5. Uygulama

Bu uygulama Türkçe metinlerdeki harflerin tekli, ikili ve üçlü ardarda gelmesi olasılıklarını ortaya koymak amacıyla internet üzerinden Türk Dil Kurumu, Türkoloji…vb Türk diline kaynak olacak sitelerden yararlanarak MS Office Access veritabanı ile MS Visual Studio ortamında C#.NET platformu ile apriori algoritması ile yapılmıştır.

Kelimelerdeki harf analizlerinin arama motorlarında kullanılan yöntemleri iyileştirmek için kullanılması öngörülebilir. Kullanma alanları Tıp, Mühendislik ve Sosyal alanlardır…

Bu tezde aşağıdaki sorulara cevap bulunmaya çalışılmıştır.

* Türkçe metinlerde hangi harf ne sıklıkla kullanılmıştır?
* Türkçe metinlerde hangi harfler ne sıklıkla yan yana gelmiştir?
* Türkçe metinlerde 3 harf ne sıklıkla yan yana gelmiştir?

Bu çalışmamız Türkçe metinlerdeki harfler hakkında istatistikî bilgi sağlayacaktır. Yazılan metinlerin türlerine göre, bilimlerine göre veya kullanıldıkları yerlere göre değişik sonuçlara ulaşılabilir. Böylece tıp, mühendislik ve sosyal alanlardaki çalışmalara hem teorik hem de yönetsel anlamda önemli katkılar sağlayacaktır.

**Uygulamanın Çalışması**

Girilen metin tek harf, iki harf ve üç harf üzerinden olasılıkları hesaplanarak uygulanır.

Metin üzerinde gerekmeyen ifadeler ayıklanır.

Boşluk karakterleri ayıklama işlemidir.

text = text.Trim();

Türkçe karakter için yer değiştirme(replace) fonksiyonu kullanarak büyük harf(upper) yapıyoruz.

text = text.Replace("ı", "I");

text = text.Replace("i", "İ");

text = text.Replace("ğ", "Ğ");

text = text.Replace("ü", "Ü");

text = text.Replace("ö", "Ö");

text = text.Replace("ş", "Ş");

Büyük harflere çeviriyoruz.

text = text.ToUpper();

Yabancı

string gecerli = "ABCÇDEFGĞHIİJKLMNOÖPQRSŞTUÜVWXYZ0123456789";

for (int i = 0; i < text.Length; i++)

{

if (gecerli.IndexOf(text[i]) == -1)

{

text = text.Replace(text[i].ToString(), " ");

}

}

Her üç olasılık hesaplama için öncelikle veritabanındaki olasılıklar sıfırlanır.

**Tek Harfler**

Girilen metin sonuna kadar döngüde kalır. Sıradaki her harf için Microsoft Access veritabanında TEKLI tablosunda olasılık kolonu bir arttırılarak kaydedilir. Fonksiyon:

private void tekliHarf(string girdi)

{

Fonksiyon Girdiğimiz metin “ab” olsun. “a”yı ve “b”yi işleyeceğiz. Döngü sayısı uzunluk (Lenght) oldu.

…

Veritabanı kodları…

Veritabanı komutu için “cmdText” değişkenini kullanıyoruz.

string cmdText = "";

Olasılık için “ihtimal” değişkenini kullanıyoruz.

int ihtimal = 0;

,

Döngü buradan başlıyor.

for (int i = 0; i < girdi.Length; i++)

{

…

Tekli tablosundan döngü sırasındaki harfin olasılığını sorgulayıp çalıştırıyoruz.

cmdText = "select OLASILIK from TEKLI WHERE HARF = '" + girdi[i] + "'";

…

Sorgu sonucu dönen değeri tamsayıya çevirip ihtimal değişkenine atıyoruz. Değişkeni bir arttırıyoruz.

ihtimal = Int32.Parse(dt.Rows[0][0].ToString());

ihtimal++;

…

cmdText = "UPDATE TEKLI SET OLASILIK = " + ihtimal + " WHERE HARF='" + girdi[i] + "'";

…

Olasılığı bir arttırılan harf için Tekli tablosunda güncelleme komutu yazılıp çalıştırıyor.

**İkili Harfler**

Girilen metin sonuna kadar döngüde kalır. Sıradaki her ikili harf için Microsoft Access veritabanında IKILI tablosunda olasılık kolonu bir arttırılarak kaydedilir. Fonksiyon:

private void ikiliHarf(string girdi)

{

Fonksiyon Girdiğimiz metin “abc” olsun. “ab”yı ve “bc”yi işleyeceğiz. Döngü sayısı uzunluğun bir eksiği (Lenght-1) olur.

…

Veritabanı kodları…

Veritabanı komutu için “cmdText” değişkenini kullanıyoruz.

string cmdText = "";

Olasılık için “ihtimal” değişkenini kullanıyoruz.

int ihtimal = 0;

,

Döngü buradan başlıyor.

for (int i = 0; i < girdi.Length-1; i++)

{

…

İkili tablosundan döngü sırasındaki ikili harfin olasılığını sorgulayıp çalıştırıyoruz.

cmdText = "select OLASILIK from IKILI WHERE HARF1 = '" + girdi[i] + "' AND HARF2 = '" + girdi[i + 1] + "'";

…

Sorgu sonucu dönen değeri tamsayıya çevirip (Int32.Parse) ihtimal değişkenine atıyoruz. Değişkeni bir arttırıyoruz.

ihtimal = Int32.Parse(dt.Rows[0][0].ToString());

ihtimal++;

…

cmdText = "UPDATE IKILI SET OLASILIK = "+ ihtimal +" WHERE HARF1='" + girdi[i] + "' AND HARF2='" + girdi[i + 1] + "'";

…

Olasılığı bir arttırılan harf için İkili tablosunda güncelleme komutu yazılıp çalıştırıyor.

**Üçlü Harfler**

Girilen metin sonuna kadar döngüde kalır. Sıradaki her harf için Microsoft Access veritabanında UCLU tablosunda olasılık kolonu bir arttırılarak kaydedilir. Fonksiyon:

private void ucluHarf(string girdi)

{

Fonksiyon Girdiğimiz metin “abcd” olsun. “abc”yı ve “bcd”yi işleyeceğiz. Döngü sayısı uzunluğun iki eksiği (Lenght-2) olur.

…

Veritabanı kodları…

Veritabanı komutu için “cmdText” değişkenini kullanıyoruz.

string cmdText = "";

Olasılık için “ihtimal” değişkenini kullanıyoruz.

int ihtimal = 0;

,

Döngü buradan başlıyor.

for (int i = 0; i < girdi.Length-2; i++)

{

…

Üçlü tablosundan döngü sırasındaki üç harfin olasılığını sorgulayıp çalıştırıyoruz.

cmdText = "select OLASILIK from UCLU WHERE HARF1 = '" + girdi[i] + "' AND HARF2 = '" + girdi[i + 1] + "' AND HARF3 = '" + girdi[i + 2] + "'";

…

Sorgu sonucu dönen değeri tamsayıya çevirip ihtimal değişkenine atıyoruz. Değişkeni bir arttırıyoruz.

ihtimal = Int32.Parse(dt.Rows[0][0].ToString());

ihtimal++;

…

cmdText = "UPDATE IKILI SET OLASILIK = "+ ihtimal +" WHERE HARF1='" + girdi[i] + "' AND HARF2='" + girdi[i + 1] + "'";

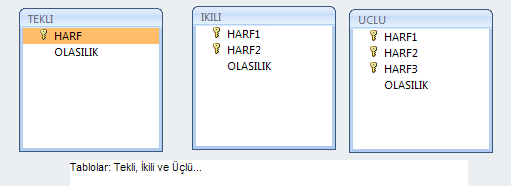
…

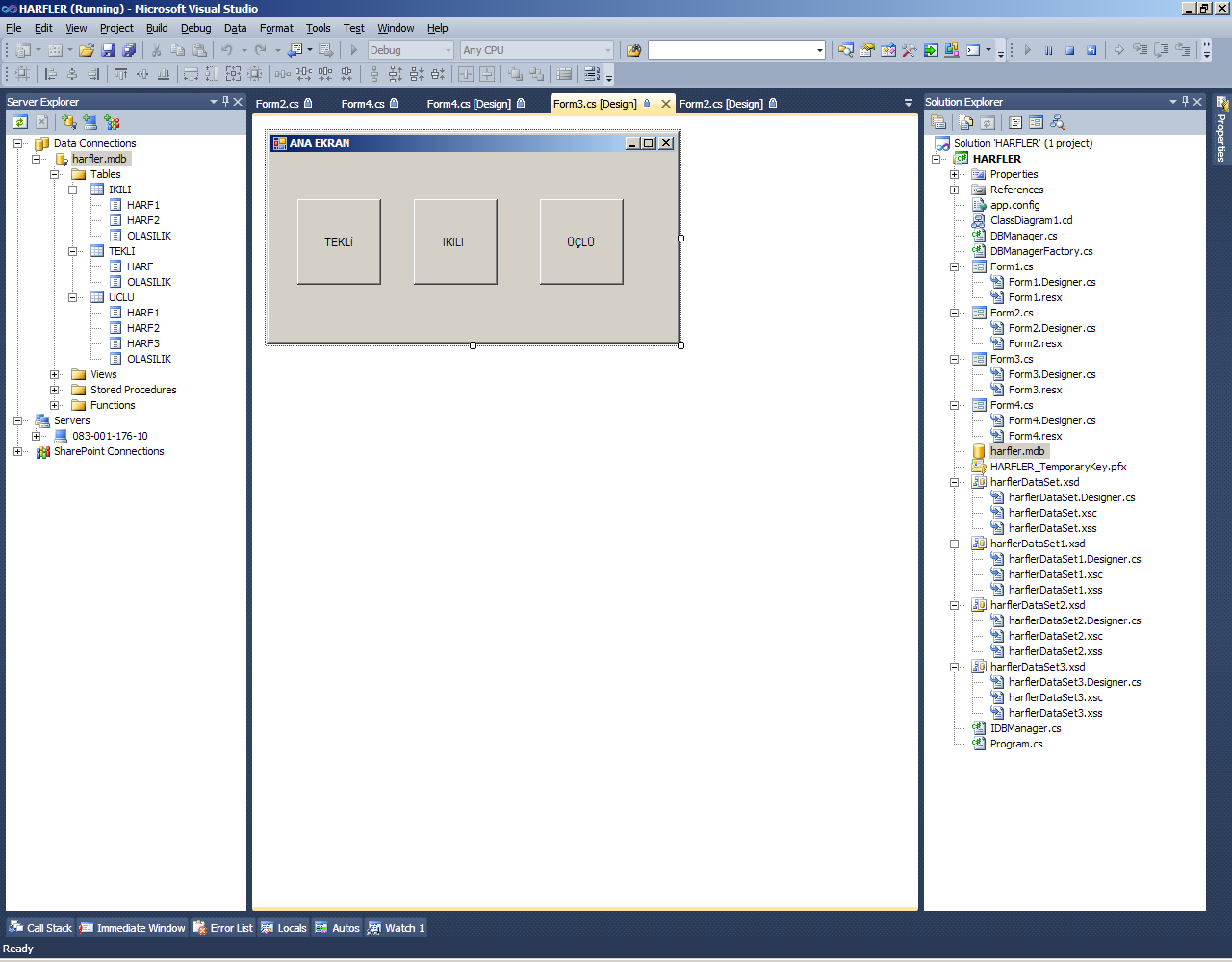
Olasılığı bir arttırılan harf için Üçlü tablosunda güncelleme komutu yazılıp çalıştırıyor.

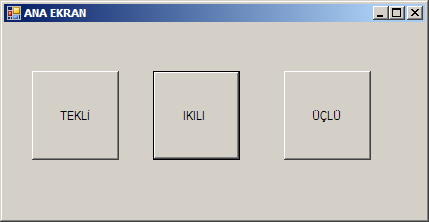
Program Windows Masaüstü (Desktop) uygulamasıdır. C# dilinde yazılmıştır. MS Office Access veritabanını kullanır. Microsoft Visual Studio uygulama geliştirme ortamı (IDE) kullanılmıştır.

Okuttuğumuz dosya içeriği : “aaabbbcccdddeeeffgğhıiijklmnoööpqrsşştuüüvvwwxxvz02132546549879,,,...\*?('/^+!'^+213çççy”.

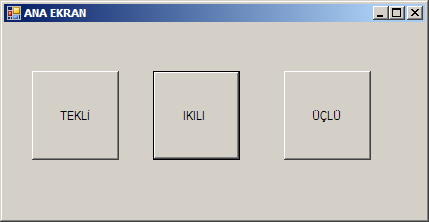
Tablolar: Tekli, İkili ve Üçlü...

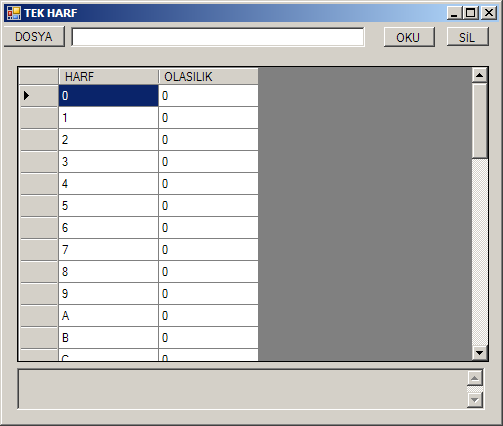




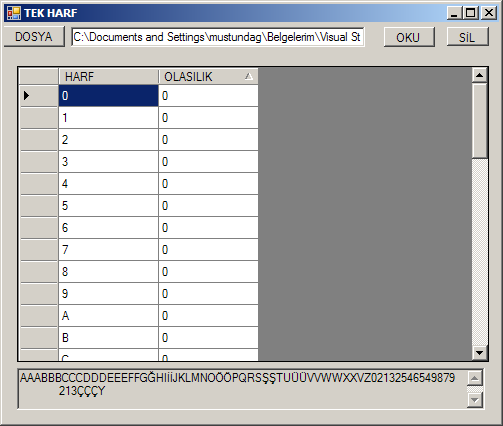


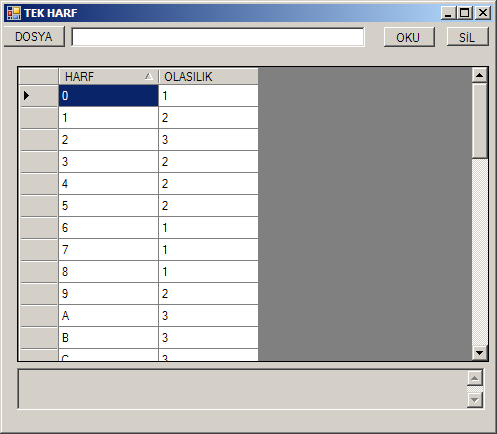
Algoritma seçimi için ana ekrandır.



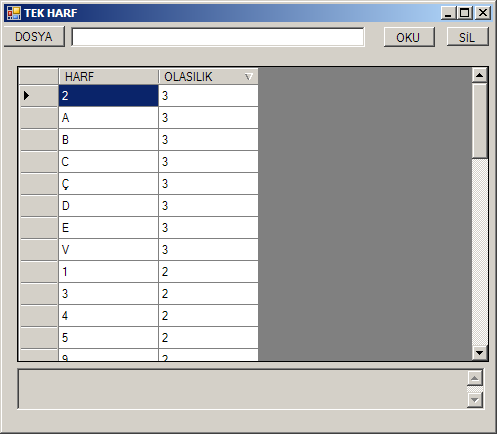


Tek harf için dosya seçilir “OKU” düğmesine basılır. Ekran kapatılır tekrar açılır.

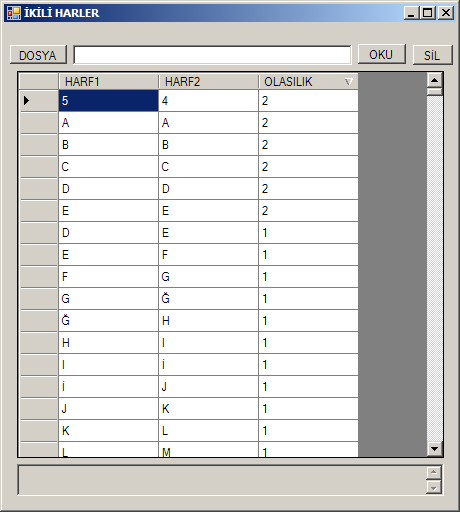




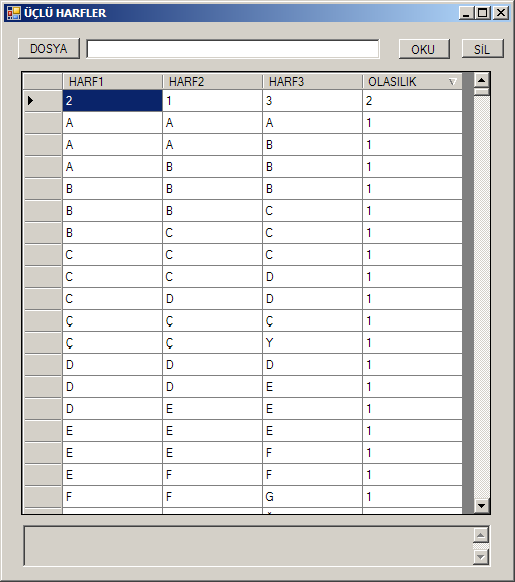
Olasılıklar ve harfler büyükten küçüğe sıralanabilir.



İkili ve üçlü harf için aynı işlemler uygulanmaktadır. Tabloları silmek için “SİL” düğmesi kullanılır.



İkili,



Üçlü.