From Pattern Matching to Suffix Trees*

W. Szpankowski
Department of Computer Science
Purdue University
W. Lafayette, IN 47907
U.S.A.

May 26, 2004

*Joint Work with P. Jacquet and M. Regnier
Outline of the Talk

1. Pattern Matching Problems
   - String Matching
   - Subsequence Matching
   - Self-Repetitive Pattern Matching
2. Suffix Trees and Its Parameters
3. Applications
4. Analysis of the Exact Pattern Matching
   - Languages and Generating Functions
   - Asymptotic Analysis
5. Analysis of Suffix Trees
   - Relation to the Exact Pattern Matching
   - Generating Function for the Depth
   - Analysis of Tries
   - Comparison of Tries and Suffix Trees
Pattern Matching

Let \( \mathcal{W} \) and \( T \) be (set of) strings generated over a finite alphabet \( \mathcal{A} \).

We call \( \mathcal{W} \) the pattern and \( T \) the text. The text \( T \) is of length \( n \) and is generated by a probabilistic source.

We shall write

\[
T^n_m = T_m \ldots T_n.
\]

The pattern \( \mathcal{W} \) can be a single string

\[
\mathcal{W} = w_1 \ldots w_m, \quad w_i \in \mathcal{A}
\]

or a set of strings

\[
\mathcal{W} = \{ \mathcal{W}_1, \ldots, \mathcal{W}_d \}
\]

with \( \mathcal{W}_i \in \mathcal{A}^{m_i} \) being a set of strings of length \( m_i \).
**Basic Parameters**

Two basic questions are:

- **how many times** $\mathcal{W}$ **occurs in** $T$,
- **how long one has to wait until** $\mathcal{W}$ **occurs in** $T$.

The following quantities are of interest:

- $O_n(\mathcal{W})$ — the number of times $\mathcal{W}$ occurs in $T$:
  \[
  O_n(\mathcal{W}) = \#\{i : T_{i-m+1}^i = \mathcal{W}, \ m \leq i \leq n\}.
  \]

- $W_{\mathcal{W}}$ — the first time $\mathcal{W}$ occurs in $T$:
  \[
  W_{\mathcal{W}} := \min\{n : T_{n-m+1}^n = \mathcal{W}\}.
  \]

**Relationship:**

\[ W_{\mathcal{W}} > n \iff O_n(\mathcal{W}) = 0. \]
Various Pattern Matching

(Exact) String Matching

In the exact string matching the pattern \( \mathcal{W} = w_1 \ldots w_m \) is a given string (i.e., consecutive sequence of symbols).

Generalized String Matching

In the generalized pattern matching a set of patterns (rather than a single pattern) is given, that is,

\[
\mathcal{W} = (\mathcal{W}_0, \mathcal{W}_1, \ldots, \mathcal{W}_d), \quad \mathcal{W}_i \in \mathcal{A}^{m_i}
\]

where \( \mathcal{W}_i \) itself for \( i \geq 1 \) is a subset of \( \mathcal{A}^{m_i} \) (i.e., a set of words of a given length \( m_i \)).

The set \( \mathcal{W}_0 \) is called the forbidden set.

Three cases to be considered:

\( \mathcal{W}_0 = \emptyset \) — one is interested in the number of patterns from \( \mathcal{W} \) occurring in the text.

\( \mathcal{W}_0 \neq \emptyset \) — we study the number of \( \mathcal{W}_i, i \geq 1 \) pattern occurrences under the condition that no pattern from \( \mathcal{W}_0 \) occurs in the text.

\( \mathcal{W}_i = \emptyset, i \geq 1, \mathcal{W}_0 \neq \emptyset \) — restricted pattern matching.
Hidden Words or Subsequence Pattern Matching

In this case we search in text for a subsequence $W = w_1 \ldots w_m$ rather than a string, that is, we look for indices $1 \leq i_1 < i_2 < \cdots < i_m \leq n$ such that

$$T_{i_1} = w_1, \ T_{i_2} = w_2, \ldots, \ T_{i_m} = w_m.$$  

We also say that the word $W$ is “hidden” in the text.

For example:

$$W = \text{date}$$

$$T = \text{hidden pattern}$$

occurs four times as a subsequence in the text as hidden pattern but not even once as a string.

Self-Repetitive Pattern Matching

In this case the pattern $W$ is part of the text:

$$W = T_1^m.$$  

We may ask when the first $m$ symbols of the text will occur again. This is important in Lempel-Ziv like compression algorithms and suffix trees.
Suffix Trees and Its Parameters

\[ S_1 = 1010010001 \]
\[ S_2 = 010010001 \]
\[ S_3 = 10010001 \]
\[ S_4 = 0010001 \]
\[ S_5 = 010001 \]

Figure 1: Suffix tree built from the first five suffixes \( S_1, \ldots, S_5 \) of \( T = 1010010001 \ldots \)

- **Depth** \( D_n \) - length of the path from the root to a randomly selected external node (suffix).
- **Height** \( H_n \) - length of the longest path.
- **Path length** \( L_n \) - sum of all paths to nodes.
The depth $D_n$ in a suffix tree can be defined in terms of the number of occurrence $O_n$ of a pattern $W = w$ as follows:

- Define $D_n(i), 1 \leq i \leq n$, to be the largest value of $k \leq n$ such that $T_{i}^{i+k-1}$ occurs at least twice in the text $T_{1}^{n}$ of length $n$; that is, $O_n(T_{i}^{i+k-1}) \geq 2$.
- The depth $D_n$ is equal to $D_n(i)$ when $i$ is randomly and uniformly selected between 1 and $n$.
- Thus for $w \in A^k$ we have

$$\Pr(D_n(i) \geq k \& T_{i}^{i+k-1} = w) = \Pr(O_n(w) \geq 2 \& T_{i}^{i+k-1} = w),$$

and

$$\sum_{i=1}^{n} \Pr(O_n(w) = r \& T_{i}^{i+k-1} = w) = r \Pr(O_n(w) = r).$$
Throughout the talk I will assume that the text is generated by a random source.

**Memoryless Source**
The text is a realization of an independently, identically distributed sequence of random variables (i.i.d.), such that a symbol \( s \in A \) occurs with probability \( P(s) \).

**Markovian Source**
The text is a realization of a stationary Markov sequence of order \( K \), that is, probability of the next symbol occurrence depends on \( K \) previous symbols.
Here is an incomplete list of results on **string pattern matching** (given a pattern $W$ find statistics of its occurrences):

- **Feller** (1968),
- **Guibas and Odlyzko** (1978, 1981),
Languages and Generating Functions

A language $\mathcal{L}$ is a collection of words satisfying some properties.

For any language $\mathcal{L}$ we define its generating function $L(z)$ as

$$L(z) = \sum_{u \in \mathcal{L}} P(u) z^{|u|}$$

where $P(w)$ is the stationary probability $u$ occurrence, $|u|$ is the length of $w$.

For Markov sources we define $\mathcal{W}$-conditional generating function:

$$L_{\mathcal{W}}(z) = \sum_{u \in \mathcal{L}} P(u | u_{-m} = w_1 \cdots u_{-1} = w_m) z^{|u|}$$

where $u_{-i}$ stands for a symbol preceding the first character of $u$ at distance $i$. 
Given a pattern $\mathcal{W}$, we define the \textbf{autocorrelation set} $S$ as:

$$S = \{ w_{k+1}^m : w_1^k = w_{m-k+1}^m, w_1^k = w_{m-k+1}^m \}$$

and $\mathcal{WW}$ is the set of positions $k$ satisfying $w_1^k = w_{m-k+1}^m$.

The \textbf{generating function} of $S$ is denoted as $S(z)$ and we call it the \textbf{autocorrelation polynomial}.

$$S(z) = \sum_{k \in \mathcal{WW}} P(w_{k+1}^m) z^{m-k}.$$ 

Its $\mathcal{W}$-\textbf{conditional generating function} is denoted $S_\mathcal{W}(z)$. For example, for a \textbf{Markov model} we have

$$S_\mathcal{W}(z) = \sum_{k \in \mathcal{WW}} P(w_{k+1}^m | w_k^k) z^{m-k}.$$
Example:

Let $W = bab$ over alphabet $A = \{a, b\}$.

\[ WWH = \{1, 3\} \text{ and } S = \{\epsilon, ab\}, \]

where $\epsilon$ is the empty word, since

\[
\begin{array}{ccc}
  b & a & b \\
  b & a & b
\end{array}
\]

For the unbiased memoryless source

\[ S(z) = 1 + P(ab)z^2 = 1 + \frac{z^2}{4}. \]

For the Markovian model of order one

\[ S_{bab}(z) = 1 + P(ab|b)z^2 = 1 + p_{ba}p_{ab}z^2. \]
We are interested in the following language:

\( \mathcal{I}_r \) – set of words that contains exactly \( r \geq 1 \) occurrences of \( \mathcal{W} \),

and its generating functions

\[
O_r(z) = \sum_{n \geq 0} \Pr\{ O_n(\mathcal{W}) = r \} z^n, \quad r \geq 1,
\]

\[
O(z, u) = \sum_{r=1}^{\infty} T_r(z) u^r = \sum_{r=1}^{\infty} \sum_{n=0}^{\infty} \Pr\{ O_n(\mathcal{W}) = r \} z^n u^r
\]

for \( |z| \leq 1 \) and \( |u| \leq 1 \).
(i) Let $T$ be a language of words containing at least one occurrence of $W$.

(ii) We define $R$ as the set of words containing only one occurrence of $W$, located at the right end. For example, for $W = aba$

$$ccaba \in R.$$ 

(iii) We also define $U$ as

$$U = \{ u : W \cdot u \in T_1 \}$$

that is, a word $u \in U$ if $W \cdot u$ has exactly one occurrence of $W$ at the left end of $W \cdot u$,

$$cba \in U, \quad ba \notin U.$$ 

(iv) Let $M$ be the language:

$$M = \{ u : W \cdot u \in T_2 \text{ and } W \text{ occurs at the right of } W \cdot u \},$$

that is, $M$ is a language such that $WM$ has exactly two occurrences of $W$ at the left and right end of a word from $M$.

$$ba \in M.$$
Lemma 1. The language $T$ satisfies the fundamental equation:

$$ T = R \cdot M^* \cdot U. $$

Notably, the language $T_r$ can be represented for any $r \geq 1$ as follows:

$$ T_r = R \cdot M^{r-1} \cdot U, $$

and

$$ T_0 \cdot \mathcal{W} = R \cdot S. $$

Here, by definition $M^0 := \{\epsilon\}$ and $M^* := \bigcup_{r=0}^{\infty} M^r$.

Example: Let $\mathcal{W} = TAT$. The following string belongs $T_3$:

$$ \overbrace{R}^{\mathcal{R}} \overbrace{CCTAT}^{\mathcal{M}} \overbrace{AT}^{\mathcal{M}} \overbrace{GATAT}^{\mathcal{M}} \overbrace{GGA}^{\mathcal{U}}. $$
Theorem 1. (i) The languages $\mathcal{M}, \mathcal{U}$ and $\mathcal{R}$ satisfy:

$$\bigcup_{k \geq 1} \mathcal{M}^k = A^* \cdot \mathcal{W} + S - \{\epsilon\},$$

$$\mathcal{U} \cdot A = \mathcal{M} + \mathcal{U} - \{\epsilon\},$$

$$\mathcal{W} \cdot \mathcal{M} = A \cdot \mathcal{R} - (\mathcal{R} - \mathcal{W}),$$

where $A^*$ is the set of all words, $+$ and $-$ are disjoint union and subtraction of languages.

(ii) The generating functions associated with languages $\mathcal{M}, \mathcal{U}$ and $\mathcal{R}$ satisfy for memoryless sources

$$\frac{1}{1 - M(z)} = S_W(z) + P(\mathcal{W}) \frac{z^m}{1 - z},$$

$$U_W(z) = \frac{M(z) - 1}{z - 1},$$

$$R(z) = P(\mathcal{W}) z^m \cdot U_W(z)$$

(Extension to Markov sources possible; cf. Regnier & WS.)
Theorem 2. The generating functions $T_r(z)$ and $T(z, u)$ are

\[
O_r(z) = R(z) M_W^{r-1}(z) U_W(z), \quad r \geq 1
\]

\[
O(z, u) = R(z) \frac{u}{1 - uM(z)} U_W(z)
\]

\[
O_0(z) P(W) = R(z) S_W(z)
\]

where

\[
M(z) = 1 + \frac{z - 1}{D_W(z)},
\]

\[
U_W(z) = \frac{1}{D_W(z)},
\]

\[
R(z) = z^m P(W) \frac{1}{D_W(z)}.
\]

with

\[
D_W(z) = (1 - z) S_W(z) + z^m P(W).
\]
Main Results: Asymptotics

Theorem 3. (i) **Moments.** The expectation satisfies, for \( n \geq m \):  
\[
\mathbb{E}[O_n(W)] = P(W)(n - m + 1),
\]

while the variance is  
\[
\text{Var}[O_n(W)] = nc_1 + c_2.
\]

with

\[
c_1 = P(W)(2S(1) - 1 - (2m - 1)P(W),
\]

\[
c_2 = P(W)((m - 1)(3m - 1)P(W)
\]
\[
- (m - 1)(2S(1) - 1) - 2S'(1)).
\]
(ii) **Case** $r = O(1)$. Let $\rho_{W}$ be the smallest root of

$$D_{W}(z) = (1 - z)S_{W}(z) + z^{m}P(W) = 0.$$ 

Then

$$\Pr\{O_{n}(W) = r\} \sim \sum_{j=1}^{r+1} (-1)^{j} a_{j} \binom{n}{j-1} \rho_{W}^{-(n+j)}$$

where

$$a_{r+1} = \frac{\rho_{W}^{m}P(W)(\rho_{W} - 1)^{r-1}}{(D'_{W}(\rho_{W}))^{r+1}},$$

and the remaining coefficients can be easily computed, too.
Central Limit and Large Deviations

(iii) **CLT** Case \( r = E O_n + x \sqrt{\text{Var} O_n} \) for \( x = O(1) \). Then:

\[
\Pr\{O_n(W) = r\} = \frac{1}{\sqrt{2\pi c_1 n}} e^{-\frac{1}{2} x^2} \left( 1 + O\left( \frac{1}{\sqrt{n}} \right) \right).
\]

(iv) **Large Deviations**: Case \( r = (1 + \delta) E O_n \). Let \( a = (1 + \delta) P(W) \) with \( \delta \neq 0 \). For complex \( t \), define \( \rho(t) \) to be the root of

\[
1 - e^t M_W(e^\rho) = 0,
\]

while \( \omega_a \) and \( \sigma_a \) are defined as

\[
-\rho'(\omega_a) = a \quad \quad -\rho''(\omega_a) = \sigma_a^2
\]

Then

\[
\Pr\{O_n(W) \sim (1 + \delta) E O_n\} = \frac{e^{-(n-m+1)I(a)+\delta_a}}{\sigma_a \sqrt{2\pi(n-m+1)}}
\]

where \( I(a) = a \omega_a + \rho(\omega_a) \) and \( \delta_a \) is a constant.
Recall that:

- For any \( i \leq n \), \( D_n(i) \) is the largest \( k \) such that \( T_{i+k-1}^i \) occurs at least twice in the text \( T_1^n \).
- The typical depth \( D_n \) is defined as
  \[
  \Pr(D_n = \ell) = \frac{1}{n} \sum_{i=1}^{n} \Pr(D_n(i) = \ell)
  \]
  for any \( 1 \leq \ell \leq n \).
- For \( w \in A^k \) let \( O_n(w) \) be, as before, the number of times \( w \) occurs in the text \( T_1^n \).
- The following is true
  \[
  \Pr(D_n(i) \geq k \& T_{i+k-1}^i = w) = \Pr(O_n(w) \geq 2 \& T_{i+k-1}^i = w),
  \]
  and
  \[
  \sum_{i=1}^{n} \Pr(O_n(w) = r \& T_{i+k-1}^i = w) = r \Pr(O_n(w) = r).
  \]
Basic Relationship Between $D_n$ and $O_n$

Recalling that $O_n(u) = \mathbf{E}[u^{O_n(w)}]$, we have

$$
\Pr(D_n \geq k) = \frac{1}{n} \sum_{i=1}^{n} \Pr(D_n(i) \geq k)
$$

$$
= \sum_{w \in A^k} \frac{1}{n} \sum_{i=1}^{n} \Pr(D_n(i) \geq k \land T_i^{i+k-1} = w)
$$

$$
= \frac{1}{n} \sum_{w \in A^k} \sum_{r \geq 2} r \Pr(O_n(w) = r)
$$

$$
= \sum_{w \in A^k} \left( \Pr(w) - \frac{1}{n} O_{n,w}'(0) \right)
$$

$$
= 1 - \frac{1}{n} \sum_{w \in A^k} \frac{d}{du} \left( \mathbf{E}[u^{O_n(w)}] \right) \bigg|_{u=0}
$$

where $O_{n,w}'(0)$ denotes the derivative of $O_n(u)$ at $u = 0$.
Generating Function of $D_n$

From previous slide we conclude that the probability generating function

$$D_n(u) = \mathbb{E}[u^{D_n}] = \sum \Pr(D_n = k) u^k$$

becomes

$$D_n(u) = \frac{1}{n} \frac{(1 - u)}{u} \sum_{w \in A^*} u^{|w|} O_{n,w}'(0),$$

and the bivariate generating function

$$D(z, u) = \sum_n nD_n(u) z^n$$

is

$$D(z, u) = \frac{1 - u}{u} \sum_{w \in A^*} u^{|w|} \frac{\partial}{\partial u} O_w(z, 0)$$

where $O_w(z, u) = \sum_{n=0}^\infty \sum_{r=0}^\infty \Pr(O_n(w) = r) z^n u^r$. 
In Theorem 1 of pattern matching we derived

$$O_w(z, u) = \frac{z^{\mid w \mid} \Pr(w)}{D_w^2(z)} \frac{u}{1 - u M_w(z)} + \frac{S_w(z)}{D_w(z)},$$

where

$$M_w(z) - 1 = \frac{z - 1}{D_w(z)}$$

$$D_w(z) = (1 - z) S_w(z) + z^{\mid w \mid} \Pr(w)$$

and $S_w(z)$ is the autocorrelation polynomial for $w$.

**Lemma 2.** The bivariate generating function for $D_n$ is

$$D(z, u) = \frac{1 - u}{u} \sum_{w \in A^*} (zu)^{\mid w \mid} \frac{\Pr(w)}{((1 - z) S_w(z) + z^{\mid w \mid} \Pr(w))^2}$$

for $|u| < 1$ and $|z| < 1$, where $S_w(z)$ is the autocorrelation polynomial for $w$. 
Define $h$ to be the entropy and $h_2 = \sum_{i=1}^{V} p_i \log^2 p_i$.

**Theorem 4.** (i) For a biased memoryless source (i.e., $p_i \neq p_j$ for some $i \neq j$) and any $\varepsilon > 0$

$$
\mathbb{E}D_n = \frac{1}{h} \log n + \frac{\gamma}{h} + \frac{h_2}{h^2} + P_1(\log n) + O(n^{-\varepsilon}),
$$

$$
\text{Var}(D_n) = \frac{h_2 - h^2}{h^3} \log n + O(1)
$$

where $P_1(\cdot)$ is a periodic function with small amplitude when the tuple ($\log p_1, \ldots, 
\log p_V$), is collinear with a rational tuple (i.e., $\log p_j / \log p_1 = r/s$ for some integers $r$ and $s$) and converges to zero otherwise.

Furthermore, $(D_n - \mathbb{E}[D_n]) / \text{Var}(D_n)$ is asymptotically normal with mean zero and variance one that is, for fixed $x \in \mathbb{R}$

$$
\lim_{n \to \infty} \Pr\{D_n \leq \mathbb{E}[D_n] + x \sqrt{\text{Var}(D_n)}\} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2} dt ,
$$

and for all integer $m$

$$
\lim_{n \to \infty} \mathbb{E}\left[ \frac{D_n - \mathbb{E}[D_n]}{\sqrt{\text{Var}D_n}} \right]^m = \begin{cases} 
0 & \text{when } m \text{ is odd} \\
\frac{m!}{2^{m/2}(m/2)!} & \text{when } m \text{ is even}
\end{cases}
$$
(ii) For the *unbiased source* (i.e., $p_1 = \cdots = p_V = 1/V$), $h_2 = h^2$, the expected value $E[D_n]$ is given above, and for any $\epsilon > 0$

$$\text{Var}(D_n) = \frac{\pi^2}{6 \log^2 V} + \frac{1}{12} + P_2(\log n) + O(n^{-\epsilon})$$

where $P_2(\log n)$ is a periodic function with small amplitude. The limiting distribution of $D_n$ does not exist, but one finds

$$\lim_{n \to \infty} \sup_x | \Pr(D_n \leq x) - \exp(-nV^{-x}) | = 0$$

for any fixed real $x$. 
1. Consider a trie built from \( n \) independently generated texts. Let \( D^T_n \) be the typical depth in such a trie. Since

\[
\Pr(D^T_n(i) < k) = \sum_{w \in A^k} \Pr(w)(1 - \Pr(w))^{n-1}.
\]

we obtain the following formulas for the probability generating function \( D^T_n(u) = \mathbb{E}[u^{D^T_n}] \) and the bivariate generating function \( D(z, u) \)

\[
D^T_n(u) = \frac{1 - u}{u} \sum_{w \in A^*} u^{|w|} \Pr(w)(1 - \Pr(w))^{n-1},
\]

\[
D^T(z, u) = \frac{1 - u}{u} \sum_{w \in A^*} u^{|w|} \frac{z \Pr(w)}{(1 - z + \Pr(w)z)^2}
\]

for all \(|u| \leq 1\) and \(|z| < 1\).
**Result on Tries**

2. This is known (for independent tries).

**Lemma 3 (P. Jacquet, M. Regnier, W.S.).** There exists $\varepsilon > 0$ such that

$$D_n^T(u) = (1 - u)n^{\kappa(u)}(\Gamma(\kappa(u)) + P(\log n, u))) + O(n^\varepsilon),$$

where $\Gamma$ is the Euler gamma function

$$u \sum_{i=1}^{V} p_i^{1-\kappa(u)} = 1$$

and $P(\log n, u)$ is periodic function with small amplitude when the vector $(\log p_1, \ldots, \log p_V)$ is collinear with a rational tuple, and converges to zero when $n \to \infty$ otherwise.

This lemma implies

$$e^{-tc_1 \log n / \sqrt{c_2 \log n}} D_n^T \left( e^{t / \sqrt{c_2 \log n}} \right) \to e^{t^2/2}$$

where $c_1 = 1/h$ and $c_2 = (h_2 - h^2)/h^3$.

That is,

$$\frac{D_n^T - c_1 \log n}{\text{Var}[c_2 \log n]} \to N(0, 1).$$
Our Goal is ...

Our goal now is to prove that $D_n(u)$ and $D_n^T(u)$ are asymptotically close as $n \to \infty$.

3. We accomplish this by proving that for some $\epsilon > 0$ and all $|u| < \beta$ for $\beta > 1$

$$D_n^T(u) - D_n(u) = (1 - u)O(n^{-\epsilon}),$$

that is,

$$|\Pr(D_n \leq k) - \Pr(D_n^T \leq k)| = O(n^{-\epsilon} \beta^{-k}).$$

for all positive integer $k$. 
4. For almost all words \( w \) the autocorrelation polynomial

\[
S_w(z) \approx 1.
\]

**Lemma 4.** There exist \( \delta < 1, \theta > 0 \) and \( \rho > 1 \) such that \( \rho \delta < 1 \) and

\[
\sum_{w \in A^k} \left[ |S_w(\rho) - 1| \leq (\rho \delta)^k \theta \right] \Pr(w) \geq 1 - \theta \delta^k.
\]

Based on the following observation: in order for \( w \in A \) to have the same prefix of length \( k - i \) (for some \( i \)) as the suffix of length \( k - i \), that is,

\[
w_{1}^{k-i} = w_{k-i+1}^{k}
\]

pattern \( w \) must have a periodic structure, that is, for some \( u \in A^i \) we have

\[
w_{1}^{k} = u \cdot u \cdot \ldots \cdot u \bar{u}
\]

where \( \bar{u} \leq i \).
5. The bivariate generating function $D(z, u)$ can be analytically continued to a larger disk.

**Lemma 5.** The generating function $D(z, u)$ can be analytically continued for all $|u| < \delta^{-1}$ and $|z| < 1$ where $\delta < 1$.

We prove this lemma by showing

$$uD(z, u) - \frac{(1 - u)}{(1 - uz)(1 - z)^2} = O \left( \frac{u - 1}{1 - \delta|u|} \right)$$

for $\delta < 1$ and $|z| < 1$
6. Define

\[
Q_n(u) = \frac{u}{1 - u} \left( D_n(u) - D^T_n(u) \right),
\]

and

\[
Q(z, u) = \sum_{n=0}^{\infty} n Q_n(u) z^n = \frac{u}{1 - u} \left( D(z, u) - D^T(z, u) \right).
\]

That is,

\[
Q(z, u) = \sum_w u^{\|w\|} \Pr(w) \left( \frac{z^{\|w\|}}{D_w(z)^2} - \frac{z}{(1 - z + \Pr(w)z)^2} \right).
\]

**Lemma 6.** For all \( 1 < \beta < \delta^{-1} \), there exists \( \varepsilon > 0 \) such that

\[
Q_n(u) = (1 - u)O(n^{-\varepsilon})
\]

uniformly for \( |u| \leq \beta \).