

KODLAMA SİSTEMLERİ ve VERİLERİN BİLGİSAYARDA TEMSİLİ

KODLAMA SİSTEMLERİNİN TANIMI :

Kodlama, iki küme elemanları arasında karşılıklı kesin olarak belirtilen kurallar bütünüdür diye tanımlanabilir. Diğer bir deyişle, görünebilen, okunabilen yazı, sayı ve işaretlerin değiştirilmesi işlemine 'kodlama' denir. Başka bir bakış açısı ile, sonlu elemana sahip bir kümenin her bir elemanına bir kod verilmesi, kodlama olarak tanımlanır.

Bu bölümde sayısal ve sayısal olmayan kodlama sistemleri ele alınacaktır.

Kodlama Sistemlerinin Avantajları :

1. Aritmetik işlemlerde kolaylık sağlar.
2. Hataların bulunmasını kolaylaştırır.
3. Hataların düzeltilmesi işlemlerini basitleştirir.
4. Bellek işlemlerinde verimliliği artırır.
5. Bilgilerin işlenmesi işleminin insanlarca kolayca anlaşılmasını sağlar.

Tanım:

Sayısal karakterlerin kodlanmasıyla ortaya çıkan kodlama sistemlerine **Sayısal Kodlar** denir

Alfabetik karakterlerin sayısal karakterlerin ve işaretlerin kodlanmasıyla ortaya çıkan **sistemlere alfasayısal** kodlar denir.

SAYISAL KODLAR

- BCD kodu
- Gray kodu
- +3 kodu
- Aiken kodu
- 5'te 2 kodu
- Bar kodu
- Eşlik Kodu

Not: Onlu bir sayının ikili sayı sistemindeki karşılığının yazılması ile oluşan kodlama sistemi, **yalın ikili kodlama** (pure binary coding) olarak isimlendirilir.

BCD Kodlama Sistemi:

Onluk sistemdeki bir sayının, her bir basamağının ikilik sayı sistemindeki karşılığının dört bit şeklinde yazılması ile ortaya çıkan kodlama yöntemine, **‘İkili Kodlanmış Onlu Sayı Kodu - BCD kodu’** (Binary Coded Decimal Code) ismi verilir. Onluk sayı sistemi 0 ile 9 arasındaki sayıları içerdiğinden, her basamaktaki sayının ikili sistemde kodlanması için 4 bite ihtiyaç vardır. Onlu bir sayıyı BCD kodlu olarak yazmak için, onlu sayının her bir basamağı 4 bitlik ikili sayı grupları şeklinde yazılır. Yazılan gruplar bir araya getirilince BCD kodlu sayı elde edilir.

Örnek 1: $(125)_{10}$ sayısını BCD kodu nedir?

Her bir basamaktaki sayının ikili karşılığı 4 bit olarak yazılırsa;

1 2 5

0001 0010 0101 sayıları bulunur.

Sayıların birleştirilmesiyle;

$(125)_{10} = (000100100101)_{\text{BCD}}$

eşitliği elde edilir.

BCD Kodlama Sistemi:

Onluk sistemdeki bir sayının, her bir basamağının ikilik sayı sistemindeki karşılığının dört bit şeklinde yazılması ile ortaya çıkan kodlama yöntemine, **‘İkili Kodlanmış Onlu Sayı Kodu - BCD kodu’** (Binary Coded Decimal Code) ismi verilir. Onluk sayı sistemi 0 ile 9 arasındaki sayıları içerdiğinden, her basamaktaki sayının ikili sistemde kodlanması için 4 bite ihtiyaç vardır. Onlu bir sayıyı BCD kodlu olarak yazmak için, onlu sayının her bir basamağı 4 bitlik ikili sayı grupları şeklinde yazılır. Yazılan gruplar bir araya getirilince BCD kodlu sayı elde edilir.

Örnek 1: $(125)_{10}$ sayısını BCD kodu nedir?

Her bir basamaktaki sayının ikili karşılığı 4 bit olarak yazılırsa;

1 2 5

0001 0010 0101 sayıları bulunur.

Sayıların birleştirilmesiyle;

$(125)_{10} = (000100100101)_{\text{BCD}}$

eşitliği elde edilir.

BCD Kodlama Sistemi:

Örnek 2: $(\underline{1001} \ \underline{0011} \ \underline{0110})_{\text{BCD}}$ sayısını onlu sisteme çevirelim.

Sayı dörderli gruplara ayrılarak her bir gruptaki ikili sayıların onlu karşılığı yazılırsa;

$$\begin{array}{ccc} (1001 \ 0011 \ 0110)_{\text{BCD}} \\ \downarrow \quad \downarrow \quad \downarrow \\ 9 \quad 3 \quad 6 \end{array}$$

sayıları bulunur. Bulunan sayıların bir arada yazılmasıyla sonuç olarak;

$$(100100110110)_{\text{BCD}} = (936)_{10}$$

sayısı elde edilir.

Gray Kodu:

Gray kodlama yöntemi, basamak ağırlığı olmayan bir kodlama yöntemidir. Basamak ağırlığının olmaması, her bir basamaktaki sayıların basamak ağırlıklarına göre karşılıklarının olmamasıdır.

Gray kodlanmış sayılarda basamak değeri olmadığından, bu kodlama yönteminin aritmetik işlemlerin olduğu yerlerde kullanılması mümkün değildir. Ancak sütun esasına göre çalışan cihazlardaki hatayı azalttığından, giriş / çıkış birimlerinde ve analog - dijital çeviricilerde tercih edilirler.

Gray kodlanmış sayılarla aritmetiksel işlemler yapılmaz. Bu sayılarla aritmetiksel işlem yapılması gerektiğinde bu sayılar önce ikili sayılara dönüştürülür.

Bu bölümde ikili sayıların gray koduna ve gray kodlanmış sayıların ikili sayılara dönüştürülmesi ele alınacaktır.

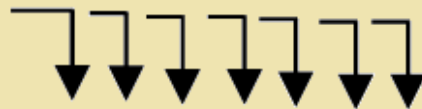
İkili sayıların gray koduna çevrilmesi

İkili sistemdeki bir sayıyı Gray kodlu sayıya dönüştürmek için, en yüksek basamak değerine sahip bitin solunda '0' olduğu kabul edilip, her bit solundaki bit ile toplanarak yazılır. Bu işleme en düşük basamak değerlikli bite kadar devam edilir. Elde edilen sayı Gray Kodu olarak yazılır.

Örnek : $(1000101)_2$ Binary sayısını Gray koduna çevirelim.

0 1 0 0 0 1 0 1

Binary Sayı



1 1 0 0 1 1 1

Gray kodlu sayı

Sonuçta;

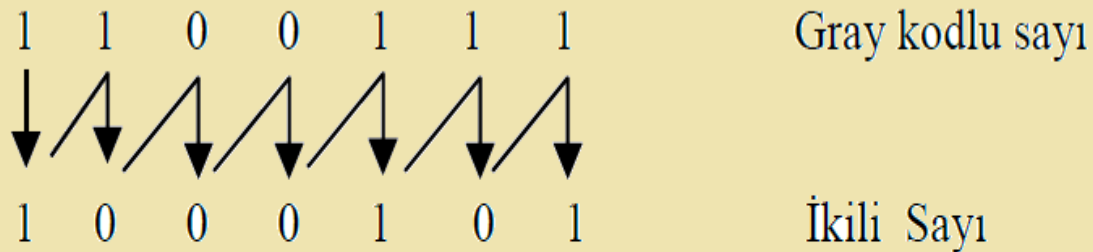
$$(1000101)_2 = (1100111)$$

eşitliği bulunur.

Gray Kodlu sayıların ikili sayılara çevrilmesi

Gray kodlu bir sayıyı ikili sistemdeki sayı şekline dönüştürmek için, en soldaki bit olduğu gibi aşağıya indirilir ve indirilen sayıyla bir sonraki basamakta bulunan sayı toplanarak yazılır. Bulunan sayı ile bir sonraki basamaktaki sayı toplanır ve bu işleme en düşük değerlikli bite kadar devam edilir.

Örnek : $(1100111)_{\text{GRAY}}$ sayısını ikili sayı sistemine çevirelim.



Sonuç olarak;

$$(1100111)_{\text{GRAY}} = (1000101)_2$$

eşitliği bulunur.

+3 Kodu:

Artı 3 kodu (+3 Code), BCD kodu ile ilgilidir ve belirli aritmetik işlemlerde işlem kolaylığı nedeniyle BCD kodu yerine kullanılır.

Bir onlu sayının Artı 3 kodundaki karşılığı, onlu sayının karşılığı olan ikili sayıya 3 eklenmiş halidir. Bu nedenle bu kodlama yöntemi, '3 fazlalık kodu' olarak ta isimlendirilir. ,

Artı 3 kodundaki sayılar, BCD kodunda olduğu gibi dört bitlik ikili sayılar şeklinde ifade edilir.

+3 Kodu:

Örnek: $(59)_{10}$ sayısının +3 koduna çevrilmesi:

$$\begin{array}{r} 5 \quad 9 \\ 3 \quad 3 \\ + \quad + \\ \hline 8 \quad 11 \end{array}$$

1000 1011

$$(59)_{10} = (10001011)_{+3}$$

Örnek: +3 kodu ile kodlanmış $(10100110)_{+3}$ sayısının onlu sistemde karşılığının bulunması:

$$(10100110)_{+3} = (1010 \ 0110)_{+3} = (10 \ 6)_{+3}$$

$$\begin{array}{r} 10 \quad 6 \\ 3 \quad 3 \\ - \quad - \\ \hline 7 \quad 3 \end{array}$$

$$\text{-----} \rightarrow (10100110)_{+3} = (73)_{10}$$

5'de 2 Kodu:

5'de 2 kodunda, her onlu sayı, icinde mutlaka iki tane '1' bulunan 5 bitlik ikili sayı ile temsil edilir. Butun sayılarda mutlaka iki tane '1' bulunduğundan hataların kolayca bulunmasını sağlar. Sayılar ikili sistemde ifade edilirken basamak değerleri '7 4 2 1 0' şeklinde sıralanır.

$(0)_{10}$ sayısını 5'te 2 kodunda ifade etmek için (11000) kombinasyonu kullanılır.

0-9 arasındaki sayıların 5'te 2 kodu aşağıdaki tabloda verilmiştir.

Desimal Sayı	5'te 2 Kodlu Sayı
	7 4 2 1 0
0	1 1 0 0 0
1	0 0 0 1 1
2	0 0 1 0 1
3	0 0 1 1 0
4	0 1 0 0 1

Desimal Sayı	5'te 2 Kodlu Sayı
	7 4 2 1 0
5	0 1 0 1 0
6	0 1 1 0 0
7	1 0 0 0 1
8	1 0 0 1 0
9	1 0 1 0 0

A ilen Kodu:

A iken kodu; 4 basamaklı ve basamak değerlerinin '2421' şeklinde ifade edildiği bir kodlama şeklidir. Onlu sistemde 5'e kadar olan sayıları kodlamak için sağ taraftaki basamaklar kullanılırken, 5'den büyük değerleri ifade etmek için sol taraftaki bitler tercih edilir.

Bu kodlama şekli simetrik kodlamaya bir örnektir. (0-4) arasındaki sayılar için normal ikili sayılar kullanılırken, (5-9) arasındaki sayılar için başlangıçtaki sayıların simetriği kullanılır

Sayı	Aiken Kodu	Sayı	Aiken Kodu
0	0000	5	1011
1	0001	6	1100
2	0010	7	1101
3	0011	8	1110
4	0100	9	1111

Parity (Eşlik) Kod (Hata Düzeltme Kodu)

Bilginin bir yerden başka bölgeye taşınması sırasında, değişik nedenlerden dolayı gürültü oluşması ve oluşan gürültünün iletilen bilgiyi bozması zaman zaman karşılaşılan durumdur. Bilgi iletimi sırasında bu şekilde oluşan hataları tespit etmek ve mümkünse düzeltmek sayısal sistemlerin özelliklerindendir.

Hataları tespit etmede kullanılan en yaygın ve en kolay yöntem eşitlik biti kodlama (parity code) yöntemidir. Bu yöntemde, hataların ortaya çıkarılmasını sağlamak amacıyla BCD kodlu sayının sağındaki veya solundaki basamağa '**eşitlik biti**' (parity bit) eklenir.

Eşitlik biti, kodlanan veride 1 yada 0'ların tek mi, çift mi olduğunu belirtir. İki türlü eşitlik biti yöntemi bulunmaktadır: Çift eşitlik (even parity) ve tek eşitlik (odd parity).

Çift Eşitlik Yöntemi

Çift eşitlik yönteminde; eşitlik bitinin değeri, kodlanacak bilgidaki 1'lerin toplam sayısı (eşitlik biti dahil) çift olacak şekilde seçilir. Kodlanacak sayıdaki 1'lerin sayısı tek ise, eşitlik biti olarak '1' eklenir. Kodlanacak bilgidaki 1'lerin sayısı çift olması durumunda ise, eşitlik biti olarak '0' eklenir.

Örnek: 10011110 sayısı çift eşitlik yöntemine göre **1**10011110 şeklinde kodlanır.

Örnek: 10011100 sayısı çift eşitlik yöntemine göre **0**10011100 şeklinde kodlanır.

Tek Eşitlik Yöntemi

Tek eşitlik bit yöntemi; aynı mantığa göre düzenlenir. Tek fark kodlanan bilgideki 1'lerin sayısı tek olmalıdır.

Örnek: 10011110 sayısı tek eşitlik yöntemine göre **0**10011110 şeklinde kodlanır.

Örnek: 10011100 sayısı tek eşitlik yöntemine göre **1**10011100 şeklinde kodlanır.

Alfa Sayısal Kodlar

Temel Olarak kodlama, iki küme arasında karşılığı tanımlanmış kurallar dizisidir. Tüm karakterler bilgisayarda kullanılırken Bazı Kodlama Sistemlerine göre kodlanmaları gerekmektedir. A'dan Z'ye karakterleri, 0'dan 9'a sayısal karakterleri ve #,& vb. karakterleri bilgisayarlarda temsil etmek için bu kodlama sistemleri kullanılır.

ASCII Kodlama Sistemi : İlk oluşturulan karakter setlerinden biridir ve bu yüzden günümüzde en yaygın olarak kullanılan karakter setidir. İlk hali 7 bit sayılarla kodlanmıştı ve 128 karakterden oluşuyordu.

ASCII tablonun ilk 32 karakteri (tablonun ilk 2 satırı), ekranda görüntülenmeyen kontrol karakterlerini içerir. Mesela; LF(Line Feed) satır besleme, CR(Carriage Return)imleci satırın en sonuna taşıyan satır başı taşıyıcı, Metin tipi veri transferinde kullanılan STX(Start-of-Text) ve ETX(End-of-Text) karakterleri, BS(BackSpace) Ekranda karakterleri geriye doğru silme karakteri bu grubun üyeleridir. 32–127 arası karakterler ise noktalama işaretlerini, harfleri, rakamları ve bazı matematiksel işaretleri içerir.

Alfa Sayısal Kodlar

7 bitlik ASCII dönüşümünden daha sonra, 8 bitlik dönüşüm tablosu tasarlandı. Sisteme yeni 128 karakter daha eklenerek oluşturulan bu yeni sete **uzatılmış (extended) ASCII** karakter seti denildi. Fakat bu eklenen 128 karakter standart olamadı. ASCII'nin değişik sürümleri ortaya çıktı.

Geliştirilen ilk çözüm bu uzatılmış ASCII setlerini, bölgesel olarak gruplamak oldu. IBM tarafından bu yeni uzatılmış ASCII setlerine **kod sayfası (code page)** denilmiştir. Kod sayfalarının ilk 128 karakteri standart ASCII ile aynıdır. Son 128 karakter o bölge veya ülkenin karakterlerini içerir. Mesela kod sayfası 737 Yunan dili karakterlerini, kod sayfası 857 ise Türkçe karakterlerini içeriyordu.

Daha sonra ISO (International Organization for Standardization) devreye girdi ve **ISO Latin-1** karakter setini geliştirdi. Bu karakter setinin ilk 128 karakteri standart ASCII ile aynıdır. Microsoft, Windows işletim sistemi için ISO Latin-1'in biraz değiştirilmiş halini **Windows Kod sayfası 1252** ismiyle kullanmaktadır.

ISO çalışmalarını daha da ileri götürerek daha çok Avrupa dillerinin karakterlerini içeren karakter setleri oluşturdu. Bu karakter setleri **ISO 8859 karakter seti grubu** adı altında toplandı. Bu set 16 gruptan oluşur. Türkçe için **ISO 8859-9** (buradaki 9, grup numarasını temsil ediyor) veya ISO Latin-5 diye bilinen karakter seti atanmıştır.

Alfa Sayısal Kodlar

ASCII karakter seti 2 noktada eksik kaldı. Birincisi, sadece 128 karakterin ifadesini mümkün kılmaktaydı. İkincisi, uluslararası ihtiyacın göz önünde bulundurulmamasıydı. Aynı şekilde ISO 8859 standart grubu da daha ziyade Avrupa dillerine hizmet veriyordu. Pek çok Asya ülkesinde 1 Byte ile ifade edilebilecek 256 farklı karakter yetersiz kalmıştır. ISO 8859 grubuna ait ikinci eksik, farklı ISO dil setini kullanan kişiler arası mesajlaşmanın doğru gösterimi vermemesidir.

Daha sonra Uluslararası ihtiyaçları esas tutan ve dünyada tek bir karakter setinin kullanımını öngören yeni bir karakter seti oluşturulması düşünüldü. Bugün için, 96 binden fazla karakterden oluşan bu yeni karakter setinin adı **Unicode**. Unicode içerisinde, görme özürlülerin kullandıkları Braille alfabesinden, matematiksel sembollere, Uzakdoğu dilleri harflerinden Arap harflerine kadar oldukça geniş bir gösterim sağlanmıştır. Başlangıçta 16 bit kullanılarak 65536 farklı karakterin gösterimi amaçlanmıştı. Fakat daha sonra bu sayı yaklaşık 1.1 milyon karaktere çıkarılmıştır.

Ses Verisi

Ses dalgasının karakteristiği, bit dizileri halinde kodlanabilirse, o sesi bilgisayar ortamında saklayabileceğimizi anlarız. Ses verisi analog formdadır. Bu yüzden her andaki ses bilgisini kaydetmek yerine (ki bu sonsuz büyüklükte yer ihtiyacını doğurur), belli anlardaki ses bilgisini kaydedebiliriz. Analog ses verisini, belli zaman aralıklarına bölüp o andaki ses verisini kaydetme işlemine **örnekleme (sampling)** ve bu işi yapan elektronik devreye de **analog dijital çevirici (Analog to Digital Converter -ADC)** denilir. Örnekleme işleminin ne kadar sık yapıldığı bilgisi **Örnekleme oranı (sampling rate)** ile belirtilir. Eğer örnekleme anları çok aralıklı olursa, ses kalitesi düşük olur.

Örnekleme sırasında, elektriksel sinyal bit dizisine çevrilir. Her örnekleme işlemi için kullanılacak bit dizisi uzunluğu **bit çözünürlüğü (bit resolution)** olarak ifade edilir. Bit dizisi ne kadar büyük olursa o kadar hassas olarak ses verisini ifade etmemiz mümkün olur. Eğer 4 bit kullanırsak bununla 16 değişik durumu ifade edebiliriz. 16 bit ile 65536 değişik seviye elde ederiz. Bit çözünürlüğü sesin kalitesi yanında sesin şiddeti için de bir ölçüttür. Kaydedilen sesin alabileceği ses şiddeti aralığı 8 bit için 48dB, 16 bit için 96 dB'dir.

Resim Verisi

Resim verilerini bilgisayarda tutmak için, resmi oluşturan noktalarına ayırmamız gerekir. Daha sonrada her noktaya (**“Red Gren Blue”-RGB**) **Kırmızı Yeşil Mavi** bileşiminden oluşan bir renk atayabiliriz. Bu noktaları yan yana ve alt alta dizdiğimizde görüntüyü oluşturabiliriz. Resmi oluşturan nokta sayısına (**“resolution”**) **çözünürlük** denir. Resmi oluşturan noktalara da (**“Picture Elements”-Pixel**) **resim ögesi** denir. Bir resim ögesi için ayırdığımız bit dizisinin büyüklüğüne (**“color depth”**) **renk derinliği** denir. 16 bitlik renk derinliği ile (**“Hi Color”**) **yüksek renk** 65536 renkten biri görüntü ögesine atanabilirken 24 bit ile 16,7 milyon (**“True Color”**) **gerçek renk** ten biri atanabilir. Resimleri resim ögesi şeklinde saklamaya (**“raster based”**) **ızgara tabanlı** saklama denilir. Resim bilgisi saklamak için yaygın kullanılan dosya biçimleri jpg , gif , png , ve bmp dir.**Jpeg/Jpg** (**“Joint Photographic Experts Group”** renkli ve gri tonlamalı fotoğraf türlerini kayıplı sıkıştırma denilen yüksek sıkıştırma oranlı saklama yapabilen bir dosya türüdür. **Gif** (**“Graphic Interchange Format”**) **Grafik Değişen Biçim** deki resimler en çok 256 renk içerebilir. Bu biçim hareketli ve arka planı saydam resimleri de destekler. **Png** (**“Portable Network Graphics”**) taşınabilir ağ grafikleri, Gif’e oranla daha yüksek sıkıştırma oranları ve saydamlığın değişik oranlarda olmasını sağlayabilir. **Bmp** (**“Bitmap”**) bit haritası Windows ortamına özgü bir resim dosya biçimidir.

Video Verisi

Video gösterimlerini elde etmek için resim karelerinin saniyede yaklaşık 25-30 tanesini ard arda göstermek, bir bakıma sinema tekniğini kullanmak gerekecektir. Ancak çok sayıda resim kullanılacağı için büyük miktarda disk alanın gerek duyulacak ve verilerin Internet hatlarında taşınabilmesi de ikinci bir sorun oluşturabilecektir. Burada sıkıştırma uygulamamız kaçınılmaz olmaktadır.

Bu tür verilerin sıkıştırılmasına **Codec (“Compressor / Decompressor”)** sıkıştırıcı açıcı denilmektedir. Yaygın kullanılan Codeclerden bazıları MPEG, DivX, Cinepak, XviD, Indeo, Real Video dur. Ayrıca hem video hemde ses verisini sıkıştırabilen Codeclerde vardır. Örneğin **AVI(“Audio Video Interleave”)**, **MOV(“Movie”)**, **RA (“Real Audio”)** , **RAM (“Real Audio Media”)** gösterilebilir.

Veri Sıkıştırma

Sıkıştırma, verileri saklama ve göndermede problem oluşturmamaları için kullanılan yöntemler bütünüdür. Bir bakıma sıkıştırma dosya boyutunu küçültme işlemidir. Sıkıştırma oranı dosyanın ne oranda küçültüldüğünün bir göstergesidir. Sıkıştırma yöntemleri (“Lossy”) kayıplı ve (“lossless”) kayıpsız olmak üzere ikiye ayrılır

Kayıpsız Sıkıştırma: Dosya içerisinde tekrar eden desenlerin daha az bir alan kapsayacak biçimde gösterilmesine dayanır. Bunlardan iki tanesini örnekleyelim: İlki **(“Run Length Encoding”) Çalışma Uzunluk Kodlaması**, tekrar eden dizinlerin değişik bir biçimde yazılması-dır.”ABABABABAB” dizini yerine “AB5” yazılabilir. Bu durumda sıkıştırma %70 olmakta, yüzde yetmişlik bir alan kazancı olmaktadır.

İkincisi ise **(“Huffman Encoding”) Huffman Kodlaması**dır. Bu yöntem veriyi oluşturan karakterler için standart bit genişliği yerine değişebilir bit genişliği kullanmayı temel alır. “AAAAAAAAAABBBBBBBBCCCCCDDDDDEE” dizininde, önce, dizin içindeki karakterlerin kaçar defa yinlendiği bir tablo içine yerleştirilir.[A,10], [B,8], [C,6], [D,5], [E,2] bu tabloya dayanan Huffman ağacı oluşturulur. Özgün dosya ile Huffman ağacı arasında %50 kazanç sağlanmış olur.

Veri Sıkıştırma

Kayıplı Sıkıştırma: (“Lossy Encoding”) Kayıplı Sıkıştırma yönteminde ilke, kodlama yapmak için bazı bilgilerin silinmesidir. Burada resim, ses ve video dosyalarında insanların işitme ve görme organlarının göremeyeceği ayrıntıların silinmesini temel alır. Ses dosyalarında MP3, görüntü dosyalarında JPEG ve video dosyalarında MPEG, DivX kayıplı sıkıştırma dosya türleridir. (Karaca,M,S-2006)