PERFORMANCE OPTIMIZATION OF TURKISH VIRTUAL KEYBOARDS

A THESIS
SUBMITTED TO THE DEPARTMENT OF
COMPUTER ENGINEERING AND
THE INSTITUTE OF ENGINEERING AND SCIENCE OF
BİLKENT UNIVERSITY
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
MASTER OF SCIENCE

By
Sinan Uşşaklı
May, 2004

I certify that I have read this thesis and that in my opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Science.

Prof. Dr. Bülent Özgüç (Supervisor)

I certify that I have read this thesis and that in my opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Science.

Prof. Dr. Cevdet Aykanat

I certify that I have read this thesis and that in my opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Science.

Assist. Prof. Dr. Kürşat Çağıltay

Approved for the Institute of Engineering and Science:

Prof. Dr. Mehmet B. Baray Director of the Institute

ABSTRACT

PERFORMANCE OPTIMIZATION OF TURKISH VIRTUAL KEYBOARDS

Sinan Uşşaklı
M. S. in Computer Engineering
Supervisor: Prof. Dr. Bülent Özgüç
May, 2004

Text input is one of the major problems of handheld and tablet computers. In this thesis we analyze Turkish language to create different digraph tables from Middle East Technical University Turkish Corpus and *ek$i sözlük*. We describe how to model human hand movement with Fitts' Law, and using it, we describe two different quantitative design methods: Metropolis Algorithm and Kernighan and Lin's heuristic to create the best virtual keyboard layout for Turkish. We discuss why Kernighan and Lin's heuristic is preferable for our problem.

Keywords: Virtual keyboard design, Metropolis Algorithm, KL heuristic, Turkish keyboard, handheld computers, digraphs

ÖZET

TÜRKÇE SANAL KLAVYELERİN PERFORMANS OPTİMİZASYONU

Sinan Uşşaklı
Bilgisayar Mühendisliği, Yüksek Lisans
Tez Yöneticisi: Prof. Dr. Bülent Özgüç
Mayıs, 2004

Tablet ve el bilgisayarları için en önemli sorunlardan biri metin girişidir. Bu tezde, Türkçeyi analiz ederek ODTÜ Derleminden ve *ek$i sözlük*ten oluşturduğumuz iki değişik çift yönlü grafik değerleri oluşturulmaktadır. Fitts'in kanunu ile insan elinin hareketlerinin nasıl modelleneceğini açıklanmakta, Metropolis algoritması ile Kernighan ve Lin'in buluşsalından nasıl en iyi sanal Türkçe klavyenin oluşturulacağı anlatılmaktadır. Kernighan ve Lin'in buluşsalının Metropolis algoritmasına göre avantajları tartışılmaktadır.

Anahtar Sözcükler: Sanal klavye tasarımı, Metropolis algoritması, KL buluşsalı Türkçe klavye, el bilgisayarları, çift yönlü grafikler

Acknowledgements

I acknowledge my supervisor Prof. Dr. Bülent Özgüç and Prof. Dr. Cevdet Aykanat for their valuable help and support they extended throughout this study and Assist. Prof Dr. Kürşat Çağıltay for providing initial ideas and valuable feedbacks.

for endless loving, caring and understanding of my mother

for the support of my family

for the memory of my father

Contents

Table of Figures

List of Tables

Chapter 1

Introduction

Today, with the increasing number of mobile devices in our daily life and with more people using them, we have more embedded, small and portable computers such as PDAs, Tablet PCs, and Smart Phones. With time, our overall computer interaction starts to depend more on them. Even though we start to use these devices, there is a major problem that impedes the usability of these devices: text input.

The text entry on both larger portable computers without keyboard, and the small PDA sized computers currently available with reduced keyboard sizes, is done through graffiti, on screen virtual keyboards, voice recognition and handwriting recognition. Most of these methods do have limitations as MacKenzie and Soukoreff [10] described in their survey.

The focus of this study is on creating an optimum design of virtual keyboard for Turkish text entry. There are several papers published, focusing on the design and optimality of virtual keyboards in English such as ATOMIC by Zhai, Hunter, Smith [19]; OPTI and OPTI II keyboard by MacKenzie and Zhang [11]; Textware™ Solutions' FITALY [16]; Lewis, Kennedy, LaLomia [8]'s layout.

Today, except for the work on Turkish F keyboard, there is no keyboard design focusing on Turkish. Therefore, the main aim of this study is to prove that there is a need for a research on Turkish virtual keyboard design, and suggest a new layout using mathematical modeling methods.

Chapter 2

Current Turkish Keyboard Layouts

Currently there are two accepted keyboard layouts for Turkish text input: the Turkish F keyboard and QWERTY keyboard with Ç, Ğ, Ş, İ, Ö, Ü key additions.

The historical reasons of English QWERTY layout is arranged in a way that the adjacent keys that are on the keyboard have low frequency of typing one after the other, when used in typewriters, this minimizes the jamming of the typewriter [1]. This layout is accepted throughout the world and in the Turkish version; Turkish characters are added to the right side of the keyboard.

Figure 1 Turkish Q keyboard [17]

The F keyboard is specifically designed for Turkish language, and therefore has a better performance when typing. The most used vowels are placed in the middle row, left hand, and most used consonants are placed at middle row right hand. This layout can be compared to Dvorak keyboard, which has similar attributes [2]. Unfortunately, the proposed Dvorak layout has only a %15 increase over QWERTY layout. The learning time penalty for Dvorak layout surpasses the %15 speedup gain.

Figure 2 Turkish F keyboard [17]

Both Turkish QWERTY and F keyboard designs are for two hands. If we consider a Virtual Keyboard where we enter text with taps of a stylus[1], both of the keyboards will have very bad performance levels.

The Turkish QWERTY keyboard has distributed vowels and is designed without regard for properties of Turkish language: for example, one of the most used letters 'İ' is added to the rightmost side of the standard QWERTY keyboard. This leads to major difficulties even at typing on physical keyboards as discussed in Section 3.3.

The F keyboard's most used vowel-consonant grouping layout will lead to move the stylus back and forth over the keyboard. In Section 3.6, performance analyses of both Turkish keyboards are shown. The change from ten finger typing on physical keyboards to stylus tapping to virtual keyboards invalidates any performance analysis made for physical keyboards.

The key to finding the optimum virtual keyboard is to find a layout, where the distance that stylus travels over the keyboard is kept at minimum.

[1] the pen for pointing on the PDA devices

Chapter 3

Performance Modeling Of Virtual Keyboards

Let a letter in alphabet expressed by ℓ_i, let k_i represents a key in a given keyboard layout and let $key(\ell_i)$ define the fact that letter ℓ_i is assigned to key k_i.

3.1 Digraph Frequencies

P_{ij} is the frequency of the letter ℓ_i to follow the letter ℓ_j among all the combinations in a digraph. In other words, P_{ij} is the ratio of the transition count from ℓ_i to ℓ_j over all the character transitions in a digraph. In order to calculate P_{ij}, we created a table containing all the character transitions including the transition count from ℓ_i to ℓ_j.

We have used three English digraph tables for the comparison of English versus Turkish digraph statistics. First table is Soukoreff and MacKenzie's [15] table, which was generated using the data provided by Mayzner and Tresselt [12]. The other two digraph tables are generated from the news articles and the chat room logs by Zhai et al. [19].

In order to create Turkish digraph tables, we have used two sources. Our first source is the corpus of first Turkish Corpus project organized by Middle East Technical University (METU) [13]. The other digraph is created using one of the most popular web sites in Turkey: *ek$i sözlük*. *Ek$i sözlük* is a Turkish term that stands for "sour dictionary". It's an online dictionary that users could contribute as in sites like Wiki[2], H2G2[3] or Everything2[4]. We used *ek$i sözlük*'s 2.6 million entries that are entered by about six thousand writers since February 1999 [5].

METU Turkish Corpus consists of a collection of 2 million words of post-1990 written Turkish samples where the words of corpus were taken from 10 different genres. At most 2 samples from one source are used, where each sample is approximately 2000 words. Some of the text is taken from Adam Publishing House, Atlas Magazine, Bilgi Publishing House, Bilim ve Ütopya Magazine, Bütün Dünya Magazine, Can Publications, Cumhuriyet Newspaper, Doğu-Batı Magazine, İletişim Publications, İş Bankası Kültür Publications, Kuraldışı Publishing House, Milliyet Newspaper, Radikal Newspaper, Yapı Kredi Publications. Therefore, this corpus is representing a wide range of formally written text.

The rationale behind the selection of *ek$i sözlük* is that we did not want to take samples from completely non-structured Turkish sentences that are sourced such as IRC chat logs, or instant messages. We would like to model the keyboard usage of the Internet users under an environment that has some common-sense rules. The entry structure of *ek$i sözlük* is generally complete, everyday sentences, mostly consist of several sentences. Instead of instant messaging or chat messaging, the text in *ek$i sözlük* is better thought and more complete. Therefore we believe it models the Turkish user's keyboard usage better. However, some percentage of the words in this site cannot be considered as perfect Turkish. This is because some of the characters of words are replaced by other characters as:

İ → I, Ö → O, Ç → C, Ğ → G, Ş → S or $ …

---

[2] http://c2.com/cgi/wiki

[3] http://www.bbc.co.uk/dna/h2g2/

[4] http://www.everything2.com

This is because of bad placement of letters İ, Ö, Ü, Ç, Ğ, Ş and some users' old habits from the days that there were only English keyboards.

With this information we have constructed two different tables for Turkish and three different tables for English.

3.2 Digraph Coloring

In order to see the difference between two Turkish and English digraphs, we created a program that colors a table with respect to the digraph values. This will help us to oversee if the resulting keyboards will change dramatically. There are two versions of this program, one in grayscale, the other in color. A more detailed color version is available; however the output is only for color prints or screen display. Figure 3, Figure 4 and Figure 5 shows comparisons between Turkish and English digraphs. The first row and the first column of these figures represent the space character; the color varies from white to black as the usage ratio of ℓ_i - ℓ_j pair increases. Y axis represent the first letter; X axis represent the second letter.

On Figure 3, there is a very small amount of difference between the left part (*ek$i sözlük*) and the right part (METU Turkish Corpus). The main thing to notice is that the usage ratios of İ, Ö, Ü, Ç, Ğ, and Ş are partly moved to the usage ratios of I, O, U, C, G, and S accordingly, we will explain the reasons behind the ratio change in letter usage later in this chapter.

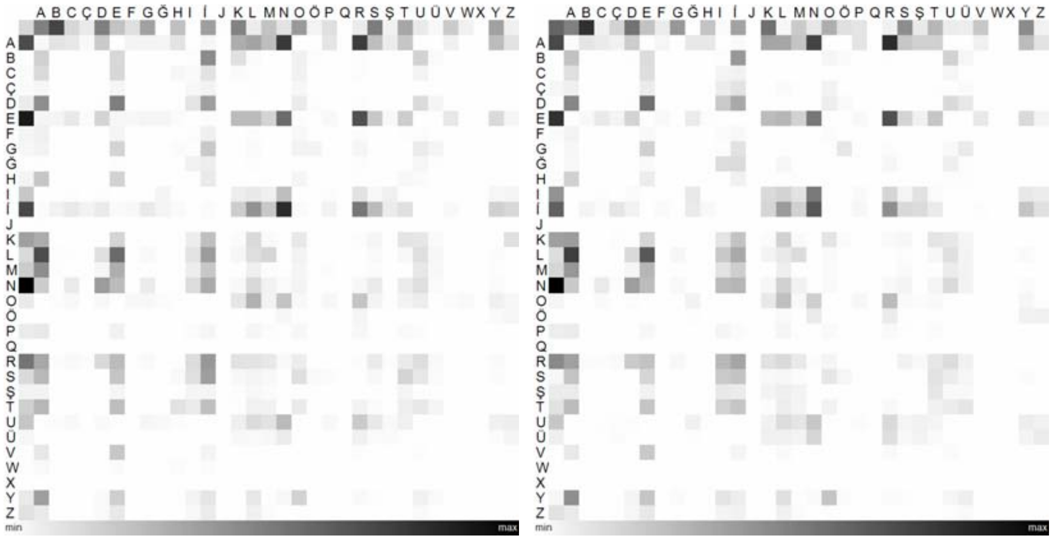

Figure 3 *ek$i sözlük* (left) and METU Turkish Corpus (right) digraph coloring results

As we compare the outputs of digraph coloring of METU Turkish Corpus and Soukoreff and MacKenzie's table [15], we immediately see totally different coloring between two tables as seen in Figure 4. As "*the*" is the most used word in English [7], we can see at the English Digraph table (Figure 4, left), that the usage ratios of Space versus T, T versus H, H versus E, E versus Space is highest. As Turkish language does not have such a frequent and short word, Turkish Digraph table (Figure 4, right) is more spread.

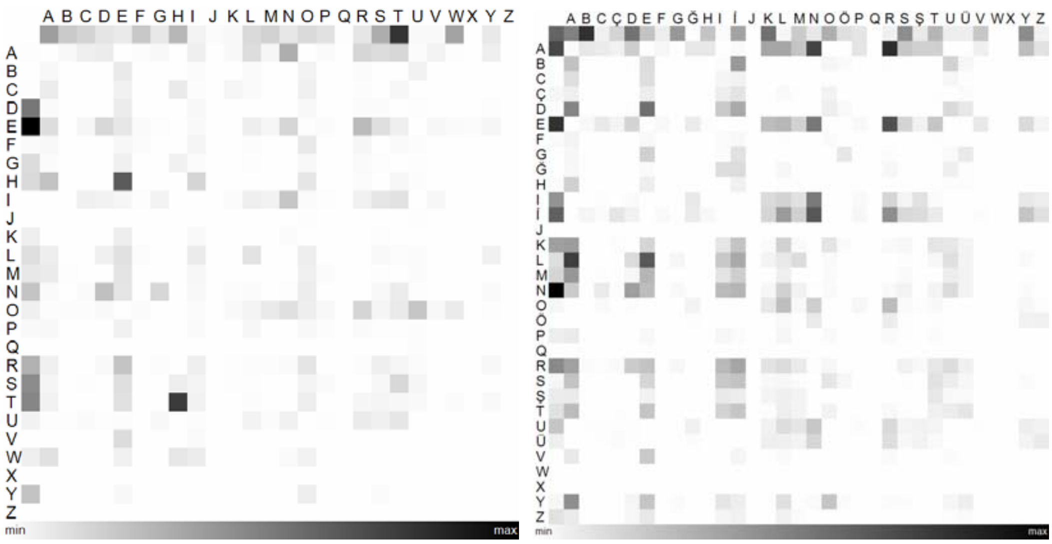

Figure 4 Soukoreff and MacKenzie (left) and METU Turkish Corpus (right) digraph coloring results

As seen in Figure 5, we see very similar results as we compare the outputs of digraph coloring of Soukoreff and MacKenzie's table [15] and Zhai et al.'s chat table [19]. Here, the chat and news Digraph results are more spread to other letters than E, H, and T. This shows that the usage of word "*the*" has decreased since 1965 as Sourkoreff and MacKenzie is using the data collected by Mayzner and Tresselt in 1965 [12].

Figure 5 Soukoreff and MacKenzie (left) and Zhai et al.'s news (middle) and chat (right) digraph coloring results

These results show us that there is a need of Turkish keyboard design, because any design that will base on digraph results of English corpora will eventually lead to performance decrease in Turkish typing. Also depending on these results, we can expect that a design that is built on an weighted average of Turkish Corpora that we collected will lead to very similar word per minute (wpm) performances on changing user profile such as for everyday use versus more formal use.

3.3 Letter usage frequency difference between formal and informal text

We also created letter usage frequency tables for both of the corpora we have, as seen on Figure 6 and Figure 7.

The usage of letters I, O, U, G, C, and S are more in *ek$i sözlük* than METU Corpus letter frequencies. The reason is that ekşi sözluk users tend to use letter I instead of letter İ, letter O instead of Ö etc., even though the resulting word can be considered as no longer Turkish. The reason of this usage habit is the popular Turkish QWERTY keyboard which is actually not designed for Turkish language, therefore İ, Ö, Ü, Ğ, Ç, Ş letters were “added” to the right hand side of the keyboard causing a non-natural layout for Turkish. Second reason is that English programs and English operating systems used to be more widespread than Turkish programs and most people use English when using computers. Finally, huge majority of Turkish words with İ, Ö, Ü, Ğ, Ç, Ş letters do not have another meaning when these letters replaced by I, O, U, G, C, and S. This results in a tendency of English character set usage in computer interaction in Turkey. However, with increasing availability of Turkish software, and as computers getting spreader each day in his situation is rapidly changing today.

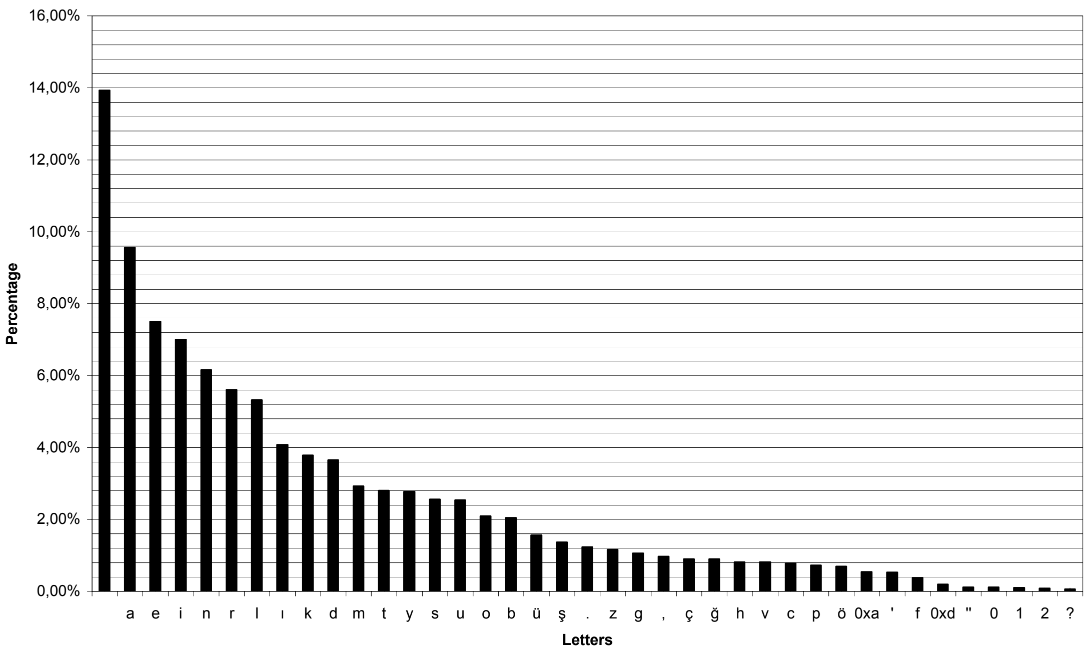

Figure 6 METU Corpus letter frequencies

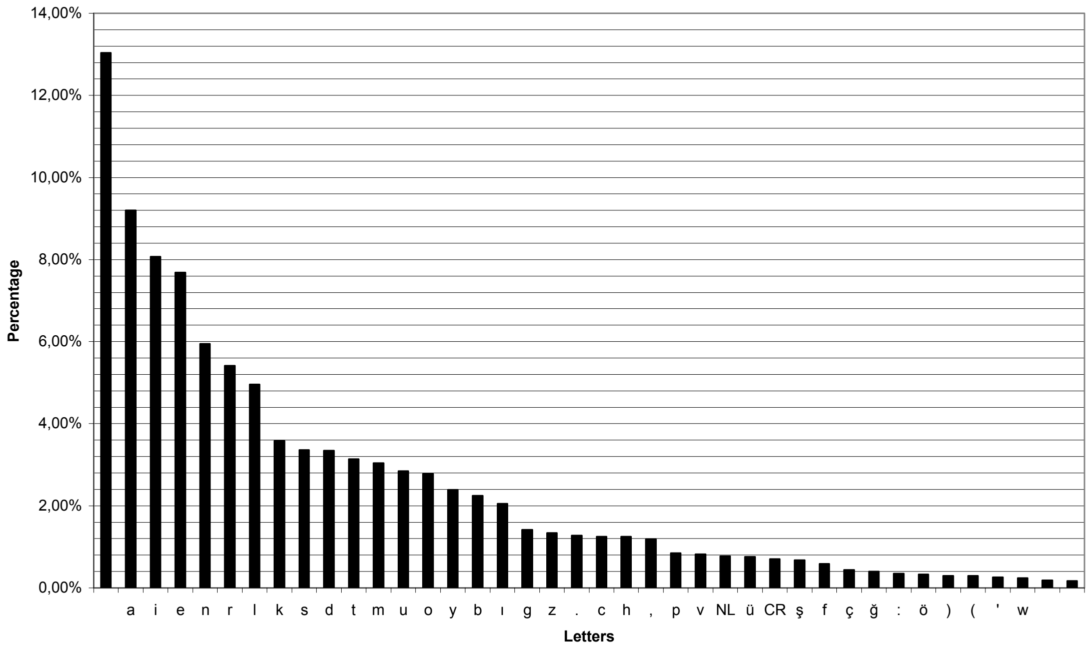

Figure 7 *ek$i sözlük* letter frequencies

3.4 Fitts' Law

The average movement time of a finger or a pointing device between a source key and a destination key can be predicted according to Fitts' Law [3] which is displayed in Figure 8.

Fitts also introduced "Index of Performance" (IP) value which is a coefficient that multiplies speed of hand movement between keys. Mackenzie et al. [9] found an IP value as 4.9 bits per second (bps). Soukoreff and MacKenzie [15] argued that 4.9bps is very conservative value and duplicated Mackenzie et al.'s work [9] with 14 bps. We will use 4.9 bps as IP in our experiments in order to compare our findings with the others [11], [19] in the literature. Other reasons using IP as 4.9 bps are: first, to be on the conservative side for our findings, as increased IP increases WPM performance of virtual keyboard; and second IP value will not be effective on the side by side performance comparisons as we will be calculating rates of performance.

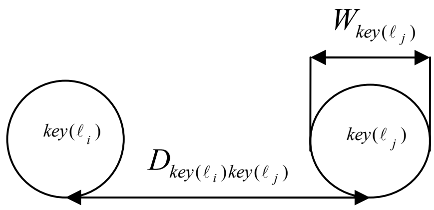

Figure 8 Fitts' model of movement time

The time to move the stylus in seconds from one key $key(\ell_i)$ to another key $key(\ell_j)$ for a given distance $D_{key(\ell_i)key(\ell_j)}$ and key size $W_{key(\ell_j)}$ is:

$$MT(\ell_i, \ell_j) = a + b\log_2\left(\frac{D_{key(\ell_i)key(\ell_j)}}{W_{key(\ell_j)}} + 1\right) \tag{1}$$

Here, *a* and *b* are empirically determined coefficients, where *a* is the waiting time on the key that we will assume to be 0, and *b* is *1 / IP*.

3.5 Mean Time for Typing a Character

As we defined P_{ij} as the frequency of the ℓ_i to follow the ℓ_j among all the digraphs, the mean time in seconds for typing a character as:

$$\begin{aligned} t_{avg} &= \sum_{i=1}^{33}\sum_{j=1}^{33} P_{ij} MT(\ell_i \ell_j) \\ t_{avg} &= \sum_{i=1}^{33}\sum_{j=1}^{33} \frac{P_{ij}}{IP}\left[\log_2\left(\frac{D_{key(\ell_i)key(\ell_j)}}{W_{key(\ell_j)}} + 1\right)\right] \end{aligned} \tag{2}$$

There are two things to note about the calculation of t_{avg}*:* First one is that Turkish keyboard should have 33 characters including space which are *A, B, C, Ç, D, E, F, G, Ğ, H, I, İ, J, K, L, M, N, O, Ö, P, R, S, Ş, T, U, Ü, V, Y, Z,* **Q, W, X**, and *space*. Even though

Q, W and *X* are not a part of Turkish alphabet, they are used in everyday Turkish and also they are included in all other Turkish keyboards; therefore we have to include them in our keyboard design. The original formula that is used by the others is for English and uses standard 27 English characters.

The second thing to note is that $MT(\ell_i, \ell_j)$ part of the formula t_{avg} will return 0 when the ℓ_i and ℓ_j are same, such as when writing “Anne”. This is the time that passes during the double tapping of the same key. According to the experiments of previous researchers [11], and [19] and according to the Soukoreff and MacKenzie’s discussion [14] we are using 0.127 as the $MT(\ell_i, \ell_j)$ time when ℓ_i and ℓ_j are same.

3.6 Turkish Keyboard Performance Analysis

In this section, we will discuss about the mean time calculation for Turkish F and Turkish QWERTY keyboards.

Using (2), we can calculate the mean time in seconds for typing a character. However, there are some keys in current physical Turkish keyboards that we cannot fully model. This is because Fitts’ Law only model circular key shapes. Fitting a circle inside a square letter keys in keyboard is not a problem, but keys such as space bar must be handled in a special way.

The long space bar in today’s keyboards can be sub-modeled by dividing the space key in keyboard into smaller buttons or replacing with a single small button (see Figure 9).

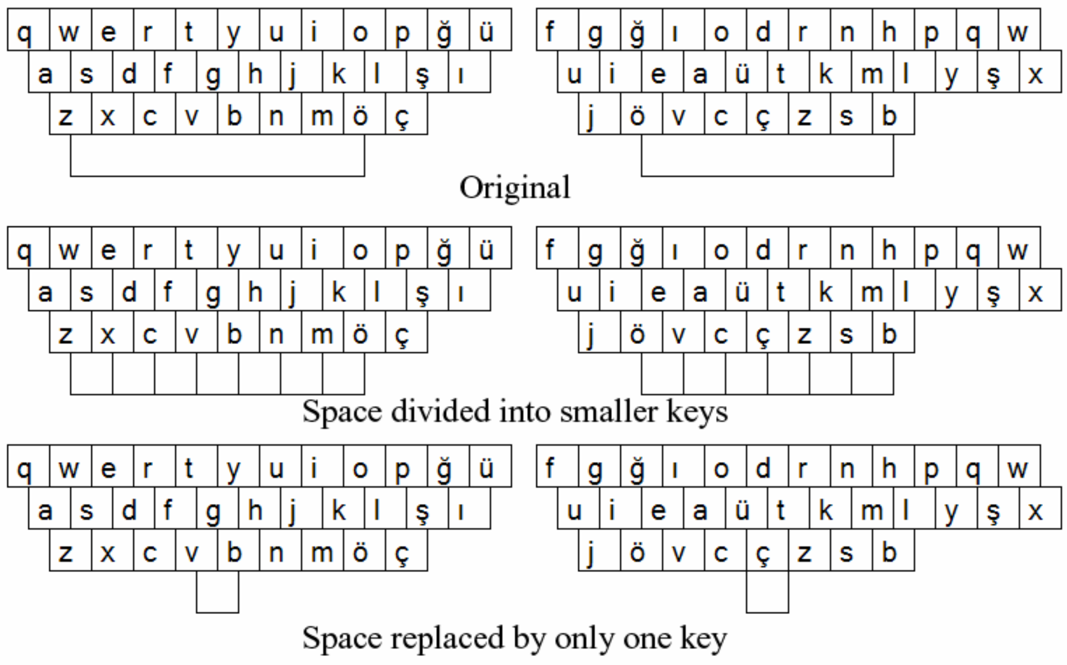

Figure 9 Modeling the space key

The optimistic solution to this several space keys would be finding the nearest space key to the originating key, and then finding the nearest space key to the destination key in a tri-graph key-space-key model (see Figure 10).

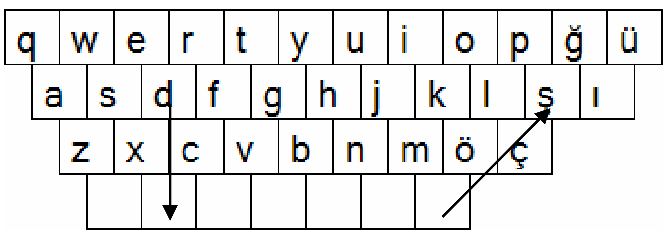

Figure 10 Space key modeled with several small keys

Also we can create a pessimistic solution by placing the long space key by a single key that has the same size with the other keys, placed in the center of the original space key (see Figure 11).

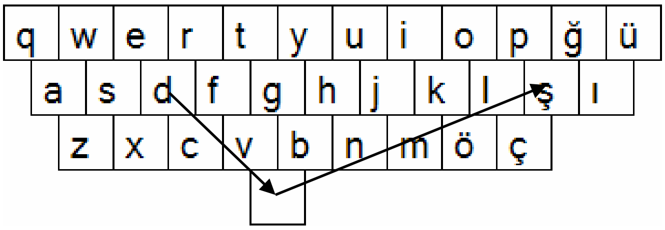

Figure 11 Space Key Modeled with Single Small Key

The former modeling will give us the upper bound, and the latter modeling will result in the lower bound. The optimum key-space-key sequence in a tri-graph of letter-space-letter will be between these two models.

However, in practice when considering optimum $key(\ell_i) \rightarrow key(\ell_j) \rightarrow key(\ell_k)$ sequence, where there are multiple alternatives of $key(\ell_j)$ and there is only one alternative of $key(\ell_i)$ and $key(\ell_k)$, users tap on optimum $key(\ell_i) \rightarrow key(\ell_j) \rightarrow key(\ell_k)$ sequence ranging only %38 to %47 of the time [11]. Therefore we will consider this performance loss when there are multiple space keys.

Table 1 shows the results of Turkish F and Turkish QWERTY keyboards considering both of the models and the multiple space usage performance loss. The results are in Word per Minute (WPM). The word per minute calculation is done assuming that we have 4 letters per word plus a space summing up to 5 characters per word ($60/5*t_{avg}$, where t_{avg} is found using (2)). This assumption is made in order to make comparisons with the other researchers in the literature and to provide a better metric for the reader. The results in parenthesis are sub-optimal user performance modeling for multiple spaces, where the user only taps optimum space key 47 percent of the time.

Table 1 Performance of Turkish keyboards

	METU Corpus (WPM)		*ek$i sözlük* Corpus (WPM)	
	Multiple Space	Single Space	Multiple Space	Single Space
Turkish QWERTY Keyboard	29.30 (28.38)	27.94	29.74 (28.80)	28.35
Turkish F Keyboard	28.46 (27.70)	27.24	28.38 (27.62)	27.13

An important point to note in Table 1 is that Turkish F keyboard is performing worse than Turkish QWERTY keyboard. This finding is expected as Turkish F keyboard is wider than Turkish QWERTY, and it is optimized so that all vowels are at left hand side, and mostly used consonants are placed at right hand side (see Figure 2). This boosts performance on physical keyboards: as we type with two hands, our fingers move least amount of distance. But this model breaks when the keyboard is a virtual keyboard and typing with a stylus requires greater distance as the stylus moves back and forth on the virtual keyboard between vowel and consonant. Turkish QWERTY, however, does not show such grouping and it's less wide and thus performs better.

3.7 Keyboard Layout

It is most likely that the best WPM performance can be achieved by a keyboard with a layout shape as round as possible, with space key at the center. However, placement of a round layout would not be easy as there are space limitations on handheld devices' rectangular screens, and most of the components used today on any graphical user interface based operating systems are rectangular. Therefore we initially created a rectangular layout for our keyboard, which has the 33 characters we defined in digraph frequencies (see Figure 13).

In creation of rectangular keyboard, we used hexagonal keys because it has been found that the average stylus hits on virtual keyboard's key surface has a circular dome (see Figure 12) (Zhai et al. [20]), and fitting a hexagon around a circle wastes less space than fitting a rectangle around a circle (see Figure 13). Another advantage of hexagons is we can create a grid with them.

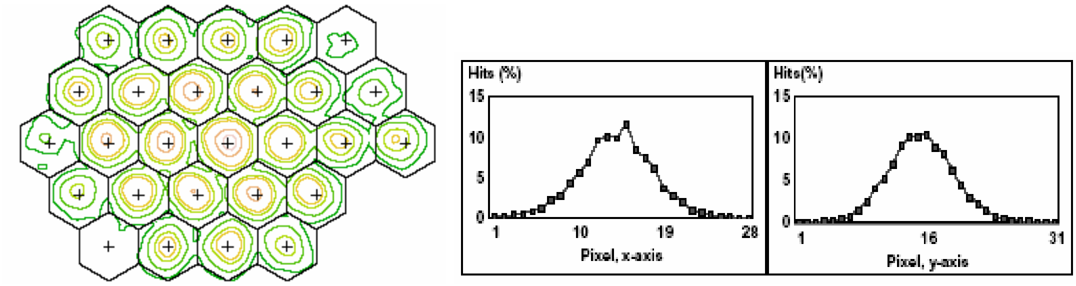

Figure 12 Stylus hit point distribution contour map over the keyboard and 1D plots of average key distribution courtesy of Zhai et al. [20]

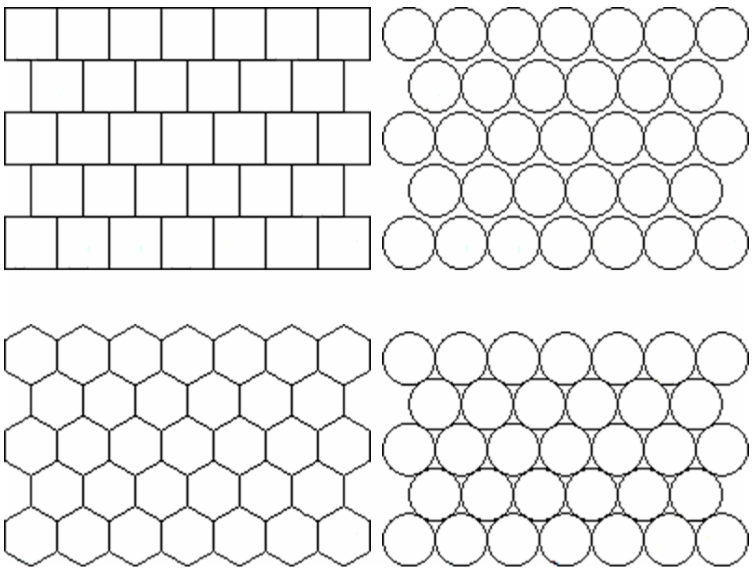

Figure 13 Rectangle versus hexagon fitting around circle

Chapter 4

Fitts-Digraph Energy and Quantitative Design Methods

The keyboard that we design should have minimum average time moving between the keys. We can think of keys of our keyboard as elements, and the distance between the keys as bonds with energy. Nearer the keys, there will be less energy between them. If we consider whole of the keyboard as a system, then nearer the more frequently used keys, less energy the system will have.

Our experiments are based on Fitts' Law (2), which is the function that defines the energy state of the keyboard. More the mean time t_{avg}, the keyboard is less likely to be at the lowest energy state therefore getting away from being optimal. The objective in mapping the letters to a given keyboard layout is the minimization of t_{avg}. Thus, the mapping problem can be modeled as an optimization problem of associating the following quality measures with a good mapping: (i) distance between each key, (ii) digraph frequency of each letter.

A mapping problem instance can be formally represented by one directed graph and one undirected graph, the digraph that we mentioned on Section 3.1, and an undirected graph that is holding the distance between the keys of a given keyboard layout that we discussed in Section 3.7. The objective is to find a one to one mapping of letters in the digraph on a keyboard layout that minimizes t_{avg} in (2).

The problem that we are dealing with is NP-hard. Therefore, we need to use heuristics for solution. In order to find minimum energy state for this system, we can use two different well-known methodologies one used in statistical physics, known as Metropolis Algorithm and the other Kernighan and Lin's heuristic [6]: a widely used technique for VLSI design. Zhai et al. [18] used the Metropolis Algorithm in their paper, we will show same results can be achieved by Kernighan and Lin's (KL) heuristic [6].

4.1 The Metropolis Algorithm

The Metropolis algorithm is actually a Simulated Annealing (SA) algorithm. SA was popularized as an alternative to greedy approaches to graph partitioning and mapping heuristics. The algorithm uses one phase approach. Metropolis algorithm consists of several numbers of these single phases, where at each phase two randomly selected keys are chosen as swapping candidate. Then without changing the configuration, the resulting energy change in t_{avg} caused by this swapping candidate is calculated. Whether this swapping candidate is swapped depends on the Metropolis function [19] shown in (3). If swapping candidate decreases the t_{avg} making a *downhill move* then it's realized. If it increases the t_{avg} making an *uphill move*, then it is realized with a probability, which decreases with the amount of increase in t_{avg}, in order to avoid trapping in local minima. Acceptance probabilities of swaps that increase the t_{avg} are controlled by a temperature parameter T which is decreased using annealing schedule. However, how parameter T is scheduled is not clear and constants changes with respect to problem definition.

$$\begin{aligned} W(O-N) &= e^{-\Delta E/kT} \quad if \quad \Delta E > 0 \\ &= 1 \quad if \quad \Delta E \leq 0 \end{aligned} \tag{3}$$

Each phase is accepted by the function W (O – N), where O denotes old configuration at previous phase, and N is the new configuration. E is the energy change, k is a determined coefficient, and T is Metropolis temperature, which could be interactively adjusted. Periodically our program set T to high values to get out of from local energy minima and then lower the *T* value to reach to the lowest point on local minima. This process is also known as annealing on Metropolis function. The complexity of a single iteration of Metropolis Algorithm we implemented is O(n), as the algorithm stops with respect to *T* value and constants it is not possible to compute overall complexity.

```
KL(key[] K, letter[] A)
random ← a random number generator
while true do
  ℓ_i ← A[random.nextInteger([0,|A|-1])]
  ℓ_j ← A[random.nextInteger([0,|A|-1])]
  s ← swapGain(key(ℓ_i), key(ℓ_j))
  if (s > 0) then
    if (e^(-s/(k*temp)) <random.nextDouble([0,1]) then
      //reject
    else
      //accept
      swap(key(b_i), key(b_j))
```

Figure 14 Metropolis algorithm adapted to our keyboard layout problem

4.2 The Kernighan and Lin's Heuristic

The Kernighan and Lin's (KL) heuristic [6] is introduced in 1970. It's considered as first good graph bisection heuristic, and widely used in VLSI designs. KL has an advantage of climbing *uphill* for getting out of local minima. KL consists of a parent phase that consists of a series of passes. Algorithm starts from an initial random state of keys. Algorithm first unlocks each key and enters series of passes. At each pass, algorithm first selects a pair of unlocked keys, which results most energy gain calculated by Fitts' Law, swaps them, records the swap operation and the energy gain, and locks them. The swapping operation is valid even though the energy gain is negative. The series continues until there are no pair of keys remains unlocked for swapping. Then, algorithm calculates a prefix sum on recorded energy gains, and undoes the swaps until the highest overall energy gain swap. Even though through the swapping process there are some swaps with negative gains may exist, but there possible proceeding swaps with positive gains which may result in a higher overall gain making an *uphill move* to avoid local minima. Such positive gains are caught by prefix sum. The complexity of our implementation of KL is $O(n^4)$.

```
KL(key[] K, letter[] A)
while true do
  best ← −∞
  unlockCount ← |K|
  E ← allocate |K| / 2 element array of reals.
  Swaps ← allocate |K| element array of keys.
  //Unlock all letters
  for each letter ℓ_i ∈ A do
    unlock(ℓ_i)
  while unlockCount >= 2 do
    //Find best swap
    for each ℓ_unlocked ∈ A as ℓ_i do
      for each ℓ_unlocked ∈ A as ℓ_j do
        if ℓ_i ≠ ℓ_j then
          if swapGain(key(ℓ_i), key(ℓ_j)) > best then
            best ← swapGain(key(ℓ_i), key(ℓ_j))
            b_i ← ℓ_i
            b_j ← ℓ_j
    //do the best swap
    swap(key(b_i), key(b_j))
    E.add(best)
    Swaps.add(key(b_i))
    Swaps.add(key(b_j))
  //find highest prefix sum on best swaps array,
  best ← E[0]
  bestIndex ← 0

  for index ← 1 to |K| / 2 do
    E[index] ← E[index] + E[index - 1]
    if E[index] > best then
      bestIndex ← index
      best ← E[index]
  //undo up to the swap that results highest prefix sum
  for index ← 15 downto bestIndex do
    swap(Swaps[index*2], Swaps[index*2+1])
  if best ← E[0] and best < 0 then
    break
save(K)
```

Figure 15 KL heuristic adapted to our keyboard layout problem

Chapter 5

Results and Comparison of Two Algorithms

5.1 Results

We applied Metropolis Algorithm and KL heuristic to a 7x5 keyboard layout, where the second and fourth rows have 6 keys leading to 33 keys that we mentioned at Section 3.7. We took the digraph frequencies are taken from METU Turkish Corpus, and let our implementation run until no improvement is seen. The resulting keyboard is shown in Figure 16. This keyboard has 40.95 word per minute (wpm) according to METU Turkish corpus, 40.56 wpm according to *ek$i sözlük* corpus.

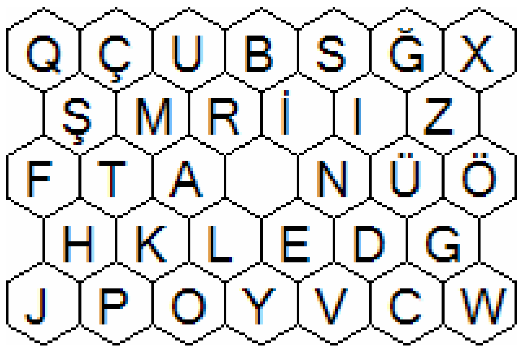

Figure 16 Turkish Keyboard optimized by using Fitts' Law on Metropolis algorithm and KL heuristic

We compared the Turkish QWERTY keyboard and Turkish F keyboard with our keyboard layout by using METU Turkish Corpus and *ek$i sözlük* corpus. As seen on Table 2, and Figure 17, our keyboard has a performance increase of %43 to %46 over Turkish QWERTY keyboard and a %50 performance increase over Turkish F keyboard.

Table 2 Comparison of current keyboards with our result

	Space Modeling	METU Corpus (WPM)	*ek$i sözlük* Corpus (WPM)	Performance of our result	
Turkish QWERTY keyboard	Multiple	29.30	29.74	140%	136%
	Single	27.94	28.35	147%	143%
Turkish F Keyboard	Multiple	28.46	28.38	144%	143%
	Single	27.24	27.13	150%	150%
Our result	-	40.95	40.56	-	-

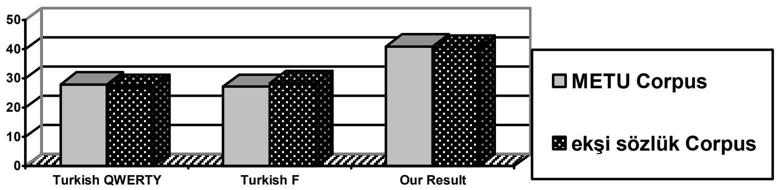

Figure 17 Performance comparison chart

5.2 The Metropolis Algorithm versus KL Heuristic

Even though we first used Metropolis Algorithm to find the best solution 40.95 WPM, we had difficulties achieving this result. There is no clear information over our problem space on how we will control the variable temperature value of Metropolis. Even though we decreased the temperature very slowly [4], there is a still unknown variable k on (3) that multiplies the effect of temperature, changing its effect, we adjusted k using common sense and examining how the keyboard's performance level reacted to the algorithm. Therefore we had to interactively adjust the variables, and spend days for finding the result. After more extended search we found that geometric cooling schedule such as setting $T = \alpha^M T_0$ (where T is the current temperature, M is number of moves made, and T_0 is initial temperature) improves the algorithms performance very effectively, and with several fine tunes the result time dropped to 8 minutes on a P4 2.2GHz machine.

However, KL heuristic required no user controlled variables, and the heuristic found 40.95 WPM result 168 times in 100,000 runs from random initial layouts. The tests ran on a P4 2.2GHz machine and it took the heuristic an average of 10 minutes to find a result. The heuristic suffers from high complexity. When compared to our initial runs of Metropolis Algorithm, and when we consider there is no variables to control, KL heuristic is a better choice.

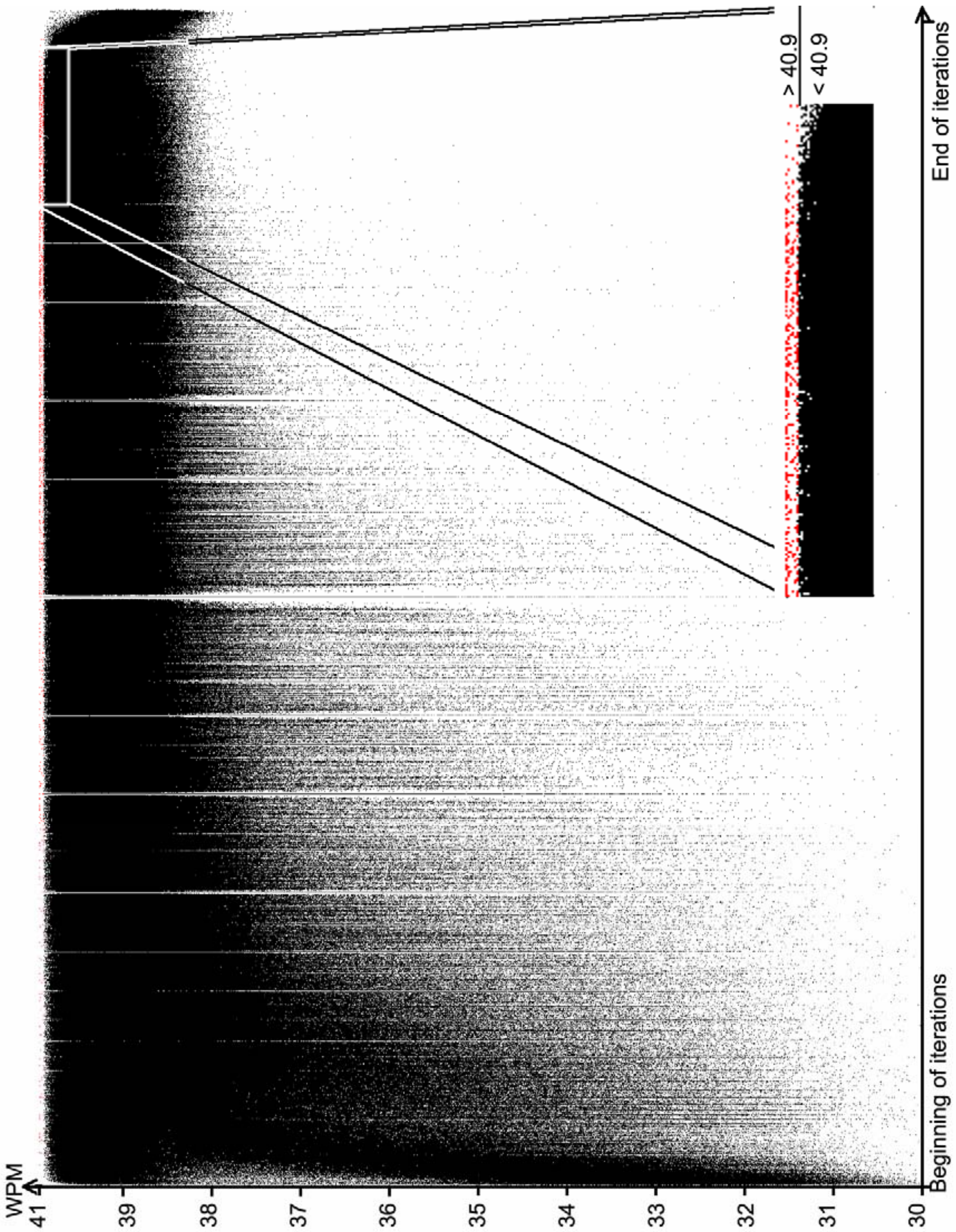

Figure 18 Trace of 100,000 runs of KL heuristic

Figure 18 is trace of 100,000 runs of KL heuristic. Each 16 swaps' WPM values of whole iterations from initial random keyboard state to final optimized keyboard state are stretched through the graph, and WPM value is represented in the graph. This graph is showing WPM values starting from 30 WPM to 41 WPM. As careful eyes notice, the heuristic converges to 40.9 level very quickly (the value 40.9 is at the end of condensed part at the top, where a part to the right corner of it is zoomed in the figure). Also, the heuristic finds several results (majority of them being 40.95, the best solution we have found) higher than 40.9, where the optimum results are getting more condensed toward the end of the iterations.

5.3 Turkish Corpus Statistics and Performance Comparisons

In order to understand how our keyboard performs, we created some statistics on METU Turkish Corpus. These statistics are: the most used words and the most used character sequences, including space, in a sequence of 2, 3, 4 and 5 characters. Table 1 shows the distance stylus moves while writing a word in units of diameter of button's inner circle, and time it takes to write the word/sequence calculated by Fitts' Law. The results are compared between our proposed keyboard layout, Turkish F keyboard and Turkish QWERTY keyboard with optimistic modeling of space (see Table 3).

Table 3 The top 10 most used words, and comparison of 40.95 versus Turkish F and Turkish QWERTY keyboards

		40,95		F (optimistic)		QWERTY (optimistic)	
Word	Frequency	Distance (diameter)	Time (sec)	Distance (diameter)	Time (sec)	Distance (diameter)	Time (sec)
bir	48798	2	0,41	11,19	1,1	6,83	0,87
ve	34469	1	0,2	1,12	0,22	2,83	0,4
bu	22905	1	0,2	7,57	0,63	2,24	0,35
da	19647	2,65	0,38	1,8	0,3	2	0,32
de	19468	1	0,2	2,69	0,38	1,12	0,22
için	10927	6,29	0,97	12,87	1,46	7,89	1,14
o	8100	0	0	0	0	0	0
çok	7899	4,61	0,65	4,93	0,73	3,35	0,57
gibi	7289	4,65	0,79	14,27	1,41	8,35	1,18
ne	7053	1	0,2	4,61	0,51	4,47	0,5

In our proposed keyboard, stylus spends least amount of distance and time compared to the other keyboards. The performance gain of our proposed keyboard comparing distance moved is more than %700. The tables for 2, 3, 4 and 5 character sequences, and more complete version of Table 3, is available at Appendix B. These tables show that our proposed keyboard outperforms current Turkish layouts on very high majority of substrings that user might tap with the stylus.

Chapter 6

Conclusion and Future Work

6.1 Conclusion

In this study we introduced the first virtual keyboard design for Turkish. We introduced several methodologies, and analyzed Turkish language for obtaining best possible results.

First, we introduced Turkish keyboards that are used today, and we mentioned a bit of history of QWERTY and F Keyboards and discussed even though Turkish F keyboard is better for 10 finger typing, its still offers no advantage to be used as a virtual keyboard.

Second, we introduced our metric: the Fitts' Law and its role in our keyboard design. We tried to keep to the same assumptions as other researchers did. We introduced mean time for typing a character, based on Fitts' Law.

Third, we introduced digraphs and created two digraph tables from METU Corpus, and *ek$i sözlük* corpus. In order to show the similarities between these two digraphs, and the distinction between them and the English corpus digraph, we created digraph coloring and did comparisons between the digraphs. We showed the letter usage frequencies in Turkish corpuses, and stated that there is a letter usage difference between more formal text and less formal text. In this chapter, finally, we discussed the ways to analyze Turkish QWERTY and Turkish F keyboards, and calculated their word per minute performances.

Fourth, we introduced the Fitts'-Digraph Energy function used with both Metropolis Algorithm and KL heuristic. We mentioned that the layout choice of our keyboard will have hexagonal keys as stylus' hits on virtual keys have circular domes, and circles fit into hexagons better. Then, we applied Metropolis Algorithm to our setting and found the best Turkish virtual keyboard. We then extended our study to other possible algorithms, and found that KL heuristic outperforms Metropolis algorithm.

Fifth, we showed that the resulting keyboard has 40.95 wpm which is a %43 to a %50 improvement over the current keyboard layouts. We analyzed KL heuristic's performance by running logs. Finally, we stated that performance of our result is very satisfying.

6.2 Future Work

There are several limitations of our work. Firstly, we have not considered any learning curve optimization for possible users of this keyboard. Zhai et al. [19] have added several energy functions to their Metropolis solution in order to ease learning curve. These are alphabetic sorting, and word connectivity enabling users to write very common words in English. These energy functions may be introduced to our keyboard.

Secondly, even though the Fitts' Law has been tested several times, we need to conduct a user test to collect data about the design, and its performance on users. Performance issues such as the angle between the successive keys and other issues such as users' hands' physical capabilities would affect the optimality of this keyboard. Therefore, it is possible to introduce a new metric for stylus input where instead of digraphs consideration we might use digraphs consisting of 3 or more nodes to optimize for any ergonomic performance of human hand with stylus. However this requires a broad research on ergonomics, and results will no longer depend on Fitts' Law.

We did some assumptions as Zhai et al. did in their work [19], for the word per minute calculation. First, every 4 letters plus a space character counts for a word. Second, the entered text of the users might not fit to the digraph frequency tables' results. Third, the user's text entry performance modeled by Fitts' Law has IP index of 4.9 bps. The Fitts' Law cannot be applied on hexagons, therefore we used the inside fitting circle of the hexagon keys in our calculations. Finally, all per minute calculations of words depend on possible advanced users of this keyboard; we did not consider any time spent searching for letters.

Bibliography

[1] Cooper, W. E. (Ed.). (1983). Cognitive aspects of skilled typewriting. *New York: Springer-Verlag.*

[2] Dvorak, A.,Merrick, N. L., Dealey,W. L., & Ford, G. C. (1936). Typewriting behavior. *New York: American Book Company.*

[3] Fitts, P. M. (1954) The information capacity of the human motor system in controlling the amplitude of movement. *J. Exp. Psych. 47*, 381-391.

[4] Hajek, B. ``Cooling schedules for optimal annealing,'' *Mathematics of Operational Research*, 13:311-329, 1988.

[5] Kapanoğlu, S. (1999) *ek$i sözlük*. Retrieved 26.10.2003 from http://sözlük.sourtimes.org

[6] Kernighan, B. W., & Lin, S. (1969) An Efficient Heuristic Procedure for Partitioning Graphs. *The Bell System Technical Journal, February 1970.*

[7] Leech G., Rayson P., & Wilson A. (2001) Word Frequencies in Written and Spoken English: based on the British National Corpus. *pp. 320, Longman, London.*

[8] Lewis, J. R., Kennedy, P. J., & LaLomia, M. J. (1999). Development of a digram-based typing key layout for single-finger/stylus input. *Proceedings of the Human Factors and Ergonomics Society 43rd Annual Meeting*.

[9] MacKenzie, I. S., Sellen, A., & Buxton,W. (1991). A comparison of input devices in elemental pointing and dragging tasks. *Proceedings of the CHI 91 Conference on Human Factors in Computing Systems.*

[10] MacKenzie, I. S., & Soukoreff, R. W. (2002). Text entry for mobile computing: Models and methods, theory and practice. *Human-Computer Interaction, 17*.

[11] MacKenzie, I. S., & Zhang, S. X. (1999). The design and evaluation of a high-performance soft keyboard. *Proceedings of the CHI 99 Conference on Human Factors in Computing Systems*.

[12] Mayzner, M.S. & M.E. Tresselt. (1965) Tables of single-letter and digram frequency counts for various word-length and letter-position combinations. *Psychonomic Monograph Supplements.*

[13] Say, B. METU Turkish Corpus (2002)

[14] Soukoreff, R. W., & MacKenzie, I. S. (2002). Using Fitts' law to model key repeat time in text entry models. Poster presented at *Graphics Interface 2002*.

[15] Soukoreff, W., & MacKenzie, I. S. (1995). Theoretical upper and lower bounds on typing speeds using a stylus and keyboard. *Behavior & Information Technology, 14, 379–379.*

[16] Textware Solutions. (1998). *The Fitaly one-finger keyboard.* Retrieved 17.11.03 from http://fitaly.com/fitaly/fitaly.htm

[17] Turkish Q and F layouts from http://www.microsoft.com/globaldev/keyboards/kbdtuq.htm and http://www.microsoft.com/globaldev/keyboards/kbdtuf.htm

[18] Zhai, S., Hunter, M., & Smith, B. A. (2000). The Metropolis keyboard: An exploration of quantitative techniques for virtual keyboard design. *Proceedings of the UIST 2000 Symposium on User Interface Software and Technology.*

[19] Zhai, S., Hunter M., & Smith, B. A. (2002). Performance Optimization of Virtual Keyboards. *Human-Computer Interaction, 17.*

[20] Zhai S., Sue A., & Accot J., (2002) Movement Model, Hits Distribution and Learning in Virtual Keyboarding. *CHI 2002.*

Appendix A

Digraph Tables

Table 1 Digraph of ek$i sözlük

		a	b	c	ç	d	e	f	g	ğ	h	ı	i	j	k	l	m
	1414018	4526513	6665949	1432988	782278	4568086	2171212	1153097	3358101	301	2255110	60587	3471384	139514	4462462	677491	2170012
a	6399944	258086	999206	833721	241986	1865890	23678	454267	417238	386756	998543	2438	312319	73406	2920244	3303992	2355452
b	40581	1706095	43149	6928	95	17357	1401720	1094	4190	34	1550	110006	4139618	9963	1140934	151361	6721
c	169925	1393182	17006	23708	39	23479	1432075	1368	4123	75	292480	195285	1044193	486	197346	127845	57099
ç	176453	317766	17293	188	1494	461	349738	17	3376	12	1133	284414	462252	9	8378	84817	65218
d	698795	4034983	10800	7416	145	121762	4600286	4496	11873	499	5665	983693	3469079	7634	7054	74789	13100
e	8217889	404480	386040	792923	167365	1505542	296636	256647	329518	334360	157266	536	87875	20273	2449542	2437192	1541063
f	375559	570612	2533	6300	2468	1669	395835	68261	2373	12	3313	108259	548160	283	9527	116980	4333
g	300345	462803	5407	2084	66	16688	1568199	1854	22780	53	148752	105280	1943560	185	1178	143490	33677
ğ	13869	82690	1327	1566	224	17391	109615	545	118	2807	249	526751	620618	5	503	132892	33165
h	372409	2059950	13935	14202	15920	8367	1578298	2395	1119	53	32984	62156	783711	818	17900	67411	64585
ı	1991854	1540	10147	122046	24128	124651	354	30629	33924	360585	5178	19749	338	307	643258	806353	570921
i	6529487	143745	484184	770540	459817	717364	209545	275734	738668	418485	185043	172	153681	15246	1813793	3876581	1974531
j	26943	86004	791	580	5	8375	70542	275	351	6	409	10478	93694	1534	1981	12535	1134
k	3418492	2997919	16077	56940	71326	27646	1540056	4646	4960	263	11888	642110	2300055	649	149287	1339022	212137
l	1352887	6420262	140654	82411	20938	1131381	5332564	40560	225005	332	16591	1028989	3516160	3179	317894	1193864	1470792
m	1831537	4048584	116079	81511	647	318743	2869288	5539	13010	49	16902	638618	2014227	898	46042	540857	114563
n	9135276	1830671	66411	754729	65724	3513786	2441455	38504	650081	106	11818	1231741	3072825	10619	242401	1397299	468041
o	754950	37248	110899	232536	7102	279888	33545	261751	283820	174242	81538	226	58463	55352	899579	2779316	464937
ö	3046	346	13428	4996	3516	20919	1187	7139	6290	65182	2893	17	617	102	39787	137968	18691
p	718125	792672	1891	13810	8364	5109	474093	2422	5312	44	62334	158435	349315	646	28636	346257	120732
q	7319	951	210	93	3	538	256	95	25	1	38	57	450	13	18	1355	157
r	4951989	3128909	129689	241529	120330	1391224	2437030	73864	205635	173	64880	1218343	3691555	27080	1027556	990320	801857
s	1351421	2508410	14101	94290	4532	11028	2078436	19515	19215	41	228514	1147659	3440728	996	284068	429345	282860
ş	465306	447650	6118	2723	10699	1986	354226	4043	6216	28	6582	370710	382220	30	131159	259598	177504
t	1716313	2614834	46779	50859	23666	9391	2346030	21093	7995	215	1049813	769703	2288375	693	82727	533081	492810
u	2073928	81498	116988	275056	52931	179571	87743	54704	342381	250720	137058	416	50901	4238	566364	1256255	1006266
ü	290000	809	12980	54658	73376	39316	8493	22408	7178	78743	14322	26	315	1657	273933	240664	320835
v	56002	847062	10296	26669	22	60231	2022766	1155	69067	15	3616	17996	412915	597	30353	98264	28808
w	119192	225802	2276	1361	2	2813	160173	1628	668	3	151462	312	175812	314	2036	8469	1976
x	47445	6328	509	5083	30	670	10061	927	184	2	693	142	14976	41	103	2750	981
y	775999	3469788	48560	23274	596	221554	1737928	25781	101659	103	12117	461471	898599	317	58905	875464	101495
z	1092135	765147	13789	70225	181	231834	820454	259	72692	40	3952	260898	683116	60	15416	566678	148942

	n	o	ö	p	q	r	s	ş	t	u	ü	v	w	x	y	z
	1034821	3623634	465427	1119695	26807	702672	4691922	522190	3186184	782085	276407	1885244	527112	21305	3392226	542547
a	7097732	19341	357	930493	3994	6959955	2358725	703305	2057656	95816	775	513106	50755	15961	2415900	1029345
b	12673	520039	107531	1596	71	127351	16209	246	4401	1600333	212531	732	2377	334	51849	8082
c	7019	517660	1371	6245	8914	115395	15459	17	107010	455700	91429	1713	462	531	14807	2916
ç	237	255941	23563	48	2	2761	6122	31	43770	21862	68805	44	2	10	1199	18
d	18178	682212	113977	3746	186	135756	64195	126	12965	1281673	391723	11241	7368	1312	75291	1148
e	5269528	71853	392	271048	9144	6297509	1989926	245200	1808318	35847	370	857098	50709	60587	1340923	335579
f	1618	282679	5924	1260	52	162780	23729	542	122890	112222	13064	209	608	802	10645	463
g	27515	460843	423177	3341	53	333347	36977	28	7741	661573	372521	657	2308	80	9321	11969
ğ	4405	2327	32	72	1	138579	2098	125	205	288411	76668	103	8	0	761	11538
h	62458	351664	7267	9500	315	89946	60547	6287	277093	159454	37461	21069	3520	130	29608	5823
ı	2377022	4783	84	128687	141	1046610	281066	550212	67683	357	34	17257	204	69	324429	428799
i	7562586	189209	232	462329	6420	4879156	2425224	699539	1135090	13995	470	216828	1340	26644	1666874	1394350
j	696	51683	2620	2598	12	1208	898	20	516	56634	1286	181	68	71	1692	79
k	93461	813647	94193	4769	39	140689	346813	23433	954721	883483	182982	5392	2695	258	37407	1141055
l	33216	460491	3355	27079	394	9303	236580	980	281011	1153611	297887	26950	9776	324	165706	12461
m	21549	326914	2251	135402	121	31162	169292	6420	14471	839143	283897	1981	2717	1021	130445	15164
n	188934	419961	5606	5397	2895	251467	747045	8494	652389	1146609	203443	17485	3224	2304	211281	70954
o	2121386	227003	84	308063	1473	2085628	406393	76263	435984	642456	51	150920	236324	12784	528475	301258
ö	292127	57	9339	29933	0	338749	50507	10848	70560	22	203	16638	169	23	206764	324233
p	5164	274223	1414	48419	179	265574	93707	1337	190309	115991	19330	803	552	436	20464	422
q	169	301	0	48	146	339	243	2	322	63070	12	78	235	35	134	21
r	222857	644545	7406	67715	2334	116854	729970	114660	743759	1196679	237102	92784	7635	1966	210767	49788
s	48894	1149928	235045	167950	6745	30672	267515	144	1541976	868563	198670	19099	28216	527	154128	1718
ş	2175	9701	25885	1003	8	7528	19958	11499	385373	101546	99267	10819	35	7	6673	600
t	18020	735504	30918	48384	103	479789	152550	258	574057	858784	426917	30195	26746	863	75318	10412
u	2450210	14746	58	264676	514	1868197	802333	226115	545460	35240	92	73036	1050	10102	533349	630019
ü	744261	856	460	58725	19	533847	147932	185387	89292	101	6612	19286	68	105	152950	405404
v	2378	69195	255	788	24	135794	51028	5837	1559	138169	15619	22695	357	1462	17667	8292
w	55409	106300	88	630	66	14420	28038	108	2254	2366	187	1425	79873	159	2492	152
x	285	2571	4	13509	69	197	1123	6	14513	1799	35	507	341	6163	4084	110
y	254755	1578928	54607	11290	51	220186	130721	2880	43505	507787	218039	40850	9039	424	27016	10355
z	7559	106312	1544	718	103	6312	38379	68	6431	206671	103521	1973	782	212	37769	42381

Table 2 Digraph of METU Corpus

		a	b	c	ç	d	e	f	g	ğ	h	ı	i	j	k	l
	170748	139508	229062	15703	48280	156057	66864	20558	114546	10	66159	5381	101947	1271	157396	9467
a	201630	4589	28355	22267	15103	55880	540	10993	1482	26943	26501	44	6257	1678	99851	99618
b	1401	67571	525	68	5	2962	38234	28	9	0	20	4389	113435	92	67	1475
c	461	35847	137	202	0	378	37151	6	8	0	1998	11389	14561	0	370	1123
ç	11066	17185	1843	11	60	8	20586	0	82	0	67	18870	30065	0	583	6398
d	2424	134992	219	24	3	4369	158024	15	179	4	156	60357	95582	73	293	1237
e	227698	1097	7136	23427	12595	48397	739	6802	1531	23175	3952	30	542	663	74790	80092
f	4529	14291	44	10	165	14	7311	598	90	0	45	4536	7957	0	508	3022
g	978	9312	14	13	0	93	51883	5	67	0	282	3829	32532	11	108	402
ğ	819	8222	206	126	16	1193	5832	17	5	1	3	37973	39563	1	8	9047
h	4398	51361	566	48	1078	424	19298	36	6	0	252	2389	15588	2	1623	3093
ı	120422	97	1321	3998	949	4890	21	1576	60	24963	215	331	27	5	40707	49632
i	175972	2395	11484	5390	27391	14825	755	5874	1991	22966	8008	27	2407	205	46283	110715
j	460	881	1	14	0	172	1237	0	1	0	0	546	2386	3	8	314
k	105932	108626	689	139	3703	775	45442	268	221	305	320	29658	66288	3	4863	50367
l	37500	214462	1540	1786	2139	40176	183773	354	9255	24	695	61084	95213	35	12623	24508
m	45879	115907	2060	2901	27	10245	81149	508	493	1	1900	31059	54146	1	1365	18670
n	282299	59773	2805	23661	4127	107325	73163	901	7607	0	481	76309	81450	152	9915	51845
o	10945	228	1823	5588	526	4039	202	1836	1705	14323	601	10	306	1964	25013	71204
ö	163	0	603	201	329	1821	0	350	85	4325	88	0	18	0	2501	7094
p	19876	22670	6	47	301	58	8830	8	51	0	725	10149	5435	0	1381	11024
q	19	4	0	0	0	0	0	0	0	0	0	1	7	0	1	0
r	131327	105163	4934	3701	6264	57673	67081	1080	8456	1	1887	77925	99540	602	32043	40036
s	9095	67102	367	575	118	115	46133	438	246	1	1041	60536	67235	0	7667	9181
ş	21529	22497	2159	48	889	180	14785	158	436	0	234	25411	20140	0	13234	18004
t	32269	74461	1407	311	2252	341	64500	901	152	2	1394	50652	63294	2	4034	18862
u	62327	1673	2349	2822	4234	3808	428	1077	2154	15869	2692	9	287	33	14636	38164
ü	17182	49	550	2707	5431	2894	205	924	33	4216	376	7	5	79	17316	17397
v	1713	25165	139	1097	2	1817	59001	39	1509	0	121	627	10695	2	512	5875
w	270	404	5	1	0	8	264	4	6	0	45	1	223	0	1	14
x	155	40	1	17	1	0	11	13	0	0	5	34	63	0	0	6
y	9555	124303	1398	356	42	11759	63338	710	4505	0	618	29505	18950	0	2353	34518
z	32672	23103	344	2132	66	10133	25119	1	3129	1	156	13063	14537	0	890	16320

	m	n	o	ö	p	q	r	s	ş	t	u	ü	v	w	x	y	z
	58243	25672	89445	34677	29825	39	16167	125706	24339	80355	20113	19548	62070	720	84	127918	14538
a	64844	207194	270	33	28858	27	234181	62324	52479	52608	650	16	17003	72	63	74877	32204
b	1028	117	10033	5996	28	0	2727	53	5	18	49910	10857	12	5	0	96	256
c	230	79	1342	41	10	58	907	49	0	150	10231	3139	6	1	0	47	170
ç	3259	2	13547	1973	5	0	185	316	0	3146	1651	4734	1	0	0	39	0
d	52	137	17963	7514	156	0	2083	417	0	116	41583	26325	123	31	0	1177	12
e	49883	149608	1312	9	6947	5	193705	47680	14656	63280	582	15	25050	332	150	40046	12251
f	28	20	4442	411	17	0	1897	261	13	3020	1422	352	2	2	1	283	9
g	269	108	1318	27396	27	0	3876	154	0	232	5619	23102	39	5	1	34	61
ğ	1501	212	106	2	0	0	9035	143	6	4	18092	4200	2	0	0	3	533
h	2328	919	2828	88	1018	2	1913	1313	283	2654	4250	3541	922	18	0	309	184
ı	32934	147371	231	2	6438	3	46115	13171	33051	4806	53	2	811	0	18	26229	20286
i	51488	185471	844	70	8985	27	124649	42376	38927	23194	105	9	3399	9	71	64350	35312
j	105	5	301	60	4	0	7	7	0	3	279	32	5	1	0	21	1
k	6072	1828	29448	5399	1581	0	6538	9934	1329	29344	26309	12638	253	5	0	1180	70
l	38554	1864	6167	107	557	0	49	4245	103	9869	34769	14214	488	2	0	6104	326
m	1687	703	4470	201	1793	0	864	6321	484	267	20616	13693	99	11	0	533	774
n	15420	3228	5757	254	98	1	7735	13341	498	12397	40379	13720	319	10	1	5440	1839
o	7373	53551	512	0	9665	4	74731	6427	4072	6898	500	0	1361	312	43	9551	2337
ö	1205	23737	0	93	1212	0	21708	2933	838	2950	0	0	1214	1	0	14679	17219
p	2660	59	6085	22	173	0	6356	2217	8	5399	1759	695	5	0	0	242	4
q	0	0	0	0	0	0	0	1	0	0	183	1	0	0	0	0	0
r	24304	3715	10557	705	1856	18	1185	14624	7119	27028	37787	16252	1653	22	88	1341	1001
s	3659	630	26668	10637	2201	29	918	3044	0	32293	20205	10649	268	34	2	5019	16
ş	14311	140	443	1663	79	0	404	1119	52	26827	6664	6892	1178	0	0	445	9
t	12981	415	14303	1771	218	0	7479	2300	0	16406	20768	27772	1455	44	6	354	160
u	27823	62557	387	0	5105	16	41812	12474	14662	11215	176	0	2218	10	62	19250	14866
ü	14101	46410	25	5	1790	0	36048	7909	14212	6165	1	20	1885	0	0	13204	22312
v	450	60	608	4	0	0	6337	468	283	17	5143	625	659	2	0	209	213
w	0	44	182	2	5	0	28	65	0	19	8	0	0	59	0	11	0
x	1	1	18	0	16	0	0	1	0	33	5	0	8	3	20	11	0
y	2357	5777	64407	5084	538	0	5430	3326	257	640	11043	18424	603	27	0	872	308
z	4571	314	3200	50	37	0	158	925	1	158	4350	5726	74	5	0	1559	494

Table 3 Digraph of Sourkernoff, & Mackenzie

	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z	
a	2	144	308	382	1	67	138	9	322	7	146	664	177	1576	1	100	0	802	683	785	87	233	57	14	319	12	50
b	136	14	0	0	415	0	0	0	78	18	0	98	1	0	240	0	0	88	15	7	256	1	1	0	13	0	36
c	368	0	13	0	285	0	0	412	67	0	178	108	0	1	298	0	1	71	7	154	34	0	0	0	9	0	47
d	106	1	0	37	375	3	19	0	148	1	0	22	1	2	137	0	0	83	95	3	52	5	2	0	51	0	2627
e	670	8	181	767	470	103	46	15	127	1	35	332	187	799	44	90	9	1314	630	316	8	172	106	87	189	2	4904
f	145	0	0	0	154	86	0	0	205	0	0	69	3	0	429	0	0	188	4	102	62	0	0	0	4	0	110
g	94	1	0	0	289	0	19	288	96	0	0	55	1	31	135	0	0	98	42	6	57	0	1	0	2	0	686
h	1164	0	0	0	3155	0	0	1	824	0	0	5	1	0	487	2	0	91	8	165	75	0	8	0	32	0	715
i	23	7	304	260	189	56	233	0	1	0	86	324	255	1110	88	42	2	272	484	558	5	165	0	15	0	18	4
j	2	0	0	0	31	0	0	0	9	0	0	0	0	0	41	0	0	0	0	0	56	0	0	0	0	0	0
k	2	0	0	0	337	0	0	0	127	0	0	10	1	82	3	1	0	0	50	0	3	0	0	0	8	0	309
l	332	4	6	289	591	59	7	0	390	0	38	546	30	1	344	34	0	11	121	74	81	17	19	0	276	0	630
m	394	50	0	0	530	6	0	0	165	0	0	4	28	4	289	77	0	0	53	2	85	0	0	0	19	0	454
n	100	2	98	1213	512	5	771	5	135	8	63	80	0	54	349	0	3	2	148	378	49	3	2	2	115	0	1152
o	65	67	61	119	34	80	9	1	88	3	123	218	417	598	336	138	0	812	195	415	1115	136	398	2	47	5	294
p	142	0	1	0	280	1	0	24	97	0	0	169	0	0	149	64	0	110	48	40	68	0	3	0	14	0	127
q	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	66	0	0	0	0	0	0
r	289	10	22	133	1139	13	59	21	309	0	53	71	65	106	504	9	0	69	318	190	89	22	5	0	145	0	1483
s	196	9	47	0	626	0	1	328	214	0	57	48	31	16	213	107	8	0	168	754	175	0	32	0	34	0	2228
t	259	2	31	1	583	1	2	3774	252	0	0	75	1	2	331	0	0	187	209	154	132	0	84	0	121	1	2343
u	45	53	114	48	71	10	148	0	65	0	0	247	87	278	3	49	1	402	299	492	0	0	0	1	7	3	255
v	27	0	0	0	683	0	0	0	109	0	0	0	0	0	33	0	0	0	0	0	1	0	0	0	11	0	0
w	595	3	0	6	285	0	0	472	374	0	1	12	0	103	264	0	0	35	21	4	2	0	0	0	0	0	326
x	17	0	9	0	9	0	0	0	10	0	0	0	0	0	1	22	0	0	0	23	8	0	0	0	0	0	21
y	11	10	0	0	152	0	1	1	32	0	0	7	1	0	339	16	0	0	81	2	1	0	2	0	0	0	1171
z	3	0	0	0	26	0	0	0	2	0	0	4	0	0	2	0	0	0	3	0	0	0	0	0	3	9	2
	1882	1033	864	515	423	1059	453	1388	237	93	152	717	876	478	721	588	42	494	1596	3912	134	116	1787	0	436	2	0

Table 4 Digraph of Zhai, Hunter, & Smith's Chat Corpus

		a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z
	4009	21745	9578	9261	7650	3973	6532	6755	10090	17781	2193	2323	5576	9862	5798	10647	6528	473	4274	12637	28897	2468	1107	14233	116	6056	75
a	7498	465	1809	2574	2828	95	402	1543	1653	2545	100	1584	7926	2721	17032	71	1150	7	8083	6395	12701	1425	2409	539	87	2854	399
b	245	1810	290	34	115	5033	8	9	10	1338	145	4	1831	7	30	2337	6	0	895	257	99	1994	71	24	0	696	10
c	777	4074	15	446	46	3020	2	9	5429	1316	3	2494	829	8	12	3576	13	6	1083	481	1771	688	3	0	0	186	26
d	14705	1610	34	108	305	4356	48	213	51	2814	14	3	228	130	520	4263	6	5	749	1038	27	758	50	27	1	582	15
e	34370	6619	426	2027	5349	3716	990	620	1038	1683	49	342	5488	2220	7563	1049	958	96	13449	8729	3606	181	2544	1036	1313	2232	106
f	5784	1597	0	6	20	1523	1122	89	2	1546	3	15	405	15	9	2872	1	0	1444	64	375	1046	0	2	9	70	0
g	5761	1275	12	5	60	2665	13	294	1875	2153	4	4	546	59	292	3949	0	1	1184	567	43	1108	4	8	0	181	1
h	4624	12788	52	33	37	21947	22	9	414	7138	6	26	57	234	342	4669	12	3	1041	117	1359	1137	13	48	8	775	0
i	6295	1759	687	4239	2615	3136	2381	2018	35	410	16	1141	3110	2719	16266	3034	633	93	2734	9766	9615	86	1543	27	106	42	203
j	51	486	3	11	2	1098	0	5	7	38	16	8	0	0	5	596	3	0	18	9	11	1072	169	18	0	0	0
k	2491	701	17	6	11	3063	15	14	381	2009	169	21	86	13	1171	300	10	0	80	944	63	72	5	50	0	203	0
l	5774	3227	230	165	2253	6082	542	144	17	5395	1	543	7169	143	27	4721	214	0	150	1510	429	716	182	175	6	3350	92
m	4209	4265	412	74	60	7113	28	13	16	1812	1	10	61	1022	219	2506	876	0	31	1145	54	1037	0	8	1	1208	1
n	12907	3056	77	1760	7938	5569	265	7641	43	2349	97	1359	616	75	1453	6352	51	19	108	2630	5478	448	180	37	28	2539	21
o	10884	282	794	1164	3161	863	4755	630	983	974	95	1122	3200	4208	11489	4163	1896	6	8040	2002	4866	11567	1130	3698	115	327	80
p	1205	1441	12	20	6	3161	12	8	708	1724	2	13	2029	46	2	1902	935	0	1911	567	598	581	1	0	1	153	0
q	10	8	2	0	0	1	0	0	0	0	0	0	1	1	1	1	0	0	0	1	0	768	1	2	0	0	0
r	8898	3884	310	750	1500	13277	210	727	100	4628	11	646	1229	511	909	4490	234	1	901	2702	1991	1387	267	133	10	1974	9
s	22762	3168	111	934	57	6075	58	647	2816	2744	11	364	595	634	616	3901	1177	30	52	2298	8368	2183	17	312	5	356	10
t	23464	3453	60	457	45	6273	48	61	25793	6133	13	12	677	107	92	8430	87	2	2334	2391	1605	1493	49	326	7	1268	38
u	4320	686	391	1247	713	1016	411	1001	174	591	19	197	2667	946	2631	137	1182	2	3885	5087	4058	98	11	10	32	492	69
v	310	474	6	7	2	7300	1	5	5	1583	0	0	8	3	9	489	0	0	7	16	8	7	2	0	0	10	0
w	2298	4759	94	3	21	3235	18	7	4332	3350	5	9	138	7	570	2410	5	0	382	365	34	73	1	242	0	19	1
x	251	234	0	114	6	125	1	2	5	330	0	1	6	0	0	17	250	0	2	7	197	42	0	3	88	91	0
y	11413	514	272	68	25	1784	101	118	35	406	1	82	66	147	61	6542	127	1	82	935	401	182	2	130	1	131	19
z	191	94	0	0	3	227	0	1	2	124	0	34	6	2	5	55	0	0	1	2	3	27	0	1	0	96	132

Table 5 Table 4 Digraph of Zhai, Hunter, & Smith's News Corpus

		a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z
	4009	21745	9578	9261	7650	3973	6532	6755	10090	17781	2193	2323	5576	9862	5798	10647	6528	473	4274	12637	28897	2468	1107	14233	116	6056	75
a	7498	465	1809	2574	2828	95	402	1543	1653	2545	100	1584	7926	2721	17032	71	1150	7	8083	6395	12701	1425	2409	539	87	2854	399
b	245	1810	290	34	115	5033	8	9	10	1338	145	4	1831	7	30	2337	6	0	895	257	99	1994	71	24	0	696	10
c	777	4074	15	446	46	3020	2	9	5429	1316	3	2494	829	8	12	3576	13	6	1083	481	1771	688	3	0	0	186	26
d	14705	1610	34	108	305	4356	48	213	51	2814	14	3	228	130	520	4263	6	5	749	1038	27	758	50	27	1	582	15
e	34370	6619	426	2027	5349	3716	990	620	1038	1683	49	342	5488	2220	7563	1049	958	96	13449	8729	3606	181	2544	1036	1313	2232	106
f	5784	1597	0	6	20	1523	1122	89	2	1546	3	15	405	15	9	2872	1	0	1444	64	375	1046	0	2	9	70	0
g	5761	1275	12	5	60	2665	13	294	1875	2153	4	4	546	59	292	3949	0	1	1184	567	43	1108	4	8	0	181	1
h	4624	12788	52	33	37	21947	22	9	414	7138	6	26	57	234	342	4669	12	3	1041	117	1359	1137	13	48	8	775	0
i	6295	1759	687	4239	2615	3136	2381	2018	35	410	16	1141	3110	2719	16266	3034	633	93	2734	9766	9615	86	1543	27	106	42	203
j	51	486	3	11	2	1098	0	5	7	38	16	8	0	0	5	596	3	0	18	9	11	1072	169	18	0	0	0
k	2491	701	17	6	11	3063	15	14	381	2009	169	21	86	13	1171	300	10	0	80	944	63	72	5	50	0	203	0
l	5774	3227	230	165	2253	6082	542	144	17	5395	1	543	7169	143	27	4721	214	0	150	1510	429	716	182	175	6	3350	92
m	4209	4265	412	74	60	7113	28	13	16	1812	1	10	61	1022	219	2506	876	0	31	1145	54	1037	0	8	1	1208	1
n	12907	3056	77	1760	7938	5569	265	7641	43	2349	97	1359	616	75	1453	6352	51	19	108	2630	5478	448	180	37	28	2539	21
o	10884	282	794	1164	3161	863	4755	630	983	974	95	1122	3200	4208	11489	4163	1896	6	8040	2002	4866	11567	1130	3698	115	327	80
p	1205	1441	12	20	6	3161	12	8	708	1724	2	13	2029	46	2	1902	935	0	1911	567	598	581	1	0	1	153	0
q	10	8	2	0	0	1	0	0	0	0	0	0	1	1	1	1	0	0	0	1	0	768	1	2	0	0	0
r	8898	3884	310	750	1500	13277	210	727	100	4628	11	646	1229	511	909	4490	234	1	901	2702	1991	1387	267	133	10	1974	9
s	22762	3168	111	934	57	6075	58	647	2816	2744	11	364	595	634	616	3901	1177	30	52	2298	8368	2183	17	312	5	356	10
t	23464	3453	60	457	45	6273	48	61	25793	6133	13	12	677	107	92	8430	87	2	2334	2391	1605	1493	49	326	7	1268	38
u	4320	686	391	1247	713	1016	411	1001	174	591	19	197	2667	946	2631	137	1182	2	3885	5087	4058	98	11	10	32	492	69
v	310	474	6	7	2	7300	1	5	5	1583	0	0	8	3	9	489	0	0	7	16	8	7	2	0	0	10	0
w	2298	4759	94	3	21	3235	18	7	4332	3350	5	9	138	7	570	2410	5	0	382	365	34	73	1	242	0	19	1
x	251	234	0	114	6	125	1	2	5	330	0	1	6	0	0	17	250	0	2	7	197	42	0	3	88	91	0
y	11413	514	272	68	25	1784	101	118	35	406	1	82	66	147	61	6542	127	1	82	935	401	182	2	130	1	131	19
z	191	94	0	0	3	227	0	1	2	124	0	34	6	2	5	55	0	0	1	2	3	27	0	1	0	96	132

Appendix B

Statistics on METU Turkish Corpus and Comparisons

Table 6 The most used words, and comparison of 40.95 versus Turkish F and Turkish QWERTY keyboards

		40,95		F (optimistic)		QWERTY (optimistic)	
Word	Frequency	Distance (diameter)	Time (sec)	Distance (diameter)	Time (sec)	Distance (diameter)	Time (sec)
bir	48798	2	0,41	11,19	1,1	6,83	0,87
ve	34469	1	0,2	1,12	0,22	2,83	0,4
bu	22905	1	0,2	7,57	0,63	2,24	0,35
da	19647	2,65	0,38	1,8	0,3	2	0,32
de	19468	1	0,2	2,69	0,38	1,12	0,22
için	10927	6,29	0,97	12,87	1,46	7,89	1,14
o	8100	0	0	0	0	0	0
çok	7899	4,61	0,65	4,93	0,73	3,35	0,57
gibi	7289	4,65	0,79	14,27	1,41	8,35	1,18
ne	7053	1	0,2	4,61	0,51	4,47	0,5
daha	6840	6,11	0,97	11,02	1,32	12	1,38
ile	6466	3	0,53	13	1,19	8,38	0,9
ama	6105	2	0,41	8	0,95	13,15	1,19
olarak	5941	5	1,02	17,99	2,21	21,5	2,25
türkiye	5893	14,94	2,14	23,92	2,52	25,73	2,72
sonra	5886	9,38	1,35	10,3	1,45	15,7	1,83
nin	5832	2	0,41	11,18	1,11	4,47	0,69
her	5073	5	0,73	9,23	1,01	4,64	0,66
kadar	5031	7,29	1,17	9,3	1,4	13,69	1,64
ın	4950	1	0,2	4	0,47	4,61	0,51
en	4902	1	0,2	4,61	0,51	4,47	0,5
olan	4719	4	0,73	13,25	1,49	14,71	1,42
ya	4663	2	0,32	6	0,57	4,61	0,51
in	4399	1	0,2	5,59	0,56	2,24	0,35
diye	4030	5,65	0,91	18,64	1,71	9,61	1,24
var	3818	3,65	0,58	3,81	0,61	6,33	0,84

dedi	3759	4	0,73	9,03	1,22	6,85	0,95
iki	3746	5,29	0,76	10	1,06	2,24	0,44
ise	3731	3,65	0,58	10,2	1,06	6,71	0,78
ben	3722	3,65	0,58	10,2	1,06	8,08	0,95
ki	3678	2,65	0,38	5	0,53	1,12	0,22
değil	3520	7,73	1,23	11,93	1,44	13,92	1,58
nın	3441	2	0,41	8	0,95	9,22	1,02
büyük	3435	11,54	1,59	15,64	1,83	21,96	2,18
a	3424	0	0	0	0	0	0
yeni	3246	3	0,61	17,2	1,68	9,71	1,25
içinde	3113	8,29	1,37	17,56	2,17	12,65	1,81
olduğunu	3002	17,9	2,56	29,62	3,27	22,76	2,84
zaman	2903	7,61	1,18	14,33	1,78	19,86	1,97
ilk	2871	3	0,53	9	0,94	2,8	0,51
olduğu	2848	12,61	1,79	16,47	2,07	18,76	2,19
önce	2824	5,73	0,94	10,79	1,3	7,24	1,08
yüzde	2641	6,65	1,11	11,73	1,56	19,12	1,81
yok	2602	2	0,41	8,28	0,94	4,12	0,63
ancak	2561	8,46	1,29	11,36	1,53	18,28	1,96
hiç	2498	6,11	0,82	10,22	1,05	4,63	0,7
bile	2495	4	0,73	19,58	1,78	11,21	1,29
göre	2481	6,61	0,98	10,35	1,3	10,03	1,2
gün	2454	2	0,41	6,33	0,84	11,96	1,14
genel	2453	5	0,94	17,02	1,89	18,21	1,98
son	2440	6,65	0,85	6,61	0,86	9,4	0,99
şey	2421	4,46	0,64	15	1,26	10,57	1,04
iyi	2373	5,29	0,76	16	1,29	4	0,65
mi	2284	2	0,32	6	0,57	2	0,32
kendi	2264	6	1,06	14,25	1,76	18,31	2,02
e	2245	0	0	0	0	0	0
ye	2213	1	0,2	7	0,61	3	0,41
yıl	2204	5,65	0,79	12,17	1,15	7,59	0,88
aynı	2197	5	0,85	12,69	1,43	11,46	1,36
başka	2192	6,73	1,15	18,61	2	22,61	2,12
böyle	2130	8,93	1,29	20,57	1,98	16,82	1,91
türk	2063	8,65	1,18	3,92	0,73	19,61	1,77
istanbul	2058	13,11	2,08	30,84	3,26	23,44	2,88
oldu	2053	6,61	0,98	12,86	1,47	10,76	1,25
bütün	2033	11,65	1,53	8,33	1,24	25,71	2,36
1	2032	0	0	0	0	0	0
arasında	1978	10,29	1,78	21,3	2,82	25,63	2,93
ırak	1968	4	0,73	8,69	1,2	17,26	1,63
3	1952	0	0	0	0	0	0
nasıl	1949	8,29	1,29	17,34	1,96	17,59	1,76
yer	1926	3	0,53	10,64	1,06	4	0,61
başkanı	1923	9,73	1,67	26,25	2,93	32,81	3,18
erdoğan	1922	17,34	2,36	13,08	1,94	24,85	2,55
tek	1901	4,65	0,7	7	0,88	7,59	0,88

abd	1881	5	0,73	8,22	0,96	7,3	0,89
2	1865	0	0	0	0	0	0
mı	1851	3	0,41	4,61	0,51	3,64	0,45
bin	1840	2	0,41	12,17	1,15	5,06	0,74
doğru	1828	10,87	1,47	12,59	1,56	17,59	1,9
şimdi	1818	10,46	1,5	21,33	2,09	13,91	1,72
hem	1814	5,65	0,79	10,59	1,08	9,03	1
karşı	1808	8	1,21	17,87	1,93	17,27	1,8
ilgili	1784	11,65	1,76	29,68	2,69	12,1	1,77
artık	1767	9,8	1,39	10,14	1,44	13,27	1,59
söyledi	1751	11,11	1,76	25,9	2,77	26,13	2,82
bana	1732	6	0,97	11,89	1,41	15,79	1,62
önemli	1729	9,65	1,56	22,99	2,4	15,46	1,98
birlikte	1723	12,02	1,99	29,88	3,24	20,98	2,72
milyon	1718	8,65	1,44	22,59	2,35	13,27	1,88
eski	1676	8,9	1,21	10,73	1,26	8,24	1,01
onu	1675	5,29	0,76	9,58	1	4,83	0,72
benim	1675	6,65	1,11	21,79	2,19	12,31	1,62
kez	1664	4,65	0,7	7,64	0,93	7,83	0,9
5	1655	0	0	0	0	0	0
onun	1636	7,94	1,14	16,15	1,6	6,83	1,04
yine	1633	4,65	0,79	18,2	1,71	8,71	1,17
uzun	1611	9,86	1,28	17,76	1,71	12,77	1,42
çünkü	1605	11,61	1,53	6,93	1,15	13,66	1,7
şu	1595	1,73	0,3	10	0,71	3,64	0,45
üzerine	1594	8,65	1,52	23,89	2,78	24,14	2,58

Table 7 The most used character sequences of length two

		40,95		F (optimistic)		QWERTY (optimistic)	
Word	Frequency	Distance (diameter)	Time (sec)	Distance (diameter)	Time (sec)	Distance (diameter)	Time (sec)
n	282161	1	0,2	3,04	0,41	1,12	0,22
ar	234180	1	0,2	2,69	0,38	2,69	0,38
b	229019	1,73	0,3	1,12	0,22	1,12	0,22
e	227684	1	0,2	2	0,32	3,04	0,41
la	214453	1	0,2	5	0,53	8	0,65
an	207081	2	0,32	3,64	0,45	5,59	0,56
a	201606	1	0,2	2	0,32	2,24	0,35
er	193704	2	0,32	3,64	0,45	1	0,2
in	185301	1	0,2	5,59	0,56	2,24	0,35
le	183754	1	0,2	6	0,57	6,58	0,6
i	175896	1	0,2	2,24	0,35	3,04	0,41
de	158024	1	0,2	2,69	0,38	1,12	0,22
k	157373	1,73	0,3	2	0,32	2	0,32
d	155971	1,73	0,3	3,04	0,41	2	0,32
en	149459	1	0,2	4,61	0,51	4,47	0,5
ın	147233	1	0,2	4	0,47	4,61	0,51
a	139492	1	0,2	2	0,32	2,24	0,35
da	134990	2,65	0,38	1,8	0,3	2	0,32
r	131168	1	0,2	3,04	0,41	3,04	0,41
y	127899	1,73	0,3	2,83	0,4	3,04	0,41
s	125667	2	0,32	1,12	0,22	2	0,32
ir	124643	1	0,2	4,61	0,51	4	0,47
ya	124296	2	0,32	6	0,57	4,61	0,51
ı	120405	1,73	0,3	3,04	0,41	3,61	0,45
ma	115887	1	0,2	4	0,47	6,58	0,6
g	114536	2,65	0,38	3,35	0,43	2	0,32
bi	113428	1	0,2	6,58	0,6	2,83	0,4
il	110711	2	0,32	7	0,61	1,8	0,3
ka	108598	1	0,2	3	0,41	7	0,61
nd	107252	1	0,2	2	0,32	3,64	0,45
k	105837	1,73	0,3	2	0,32	2	0,32
ra	105138	1	0,2	2,69	0,38	2,69	0,38
i	101924	1	0,2	2,24	0,35	3,04	0,41
ak	99851	1	0,2	3	0,41	7	0,61
al	99613	1	0,2	5	0,53	8	0,65
ri	99492	1	0,2	4,61	0,51	4	0,47
di	95575	2	0,32	3,64	0,45	4,61	0,51
li	95135	2	0,32	7	0,61	1,8	0,3
o	89435	2	0,32	3,04	0,41	3,04	0,41
ni	81415	1	0,2	5,59	0,56	2,24	0,35
me	81141	2,65	0,38	5	0,53	5,39	0,55
t	80333	2	0,32	2	0,32	3,04	0,41
el	80087	1	0,2	6	0,57	6,58	0,6
rı	77877	2	0,32	3	0,41	7,57	0,63
nı	76215	1	0,2	4	0,47	4,61	0,51
ay	74790	2	0,32	6	0,57	4,61	0,51
ek	74790	2	0,32	4	0,47	5,59	0,56
or	74731	2,65	0,38	2	0,32	5	0,53

ta	74456	1	0,2	2	0,32	3,64	0,45
ne	73149	1	0,2	4,61	0,51	4,47	0,5
ol	71203	1	0,2	4,61	0,51	1,12	0,22
ba	67569	2	0,32	4,61	0,51	4,61	0,51
si	67202	1	0,2	5,59	0,56	5,59	0,56
sa	67089	2,65	0,38	3,64	0,45	1	0,2
re	67065	2	0,32	3,64	0,45	1	0,2
e	66856	1	0,2	2	0,32	3,04	0,41
ki	66260	2,65	0,38	5	0,53	1,12	0,22
h	66139	2,65	0,38	3,04	0,41	2	0,32
am	64844	1	0,2	4	0,47	6,58	0,6
te	64485	2,65	0,38	3	0,41	2	0,32
yo	64407	1	0,2	5,59	0,56	3	0,41
iy	64329	2,65	0,38	8	0,65	2	0,32
ye	63333	1	0,2	7	0,61	3	0,41
et	63280	2,65	0,38	3	0,41	2	0,32
ti	63240	2,65	0,38	4	0,47	3	0,41
un	62460	2,65	0,38	6,58	0,6	2	0,32
u	62323	2	0,32	2,83	0,4	3,04	0,41
as	62316	2,65	0,38	3,64	0,45	1	0,2
v	62068	2	0,32	1,12	0,22	1,12	0,22
lı	61067	2,65	0,38	5,59	0,56	2	0,32
sı	60516	1	0,2	4,47	0,5	9	0,68
dı	60355	1,73	0,3	2	0,32	8	0,65
na	59735	2	0,32	3,64	0,45	5,59	0,56
ve	59001	1	0,2	1,12	0,22	2,83	0,4
m	58220	1,73	0,3	2	0,32	1,12	0,22
rd	57620	2,65	0,38	1	0,2	1,12	0,22
ad	55699	2,65	0,38	1,8	0,3	2	0,32
mi	54117	2	0,32	6	0,57	2	0,32
on	53534	2,65	0,38	3	0,41	2,83	0,4
at	52608	1	0,2	2	0,32	3,64	0,45
aş	52479	1,73	0,3	7	0,61	9	0,68
ge	51883	2	0,32	1,8	0,3	2,69	0,38
nl	51815	1,73	0,3	1,8	0,3	2,69	0,38
im	51488	2	0,32	6	0,57	2	0,32
ha	51357	1,73	0,3	4,61	0,51	5	0,53
tı	50623	3,61	0,45	2,69	0,38	6,58	0,6
kl	50269	1	0,2	2	0,32	1	0,2
bu	49910	1	0,2	7,57	0,63	2,24	0,35
em	49881	2,65	0,38	5	0,53	5,39	0,55
ıl	49622	2,65	0,38	5,59	0,56	2	0,32
ç	48276	2,65	0,38	1,12	0,22	1,8	0,3
ed	48262	1	0,2	2,69	0,38	1,12	0,22
es	47677	2,65	0,38	4,61	0,51	1,12	0,22
ün	46378	1	0,2	2,69	0,38	5,39	0,55
ik	46283	2,65	0,38	5	0,53	1,12	0,22
ır	46115	2	0,32	3	0,41	7,57	0,63
se	46104	2,65	0,38	4,61	0,51	1,12	0,22

Table 8 The most used character sequences of length three

		40,95		F (optimistic)		QWERTY (optimistic)	
Word	Frequency	Distance (diameter)	Time (sec)	Distance (diameter)	Time (sec)	Distance (diameter)	Time (sec)
lar	100428	2	0,41	7,69	0,91	10,69	1,03
bi	83699	2,73	0,5	7,69	0,82	3,95	0,62
ler	80275	3	0,53	9,64	1,02	7,58	0,8
an	69257	3	0,53	6,68	0,86	6,71	0,78
eri	67751	3	0,53	8,25	0,96	5	0,68
bir	66288	2	0,41	11,19	1,1	6,83	0,87
ya	64916	3,73	0,62	8,83	0,97	7,65	0,92
in	64455	2	0,41	8,63	0,97	3,35	0,57
arı	63779	3	0,53	5,69	0,79	10,26	1,02
ka	63356	2,73	0,5	5	0,73	9	0,94
en	62251	2	0,41	7,65	0,92	5,59	0,72
da	57360	3,65	0,58	3,8	0,63	4,24	0,67
ir	55901	2	0,41	7,65	0,92	7,04	0,89
de	54881	2,73	0,5	5,73	0,8	3,12	0,54
de	53380	2	0,41	4,69	0,71	4,16	0,63
ara	49265	2	0,41	5,39	0,77	5,39	0,77
ba	48456	3,73	0,62	5,73	0,73	5,73	0,73
ın	48123	2	0,41	7,04	0,89	5,73	0,73
ol	47481	3	0,53	7,65	0,92	4,16	0,63
ve	46793	3	0,53	2,24	0,44	3,95	0,62
nda	45063	3,65	0,58	3,8	0,63	5,64	0,78
yor	44544	3,65	0,58	7,59	0,88	8	0,94
ge	44037	4,65	0,7	5,16	0,74	4,69	0,71
bu	41538	2,73	0,5	8,68	0,85	3,35	0,57
ini	39313	2	0,41	11,18	1,11	4,47	0,69
ını	37303	2	0,41	8	0,95	9,22	1,02
ile	36850	3	0,53	13	1,19	8,38	0,9
ası	36001	3,65	0,58	8,11	0,95	10	0,88
rin	35581	2	0,41	10,2	1,06	6,24	0,82
ve	35293	2	0,41	3,12	0,54	5,87	0,81
nde	34157	2	0,41	4,69	0,71	4,76	0,67
n b	34101	2,73	0,5	4,16	0,63	2,24	0,44
sa	33079	4,65	0,7	4,76	0,67	3	0,53
lan	33027	3	0,53	8,64	0,98	13,59	1,2
le	32874	2	0,41	8	0,9	9,62	1,01
den	32352	2	0,41	7,3	0,89	5,59	0,72
da	32350	4,38	0,68	4,84	0,71	4	0,65
ha	31827	4,38	0,68	7,65	0,92	7	0,85
ama	31591	2	0,41	8	0,95	13,15	1,19
ınd	31513	2	0,41	6	0,8	8,25	0,96
rın	31447	3	0,53	7	0,88	12,18	1,14
anı	30563	3	0,53	7,64	0,93	10,2	1,06
ar	29922	2	0,41	5,73	0,8	5,73	0,8
ne	29893	2	0,41	6,61	0,83	7,51	0,91
esi	29004	3,65	0,58	10,2	1,06	6,71	0,78
ak	28904	2,73	0,5	5	0,73	9	0,94
er	28753	3	0,53	6,68	0,86	4,04	0,62

ni	28001	2	0,41	7,83	0,9	5,28	0,76
be	27950	4,38	0,68	6,71	0,78	4,72	0,67
nla	27797	2,73	0,5	6,8	0,83	10,69	1,03
e b	27596	2,73	0,5	3,12	0,54	4,16	0,63
nı	27494	2,73	0,5	7,04	0,89	8,22	0,96
ta	27300	3	0,53	4	0,65	6,68	0,86
edi	27140	3	0,53	6,33	0,84	5,73	0,73
aya	27074	4	0,65	12	1,15	9,22	1,02
ind	26392	2	0,41	7,59	0,88	5,88	0,8
gö	26189	3,65	0,58	5,59	0,78	5,64	0,78
ele	25844	2	0,41	12	1,15	13,15	1,19
dan	25419	4,65	0,7	5,44	0,76	7,59	0,88
a b	25338	2,73	0,5	3,12	0,54	3,35	0,57
nin	25317	2	0,41	11,18	1,11	4,47	0,69
ki	25138	3,65	0,58	7,24	0,87	4,16	0,63
nın	25037	2	0,41	8	0,95	9,22	1,02
na	24886	3	0,53	5,64	0,78	7,83	0,9
ala	24817	2	0,41	10	1,06	16	1,29
ili	24598	4	0,65	14	1,22	3,61	0,61
so	24441	6	0,8	4,72	0,67	8,58	0,92
ola	24395	2	0,41	9,61	1,04	9,12	0,87
ko	24224	2,73	0,5	4,69	0,71	3,12	0,54
eni	24101	2	0,41	10,2	1,06	6,71	0,85
ığı	23758	2	0,41	2	0,41	2,24	0,44
ere	23469	4	0,65	7,28	0,9	2	0,41
sın	23388	2	0,41	8,47	0,97	13,61	1,19
anl	22972	3,73	0,62	5,44	0,76	8,28	0,94
eli	22554	3	0,53	13	1,19	8,38	0,9
bil	22486	3	0,53	13,58	1,21	4,63	0,7
ada	22481	5,29	0,76	3,61	0,61	4	0,65
kar	22464	2	0,41	5,69	0,79	9,69	1
bu	22449	3	0,53	10,39	1,03	5,28	0,76
iye	22147	3,65	0,58	15	1,26	5	0,73
i b	22135	2,73	0,5	3,35	0,57	4,16	0,63
n d	21825	2,73	0,5	6,08	0,82	3,12	0,54
al	21707	2	0,41	7	0,85	10,24	0,99
yle	21700	2	0,41	7	0,78	10,22	1,05
se	21603	4,65	0,7	5,73	0,73	3,12	0,54
ine	21541	2	0,41	10,2	1,06	6,71	0,85
sin	21513	2	0,41	11,18	1,11	7,83	0,9
kla	21481	2	0,41	7	0,85	9	0,85
n k	21432	2,73	0,5	5,04	0,73	3,12	0,54
ri	21139	2	0,41	6,85	0,85	7,04	0,89
n a	21045	2	0,41	5,04	0,73	3,35	0,57
alı	21036	3,65	0,58	10,59	1,08	10	0,97
baş	20977	3,73	0,62	11,61	1,12	13,61	1,19
iği	20576	3,46	0,59	2,24	0,44	6	0,82
ek	20429	3,73	0,62	6	0,8	7,59	0,88

Table 9 The most used character sequences of length four

		40,95		F (optimistic)		QWERTY (optimistic)	
Word	Frequency	Distance (diameter)	Time (sec)	Distance (diameter)	Time (sec)	Distance (diameter)	Time (sec)
bir	60949	3,73	0,7	12,3	1,32	7,95	1,09
bir	50404	3	0,61	14,23	1,52	9,87	1,28
ları	48837	4	0,73	10,69	1,32	18,26	1,66
leri	43967	4	0,73	14,25	1,53	11,58	1,27
ve	34196	4	0,73	4,24	0,77	6,99	1,03
erin	31800	4	0,73	13,84	1,51	7,24	1,02
arın	30574	4	0,73	9,69	1,27	14,87	1,52
ında	30292	4,65	0,79	7,8	1,1	10,25	1,28
inde	24326	3	0,61	10,28	1,26	6,99	1,02
nda	23558	4,65	0,79	5,8	0,95	7,88	1,12
nin	22705	3	0,61	14,22	1,52	5,59	0,91
den	21957	3	0,61	10,34	1,3	6,71	0,94
dan	21094	5,65	0,91	8,48	1,17	8,71	1,1
bu	20680	4,73	0,82	11,51	1,25	6,4	0,98
nın	19944	3	0,61	11,04	1,36	10,34	1,24
ini	19684	3	0,61	13,42	1,46	7,51	1,1
baş	19008	5,46	0,92	12,73	1,34	14,73	1,41
ını	18856	3,73	0,7	11,04	1,36	12,83	1,47
iyor	18421	6,29	0,97	15,59	1,53	10	1,26
ola	18235	4	0,73	12,65	1,45	12,16	1,28
nlar	18111	3,73	0,7	9,5	1,22	13,39	1,42
ine	17731	3	0,61	12,2	1,39	9,75	1,26
eri	17707	4	0,73	10,49	1,31	8,04	1,09
arı	17300	4,73	0,82	8,73	1,2	13,86	1,47
nde	17263	3	0,61	6,69	1,03	7,8	1,08
içi	16755	6,29	0,97	9,52	1,25	8,7	1,2
anla	16466	4,73	0,82	10,44	1,28	16,28	1,59
lar	16222	3	0,61	10,73	1,32	13,73	1,44
için	15949	6,29	0,97	12,87	1,46	7,89	1,14
asın	15500	4,65	0,79	12,11	1,43	14,61	1,39
yap	15418	5,73	0,94	14,42	1,52	16,21	1,58
kar	15286	3,73	0,7	7,69	1,12	11,69	1,32
da	14625	5,38	0,88	6,84	1,04	6,24	0,99
iler	14541	5	0,85	16,64	1,64	9,38	1,1
ıyor	14511	6,65	0,99	14,17	1,47	13,59	1,49
de	14465	3,73	0,7	7,73	1,12	6,16	0,96
gör	14082	7,25	1,03	10,06	1,28	11,03	1,32
alar	13852	3	0,61	12,69	1,44	18,69	1,68
n bi	13599	3,73	0,7	10,74	1,23	5,06	0,84
ler	13500	4	0,73	12,68	1,44	10,62	1,21
klar	13480	3	0,61	9,69	1,24	11,69	1,24
ına	13387	4	0,73	9,64	1,25	12,44	1,41
ması	13130	4,65	0,79	12,11	1,43	16,58	1,48
eler	13110	4	0,73	15,64	1,6	14,15	1,4
inin	13107	3	0,61	16,77	1,67	6,71	1,04
endi	13009	4	0,73	10,25	1,28	12,72	1,46
gel	12932	5,65	0,91	11,16	1,31	11,27	1,3
son	12899	8,65	1,18	7,72	1,08	11,4	1,31

lara	12645	3	0,61	10,39	1,3	13,39	1,42
ndan	12205	5,65	0,91	7,44	1,08	11,23	1,33
öyle	12023	5,46	0,85	14,57	1,41	13,82	1,5
esin	11981	4,65	0,79	15,79	1,62	8,94	1,12
ının	11809	3	0,61	12	1,42	13,83	1,52
bil	11521	4,73	0,82	14,69	1,43	5,75	0,92
yle	11519	3	0,61	9	1,1	13,26	1,46
ası	11456	5,38	0,88	11,15	1,36	13,61	1,33
lan	11270	4	0,73	11,68	1,39	14,71	1,42
rini	11269	3	0,61	15,79	1,62	8,47	1,17
dığı	11137	3,73	0,7	4	0,73	10,24	1,09
kon	10949	5,38	0,88	7,69	1,12	5,95	0,94
oldu	10944	6,61	0,98	12,86	1,47	10,76	1,25
arak	10926	3	0,61	8,39	1,18	12,39	1,38
rını	10913	4	0,73	11	1,36	16,79	1,65
rak	10768	3,73	0,7	7,69	1,12	11,69	1,32
ara	10608	3	0,61	7,39	1,09	7,62	1,12
değ	10526	5,73	0,91	6,85	1,02	11,12	1,19
tür	10428	8,65	1,18	4,8	0,83	18,04	1,67
ver	10377	5	0,85	5,88	0,89	4,95	0,82
old	10320	5	0,85	11,29	1,37	10,16	1,21
ger	10305	6,65	1,03	8,8	1,19	5,69	0,91
çin	10186	4,65	0,79	12,27	1,42	6,18	0,96
diği	10146	5,46	0,92	5,88	0,89	10,61	1,32
duğu	10092	9,61	1,27	8,22	1,11	11,64	1,4
dı.	10033	1,73	0,3	2	0,32	8	0,65
türk	10021	8,65	1,18	3,92	0,73	19,61	1,77
rın	9934	4	0,73	10,04	1,29	13,29	1,36
n ya	9916	4,73	0,82	11,87	1,38	8,77	1,14
ığı	9879	3,73	0,7	5,04	0,82	5,84	0,89
olm	9871	5	0,85	8,65	1,12	5,96	0,94
e ka	9782	3,73	0,7	7	1,06	12,04	1,35
i bi	9734	3,73	0,7	9,93	1,16	6,99	1,03
a ka	9722	3,73	0,7	7	1,06	11,24	1,28
ist	9687	5,46	0,85	9,63	1,2	11,32	1,35
ile	9604	4	0,73	15	1,51	11,42	1,31
esi	9554	4,65	0,79	12,44	1,41	9,75	1,19
kler	9529	4	0,73	11,64	1,35	8,58	1
an b	9524	4,73	0,82	7,8	1,08	7,83	1
acak	9487	7,93	1,09	5,24	0,85	12,39	1,38
aya	9465	5	0,85	14	1,47	11,46	1,36
erek	9451	6	0,97	11,28	1,38	7,59	0,96
anın	9312	4	0,73	11,64	1,4	14,81	1,57
e ya	9225	4,73	0,82	10,83	1,29	10,69	1,33

Table 10 The most used character sequences of length five

		40,95		F (optimistic)		QWERTY (optimistic)	
Word	Frequency	Distance (diameter)	Time (sec)	Distance (diameter)	Time (sec)	Distance (diameter)	Time (sec)
bir	47677	4,73	0,91	15,35	1,74	10,99	1,5
ların	27283	5	0,94	14,69	1,79	22,87	2,17
lerin	24124	5	0,94	19,84	2,09	13,81	1,62
ında	15942	5,65	0,99	9,8	1,42	12,49	1,63
için	15705	7,29	1,17	15,11	1,8	10,93	1,55
ları	14872	5,73	1,03	13,73	1,73	21,86	2,11
inde	13632	4	0,82	12,28	1,59	10,04	1,43
leri	13458	5	0,94	16,49	1,88	14,62	1,69
inin	12230	4	0,82	19,81	2,08	7,83	1,26
ndan	11101	6,65	1,11	10,48	1,49	12,35	1,55
n bir	11095	4,73	0,91	15,35	1,74	9,06	1,31
ının	11015	4	0,82	15,04	1,83	14,95	1,74
arını	10720	5	0,94	13,69	1,74	19,48	2,03
oldu	10318	8,61	1,3	15,9	1,88	13,8	1,66
erini	10104	5	0,94	19,43	2,07	9,47	1,37
için	10009	7,29	1,17	15,91	1,87	9,01	1,36
arın	9766	5	0,94	12,73	1,68	15,99	1,75
kları	9453	5	0,94	12,69	1,64	19,26	1,87
nları	8951	5,73	1,03	12,5	1,62	20,95	2,05
arak	8929	4,73	0,91	10,39	1,5	14,39	1,7
ından	8729	6,65	1,11	11,44	1,55	15,84	1,84
konu	8724	8,02	1,26	14,27	1,71	7,95	1,26
olma	8708	6	1,06	12,65	1,6	12,54	1,53
türk	8574	10,65	1,5	5,92	1,05	22,65	2,18
anlar	8558	5,73	1,03	13,14	1,67	18,98	1,97
sında	8548	5,65	0,99	12,27	1,6	19,25	1,96
deği	8443	7,46	1,2	7,97	1,24	14,12	1,6
yordu	8153	9,9	1,42	13,2	1,59	12,76	1,61
aları	7820	5	0,94	15,69	1,85	26,26	2,31
gibi	7700	7,29	1,17	17,63	1,85	10,35	1,5
erin	7666	5	0,94	16,88	1,93	8,35	1,24
lduğu	7592	11,61	1,59	11,86	1,57	17,64	1,97
çok	7533	8,98	1,33	8,05	1,27	7,16	1,19
kendi	7462	6	1,06	14,25	1,76	18,31	2,02
nden	7366	4	0,82	12,34	1,63	10,35	1,39
ileri	7256	6	1,06	21,25	2,14	13,38	1,58
kend	7240	5,73	1,03	12,61	1,63	15,7	1,83
eleri	7237	5	0,94	20,25	2,11	18,15	1,87
i bir	7193	4,73	0,91	14,54	1,67	10,99	1,5
rinde	7178	4	0,82	14,89	1,77	10,99	1,49
olduğ	7100	9,61	1,39	14,66	1,77	14,76	1,72
sonra	7005	9,38	1,35	10,3	1,45	15,7	1,83
başka	6979	6,73	1,15	18,61	2	22,61	2,12
a bir	6932	4,73	0,91	14,3	1,65	10,18	1,44
e bir	6809	4,73	0,91	14,3	1,65	10,99	1,5
sonr	6799	10,38	1,47	8,72	1,28	15,01	1,76
larak	6777	4	0,82	13,39	1,7	20,39	2,03
daha	6770	7,84	1,27	14,06	1,73	14	1,7

erind	6742	5	0,94	15,84	1,84	10,88	1,48
daha	6724	7,11	1,18	13,02	1,64	14,24	1,72
anla	6701	5,73	1,03	12,44	1,61	18,52	1,93
yapı	6534	10,09	1,44	20,42	2,1	18,01	1,89
larda	6497	7,29	1,17	10,5	1,42	13,81	1,58
kleri	6415	5	0,94	16,25	1,86	12,58	1,48
asınd	6408	5,65	0,99	14,11	1,75	18,25	1,84
başk	6400	7,46	1,24	16,73	1,81	16,73	1,73
ile	6349	5	0,94	17,24	1,85	14,46	1,72
olara	6201	4	0,82	14,99	1,8	14,5	1,64
daki	6172	7,29	1,17	12,04	1,58	13,16	1,57
türki	6170	11,29	1,56	8,92	1,26	20,73	1,99
görü	6162	9,9	1,42	11,87	1,58	19,03	1,97
rkiye	6152	8,29	1,29	21,12	2,01	10,73	1,46
ürkiy	6149	9,94	1,47	15,92	1,7	15,73	1,7
rini	6066	4	0,82	18,03	1,96	11,51	1,58
erine	6058	5	0,94	18,45	2,02	11,71	1,52
rine	6013	4	0,82	16,81	1,89	13,75	1,73
olar	5933	5	0,94	15,34	1,83	14,85	1,66
olan	5881	6	1,06	16,29	1,9	17,75	1,83
inden	5873	4	0,82	14,89	1,77	11,47	1,52
söyl	5862	9,11	1,35	14,68	1,58	15,82	1,82
gibi	5857	5,65	0,99	16,51	1,76	11,39	1,59
masıı	5840	6,38	1,08	15,15	1,84	20,18	1,93
rını	5808	5,73	1,03	14,04	1,77	20,39	2,1
kara	5752	4,73	0,91	10,39	1,5	14,39	1,7
karşı	5727	8	1,21	17,87	1,93	17,27	1,8
yorum	5707	5,65	0,99	20,18	2,05	13,24	1,69
karş	5705	5,73	1,03	12,3	1,62	18,27	1,92
lerde	5654	6,65	1,11	13,33	1,61	9,81	1,24
iyle	5648	5,65	0,99	17	1,75	15,26	1,78
a da	5580	6,38	1,08	8,84	1,36	8,47	1,34
başl	5471	8,11	1,3	14,73	1,66	15,73	1,61
söyle	5440	8,11	1,23	19,57	1,94	20,4	2,09
anın	5434	5	0,94	14,68	1,81	15,93	1,79
ığını	5420	4	0,82	10	1,36	11,46	1,46
ğını	5385	4,73	0,91	12,04	1,56	13,94	1,69
diye	5382	7,38	1,2	21,68	2,12	11,61	1,56
arınd	5334	5	0,94	11,69	1,59	18,51	1,98
rında	5301	6,65	1,11	10,8	1,51	17,82	1,92
iste	5292	8,11	1,23	12,63	1,61	13,32	1,67
büyü	5282	9,67	1,44	14,76	1,73	19,44	1,95
deki	5266	6,65	1,11	13,93	1,73	10,87	1,41
kada	5257	8,02	1,26	8,61	1,34	13	1,58
masın	5246	5,65	0,99	16,11	1,9	21,19	1,99