathcal{L}_1, \mathcal{L}_2) \subseteq ESpl(\mathcal{L}'_1, \mathcal{L}'_2)$.

Lemma:
If a language family $\mathcal{L}$ is closed under concatenation with symbols, then $\mathcal{L} \subseteq ESpl(\mathcal{L}, \mathcal{L}(FIN))$.

Lemma:
$\mathcal{L}(REG) \subseteq ESpl(\mathcal{L}(FIN), \mathcal{L}(FIN))$.

Some Lemmas IV

Lemma:
For any family $\mathcal{L}$ which is closed under union, concatenation, Kleene-closure, homomorphisms, inverse homomorphisms and intersections with regular sets, $ESpl(\mathcal{L}, \mathcal{L}(FIN)) \subseteq \mathcal{L}$.

Lemma:
For any recursively enumerable language $L \subseteq T^*$, there is an extended splicing system $G = (V, T, R, A)$ with a finite set A and a regular set R of splicing rules such that $L(G) = L$.

Lemma:
For any extended splicing system $G = (V, T, R, A)$, $L(G)$ is a recursively enumerable set.

Some Measures of Descriptional Complexity – Definitions

Definition: i) For a splicing system $G = (V, R, A)$ or an extended splicing system $G = (V, T, R, A)$ we define the complexity measures $r(G)$, $a(G)$ and $l(G)$ by

$r(G) = \max\{|u| \mid u = u_i$ for some $u_1\#u_2\$u_3\#u_4 \in R,\ 1 \leq i \leq 4\}$,

$a(G) = \#(A)$,

$l(G) = \max\{|z| \mid z \in A\}$.

ii) For a language family $\mathcal{L}$ and $n \geq 1$ and $m \in \{a, l\}$, we define the families $\mathcal{L}_n(r, \mathcal{L})$ and $\mathcal{L}_n(m, \mathcal{L})$ as the set of languages $L(G)$ where $G = (V, R, A)$ is a splicing system with $r(G) \leq n$ and $A \in \mathcal{L}$ and with $m(G) \leq n$ and $R \in \mathcal{L}$, respectively.

Analogously, for $m \in \{r, a, l\}$ and extended splicing systems, we define the sets $\mathcal{L}_n(em, \mathcal{L})$.

Results on Descriptional Complexities – Results

Theorem: For any $n \geq 1$,
i) $\mathcal{L}(FIN) \subset \mathcal{L}_n(r, \mathcal{L}(FIN)) \subset Spl(\mathcal{L}(FIN), \mathcal{L}(FIN))$,
ii) $\mathcal{L}_n(r, \mathcal{L}(FIN)) \subset \mathcal{L}_{n+1}(r, \mathcal{L}(FIN))$:

Theorem: For $\mathcal{L} \in \{\mathcal{L}(REG), \mathcal{L}(CF), \mathcal{L}(RE)\}$ and $n \geq 1$,
$\mathcal{L}_n(r, \mathcal{L}) = \mathcal{L}$.

Theorem: For any $n \geq 1$,
$\mathcal{L}_n(ea, \mathcal{L}(REG)) = ESpl(\mathcal{L}(FIN), \mathcal{L}(REG))$.

Theorem: For any $n \geq 2$,
$\mathcal{L}_1(el, \mathcal{L}(REG)) \subset \mathcal{L}_n(el, \mathcal{L}(REG)) = ESpl(\mathcal{L}(FIN), \mathcal{L}(REG))$: