એકમ–13 બહુભાષીય વિષયવસ્તુ વિકાસ (યુનિકોડના ઉપયોગ દ્વારા) Multilingual Content Development (Using Unicode)

: રૂપરેખા :

13.0 ઉદ્દેશો
13.1 પ્રસ્તાવના
13.2 અક્ષર(Character)ની કમ્પ્યૂટરમાં રજૂઆત
13.3 માહિતીની આદાનપ્રદાન માટે અમેરિકન સ્ટાન્ડર્ડ કોડ (ASCII)
13.1.1 ASCIIના પ્રશ્નો
13.4 ભારતીય પરિદશ્ય અને માહિતી આદાનપ્રદાન માટે ઈન્ડિયન સ્ટાન્ડર્ડ કોડ (ISCII)
13.4.1 GIST કાર્ડ
13.4.2 વેબ પ્રયોજિતતા અને GIST
13.4.3 ISCII સામેના અવરોધો
13.5 યુનિકોડ (UNICODE)
13.5.1 યુનિકોડ રચનાના સિદ્ધાંતો
13.5.2 યુનિકોડનું માળખું
13.5.3 યુનિકોડ અને તેનાં લક્ષણો
13.6 યુનિકોડ દ્વારા વેબ ઉપર વિષયવસ્તુ વિકાસ
13.7 યુનિકોડની પ્રયોજિતતાઓ
13.7.1 બહુભાષી સાંકેતીકરણ
13.7.2 બહુભાષી લિપ્યંતર
13.7.3 બહુભાષી કાર્યો
13.8 યુનિકોડના ગ્રંથાલયોમાં પ્રયોજિતતા
13.9 યુનિકોડ સાથે સંકળાયેલા પ્રશ્નો
13.10 સારાંશ
13.11 તમારી પ્રગતિ ચકાસોના ઉત્તરો
13.12 ચાવીરૂપ શબ્દો
13.13 સંદર્ભો અને વિશેષ વાચંન

13.0 ઉદ્દેશો (OBJECTIVES)

અગાઉના એકમમાં તમે વેબ વિષયવસ્તુ વિકાસની ટેક્નોલોજી, સાધનો અને તકનીકોની વિવિધ બાબતો વિશે શીખ્યા છો. આવી વિષયવસ્તુ જુદી જુદી લિપિમાં હોઈ શકે છે. અત્રે જાણવું આવશ્યક છે કે કમ્પ્યૂટર અક્ષરોની રજૂઆત કેવી રીતે કરવામાં આવે છે. અને પ્રમાણિત આંતરરાષ્ટ્રીય કોડ અન્ય લિપિઓ ખાસ કરીને ભારતીય લિપિઓના અક્ષરોને સુવિધા આપી શકે છે. આ એકમમા આપણે લિપિઓ હસ્તગત કરવા ખાસ કરીને ભારતીય લિપિઓ સંદર્ભમાં પ્રાપ્ય પ્રમાણિત સંકેત સમૂહો (Standard code)ની ચર્ચા કરીશું.

- **આ એકમના અધ્યયન બાદ તમે આ બાબતોથી સક્ષમ બનશો.**
- કમ્પ્યૂટરમાં અક્ષરોની રજૂઆતની પદ્ધતિની સમજ પ્રાપ્ત કરશો.

- આતંરરાષ્ટ્રીય પ્રમાણિત સંકેત સમૂહ જેવા કે, આસ્કી-2(ASCII) શીખશો.
- બિનલેટિન અક્ષરોના સાંકેતીકરણની જરૂરિયાત અને પડકારો સમજશો.
- કમ્પ્યૂટરમાં ભારતીય ભાષાઓની રજૂઆત સમજશો.
- કમ્પ્યૂટરમાં ભારતીય ભાષાઓને હસ્તગત કરવા માટેની તકનીકો અને માનાંકો જાણશો.
- યુનિકોડના લક્ષણો સમજશો.
- યુનિકોડની શક્ય પ્રયોજિતાઓ અને મર્યાદાઓ.

13-1 પ્રસ્તાવના(INTRODUCTION)

સાપ્રંત સમયમાં માહિતી પ્રત્યાયન સાથે માહિતી ટેક્નોલૉજી પ્રવેશવા પામી છે તે જરૂરિયાતમાં મદદ અને સાધનો પૂરા પાડે છે. આજે પ્રત્યાયનનું એક સાર્વત્રિક સ્વરૂપ નેટવર્ક ઉપર છે પરંતુ માહિતી માત્ર અંગ્રેજીમાં છે. કારણ કે તાજેતર સુધીમાં જુદી જુદી ભાષાઓની કમ્પ્યૂટરમાં રજુઆત માટેના રજૂઆતલક્ષી માનાંકોની ઉણપ છે. વિચારવસ્તુની લેખિત સ્વરૂપે રજૂઆત કરવા માટે લિપિની જરૂરી છે. વિચારોને સપાટી ઉપર સંકેતો દ્વારા રજૂઆત કરવી એ લિપિકરણ છે. લિપિઓની શોધ અંગે ઘણી વિચારધારાએ છે. ઘણાં માને છે કે તેનો જન્મ મેસોપોટેમિયામાં થયો હતો. થોડાક માને છે કે, તેનો જન્મ અલગ અલગ રીતે સમાંતર હિન્દુ, મેસોપોટેમિયામાં, મેસો અમેરિકા અને ચીનમાં થયો હતો.(Anclentscripts. com) ભારતમાં સંસ્કૃત સમયાંતરે જુદી જુદી લિપિમાં લખવામાં આવતી હતી. બ્રાહ્મી(ઈ.સ પૂર્વે 7મી સદી) અને ખારોશી(ઈ.સ પૂર્વે 5મી સદી દરમ્યાન) પ્રાચીન જાણીતી લિપિઓ છે. જેમાંથી બ્રાહ્મીએ ઘણી ભારતીય લિપિઓ માટે માતૃલિપિ થઈ છે. તે સમય દરમ્યાન ઘણી લિપિઓ જેવી કે ગ્રંથા અને ત્યારબાદ નાગરી અર્થાત્ દેવનાગરી અનુક્રમે 5મી અને 7મી સદીમાં અસ્તિત્વમાં આવી. સમયના પ્રવાહની સાથે સંસ્કૃતિના સ્થળાંતરને કારણે ઘણી ભાષાઓ વિકાસ પામી અને તે પ્રમાણે ભારતીય ઉપખંડમાં જુદી જુદી લિપિઓ બ્રાહ્મીમાંથી વિકાસ પામી.

ઈન્ટરનેટનો માહિતીના ત્વરિતગતિ માર્ગ(Super Highway) તરીકે વિકાસ થતાં તમે લાગે છે કે માહિતીનો પ્રવાહ ભાષાંતરને બદલે સ્રોતની ભાષા લિપિમાં થવો જોઈએ અને બાકીની દુનિયાને પણ તે માહિતી પ્રાપ્ત થવી જોઈએ. લિપિના અક્ષરોની કમ્પ્યૂટરમાં પ્રયોજિતતા એ જટિલ પ્રક્રિયા છે. અને તેના માટે માનાંકો છે. આસ્કી(American Standerd code for Informetion Interchange - ASCII) એ તે પૈકીનો એક માનાંક છે જેની રચના સૉફ્ટવેર માધાતાઓની કમિટી દ્વારા 1962માં કરવામાં આવી અને 1967માં તેની પુનઃરચના કરવામાં આવી. પરતું તે ભાગ્યે જે એક કરતાં વધુ લિપિની રજૂઆતના પ્રશ્નોને સંબોધી(રજૂ કરી) શકે છે. આ રીતે તે અક્ષરોના સાંકેતીકરણની વૈશ્વિક પદ્ધતિ તરીકેતે નિષ્ફળ જાય છે. ASCII એ અક્ષરોની રજૂઆત માટે એક બાઈટ કોડ(One byte) છે. જે ફક્ત 256(0-256) અક્ષરોની વધારેલ વિસ્તારસ્વરૂપમાં રજૂઆત કરી શકે છે. પરંતુ તે ફક્ત 7 બીટ્સ(Bits) નો ઉપયોગ કરે છે. અર્થાત્ 128 અક્ષરોને સુવિધા પ્રદાન કરી શકાય છે. ASCII ના સિદ્ધાંત પ્રમાણે અન્ય દેશોએ તેમની ભાષાની લિપિ માટેના વ્યક્તિગત માનાંકની રચના કરી. તેજ માર્ગદર્શિકા પ્રમાણે ભારતમાં પ્રશ્નોની રજૂઆત માટે બ્યુરો ઑફ ઈન્ડિયન સ્ટાન્ડર્ડ(BIS) એ ISCII ઈસ્કી(Indian Standard code for Information Interchange)ની રચના કરી.

13.2 કમ્પ્યૂટરમાં અક્ષરોની રજૂઆત(CHARACTER REPRESENTATION IN COMPUTER)

કમ્પ્યૂટરમાં દરેક અક્ષરને એક જ નંબરનો દરજ્જો અપાય છે; કારણ કે, કમ્પ્યૂટર ફક્ત વિદ્યુતપ્રવાહ સ્વરૂપમાં જ ફક્ત નંબરને સમજી શકે છે, પરંતુ વાસ્તવિક પ્રક્રિયા તે કરતાં પણ ઘણી જટિલ છે કારણ કે કમ્પ્યૂટર એ વિજાણું પ્રયુક્તિ છે અને તે પણ ફક્ત દબાણ ON અને OFF સમજે છે. વિદ્યુત દબાણ ON અને OFF ની રજૂઆત તરીકે અનુક્રમે 1 અને 0 સંજ્ઞા વિકસાવવા પાછળનું આ એક કારણ છે. સંજ્ઞાની આ વ્યાસ પદ્ધતિને 'દ્વિઅંકી પદ્ધતિ' તરીકે ઓળખાવામાં આવે છે અને આવી પદ્ધતિનો ઉપયોગ કરીને અક્ષરોનું સાંકેતીકરણ કરવાની પદ્ધતિને 'અક્ષર સાંકેતીકરણ' તરીકે ઓળખવામાં આવે છે. ACSII માં જ્યારે 'A' સૂચક ચાવી દબાવવામાં આવે ત્યારે કમ્પ્યૂટર તેને '01000001' તરીકે(અથવા દશાંશ 65) સમજે છે. અને પડદા ઉપર 'A' છાપ છે.

Table 13.1: Represention of Character in Computer

Character on the screen	Binary value used to process it	Character on the screen	Binary value us to process it
1	0110001	A	1000001
2	0110010	B	1000010
3	0110011	C	1000011
4	0110100	D	1000100
5	0110101	E	1000101

◆ **તમારી પ્રગતિ ચકાસો.**

(1) અક્ષરો કમ્પ્યૂટરમાં કેવી રીતે રજૂ કરવામાં આવે છે ? સમજવો.

નોંધ :- (1) નીચે આપેલી જગ્યામાં તમારો ઉત્તર લખો.
(2) એકમને એને આપેલ ઉત્તરો સાથે તમારો ઉત્તર ચકાસો.

....
....
....
....
....
....

13.3 અમેરિકન સ્ટાન્ડર્ડ કોડ ફોર ઈન્ફોરમેશન ઈન્ટરચેન્જ (ASCII) [AMERICAN STANDARD CODE FOR INFORMATION INTERCHANGE (ASCII)]

ASCII 1963(ASCII-1963)

ASCIIની રચના IBM, ATST અને તેની પેટાકંપની Teletype દ્વારા રચાયેલ કાર્યશીલ જૂથ જેને 3.4 તરીકે ઓળખવામાં આવે છે. તેના દ્વારા કરવામાં આવે છે. મૂળભૂત રીતે આ કોર્ડ 8 બીટ અક્ષર સાંકેતીકરણ માટે હતો અને ગોઠવણી તાર્કિક હતી. અર્થાત્ જો કોઈ એક વ્યક્તિ જુદા પાડવા અંગેનો કાર્યક્રમ લખવાની ઈચ્છા રાખે તો તે ASCII નો ઉપયોગ કરી તેનું સંચાલન કરી શકે છે. પરંતુ કોડનું ખાસ આ સંસ્કરણ ફક્ત મોટા અક્ષરો(કેપીટલ) સ્વીકારનું હતું અને વચ્ચે ઘણી ખાલી જગ્યાઓ એવી ધારણા સાથે છોડી દેવામાં આવેલી કે તેનો ઉપયોગ રોમન નાના અક્ષરો માટે કરી શકાશે.(વધુ ઊંડાણપૂર્વક વાચન વિભાગ 3માં)

ASCII - 1967(ASCII-1967)

તે જ કમિટિએ ASCII નું સુધારેલ સંસ્કરણ 1967માં તૈયાર કર્યું. આ સંસ્કરણ રોમન નાના અક્ષરો અને થોડાક વધારાના અક્ષરો જે ASCII 1963માં નહતા જેવા કે {,}, |, ~ વગેરેનો સમાવેશ કરવામાં આવ્યો. ચોક્કસપણે કહી શકાય કે તે 7 બીટ કોડ છે. 8 બીટ કોડ નથી. અર્થાત્ તે 128 અક્ષરોના સ્થાન ધરાવે છે. તે નોંધવું જોઈએ કે 0 થી 9 સુધી સંખ્યાઓને 48 થી 57; કેપીટલ અક્ષરો A-Z ને 65 થી 90 અને નાના અક્ષરો A-Z ને 97 થી 122 નંબર જ્યારે અક્ષરો અને રેખાઓ જુદા પાડવામાં આવે ત્યારે સાચાં પરિણામ પ્રાપ્ત કરવા આપવામાં આવ્યા છે.

વિસ્તારેલ ASCII (લેટીન -+) [Extended ASCII(Latin - +)] અંગત કમ્પ્યૂટરનું નેટવર્ક સામાન્ય બન્યું તે પહેલાં રોમન લિપિ સિવાયની લિપિમાં લખાયેલ ભાષાઓના અક્ષરો અને સોફટવેર માટેના ઉત્પાદકોએ બિન રોમન અક્ષરોને સુવિધા પ્રદાન કરવા સામાન્ય રીતે ASCII ના વિસ્તારેલ 128-256 અક્ષર સ્થાનનો ઉપયોગ કરી બિન પ્રમાણભૂત અક્ષર આલેખન શરૂ કર્યું. બિન પ્રમાણભૂત અક્ષર આલેખનની સાથે સ્વતંત્ર પ્રયોજિતતા અને ત્યારબાદ વિશિષ્ટ અક્ષર સ્વરૂપો અને તે વેબમાં પણ ઉમેરવા માટે શક્ય બનાવ્યું. પરંતુ બિનપ્રમાણભૂત યોજનાઓ સાથે આંતરપ્રચાલનમાં પ્રશ્નો રહેવા પામે છે.(Monotype Imaging Inc)

ઇન્ટરનેટની વધતી જતી લોકપ્રિયતાએ આંતરપ્રચાલન એ મુખ્ય બનવા પામ્યો છે. અક્ષરોના સાંકેતિકરણ માટે, HTML(RFCI 1866) એ ISO-8859-I જે ASCIIનું વિસ્તારેલ સંસ્કરણ જેને લેટિન -1 તરીકે ઓળખવામાં આવે છે. તેના ઉપર આધારિત છે, જે ફક્ત અંગ્રેજી અને યુરોપની ભાષાઓ માટે જ સુયોગ્ય છે. તેમ છતાં ISO - 8859 એ કેટલીક વ્યાપક ઉપયોગ ધરાવતી લિપિઓ માટે અક્ષર

સમૂહોનું પ્રમાણિત કરવા શક્ય બનાવ્યું છે અને તે હવે પશ્ચિમી વર્ણાનુક્રમિક ભાષાઓમાં લખાણ માટે 10 પ્રમાણિત અક્ષર સમૂહ શ્રેણીનો સમાવેશ કરે છે. આ ભાષાઓ :

(1) લેટિન 1(પશ્ચિમી યુરોપ)

(2) લેટિન 2(પૂર્વ યુરોપ)

(3) લેટિન 3(દક્ષિણ યુરોપ)

(4) લેટિન 4(ઉત્તર યુરોપ)

(5) લેટિન - સિરિલિડ

(6) લેટિન - અરબી

(7) લેટિન-ગ્રીક

(8) લેટિન-હિબ્રુ

(9) લેટિન 5(તુર્કી)

(10) લેટિન 6(નોરડીક)

યુનિકોડની સરખામણીમાં 150-8859 તદ્દન મર્યાદિત છે પરંતુ તે ઇન્ટરનેટ ઉપર ઉપયોગી છે. ઉદાહરણ તરીકે MIME સોફ્ટવેર સાથે, અને તેઓ 7 બીટ ASCIIમાં ઘણાં સુધારા છે. અમેરિકન ASCII(US-ASCII) માં અક્ષરો 0 થી 127 કાયમ અભિન્ન-અનન્ય છે.(આરક્ષિત છે) અને ઉપરના સ્થાનો અન્ય લિપિઓના અક્ષરો ધરાવે છે.(MC Mohan 1997)

13.3.1 ASCII સાથેના પ્રશ્નો(Problems with ASCII)

ASCIIએ મહાન શોધ હતી તે નિશંક બાબત છે. ઇન્ટરનેટની શોધ બાદ વિશ્વે વૈશ્વિક ગામ વિશે વાતો શરૂ કરી અને ઇન્ટરનેટ ઉપર અંગ્રેજી લાદી દીધેલ ભાષા તરીકે લાગે તે ઓછું ઈચ્છનીય છે. આ રીતે દરેક વ્યક્તિએ બહુભાષીયતા વાદ વિશે વાતો શરૂ કરી.

જ્યારે આસ્કી 1 બાઈટ અક્ષરોનું સાંકેતીકરણમાં ઉપયોગ કરે છે તો 256 અક્ષરોને હસ્તગત કરે છે અને અર્થ એ થયો કે 128 અક્ષર પછી વધુ 128 અક્ષરોને સ્થાન આપી શકે છે. તેથી તે ભાગ્યે જ એક અથવા વધુમાં વધુ બે ભાષાઓને હસ્તગત કરી શકે છે અને તે બાબત દુનિયાની તમામ ભાષાઓને હસ્તગત કરવાની મર્યાદા દર્શાવે છે. તેથી આ બિંદુએ આસ્કી ચોક્કસ નિષ્ફળ જાય છે અને યુનિકોડ એ આશ્ચર્ય પમાડે તેવો ઉકેલ લાગે છે જે દુનિયાની તમામ ભાષાની લિપિઓને આવરી લે છે.

◆ **તમારી પ્રગતિ ચકાસો.**

(2) અક્ષરોના સાંકેતિકરણ માટે ASCII માનાંકની ચર્ચા કરો.

નોંધ :- (1) નીચે આપેલી જગ્યામાં તમારો ઉત્તર લખો.

(2) એકમને અંતે આપેલ ઉત્તરો સાથે તમારો ઉત્તર ચકાસો.

....

....

....

....

....

....

13.4 - ભારતીય પરિદૃશ્ય અને ઇન્ડિયન સ્ટાન્ડર્ડ કોડ ફોર ઈન્ફોરમેશન ઈન્ટરચેઈન્જ - ASCII [INDIAN SCENARIO AND INDIAN STANDARD CODE FOR INFORMATION INTERCHANGE [ISCII]

બીસ(BIS - Bureau of Indian Standard) માનાંક IS 13194:1991 એ 8 બીટ ધરાવતા કોડ છે. તે બ્રાહ્મીમાંથી વિકાસ પામેલ 10 ભારતીય ભાષાઓને સમાવેશ કરે છે. ASCII(ASCII - Indian Standard Code For Information Interchange) એ ASCIIનું વિસ્તારેલ સ્વરૂપ છે. તે છેલ્લા 128 અક્ષરોના સ્થાન ભારતીય લિપિઓમાં રજૂઆત માટે ઉપયોગ કરે છે. તે ASCIIના જેવા જે

પ્રકારનું તે જ કીબોર્ડ ઉપર તેની સાથે અન્ય 10 ભારતીય ભાષાઓના અભિલેખનો ઉપયોગ કરે છે. તે નિયંત્રણ અક્ષર SO(Shift Out) અને SI(Shift in) નો અનુક્રમે ASCII અને ISCII માટે ઉપયોગ કરે છે. ઈસ્કીમાં અક્ષરોની ગોઠવણી ધ્વનિશાસ્ત્ર અનુસાર છે. એક ભાષામાંથી બીજી ભાષામાં જતાં, પાઠનું લિપિયાંતર કરે છે.(http:// tdil.mit.gov.in) ભારતીય પરિદૃશ્ય સંદર્ભમાં ઈસ્કીએ એક ઉકેલ છે તે નિશંક બાબત છે પરંતુ જ્યારે વૈશ્વિકકક્ષાએ અક્ષરોનું પ્રતિનિધિત્વ કરવા આવે છે ત્યારે ASCII અને ISCII બંને વિસ્તારેલ અક્ષરોની રજૂઆતમાં નિષ્ફળ જાય છે, કારણ કે બંને એક બાઈટનો ઉપયોગ કરે છે.

ISCIIમાં ASCIIના અક્ષરો 8 બીટ કોડના ટેબલના નીચેના અડધા ભાગમાં(0-127) સ્થાન આપેલ છે જ્યારે ભારતીય લિપિઓના અક્ષરો ઉપરના અડધા ભાગ(160-225) માં સ્થાન આપેલ છે. ASCII નીચેની 10 ભારતીય લિપિઓને સેવા આપે છે. દેવનાગરી, ગુજરાતી, પંજાબી, બંગાળી, આસામીઝ(અસમીયા), ઉરિયા, તેલુગુ, તમિલ, મલાયલમ, કન્નડ ઈસ્કી કોડ ટેબલ એ ઉપર દર્શાવેલી લિપિઓ માટે જરૂરી તમામ અક્ષરોનો સુયોગ્ય સમૂહ છે. પ્રથમ સંસ્કરણ 1983માં પ્રકાશિત થયેલ ત્યારબાદ 1986 અને 1988 માં સુધારા થયા બાદ બ્યુરો ઑફ ઈન્ડિયન સ્ટાન્ડર્ડઝ(BIS) એ તેને સ્વીકાર્યો.

13.4.1 ગીસ્ટકાર્ડ(The GIST Card)

ગ્રીસ્ટ(Graphics and Intelligence besed Technology) ની રચના સી.ડેક(Center For Development Of Advanced Computing) દ્વારા મુખ્ય ભારતીય ભાષાઓને સહાય આપવા કરવામાં આવી.ગીસ્ટ ટેક્નોલૉજી ASCII કોડનો ઉપયોગ અક્ષરોની રજૂઆત કરવા ASCIIના વિસ્તારેલ ટેબલનો ઉપયોગથી કરે છે. ગીસ્ટ ઉત્પાદન નીચે દર્શાવેલ માનાંકો જેવા કે ISCII, ISFOC (અક્ષરોના કદની પડદા ઉપર રજૂઆત માટે) અને INSCRIPT(Common Keyboard layout For Indian Scripts)નો ઉપયોગ કરે છે. ભારતીય ભાષાઓ બે વિભાગમાં વિભાજિત કરી શકાય છે. સ્વર અને વ્યંજન તમામ ભારતીય ભાષાઓનો બાહ્ય દેખાવ એક સરખો છે. ગીસ્ટની રચના 8 બીટ સાંકેતીકરણ ઉપર કરવામાં આવી છે. અર્થાત્ તે વિસ્તારેલ ASCII પદ્ધતિ છે તે જુદી જુદી ભારતીય ભાષાઓમાં એક જ સરખા સ્વર અને વ્યંજન માટે એક જ સંકેતનો ઉપયોગ કરે છે. આ રીતે લિપિયાંતર કરવું ઘણું સરળ બને છે. તેમાં ફક્ત ભાષાની રીત બદલવાની જરૂરિયાત રહે છે.(અથવા સ્વીચો - બટનો) ભારતીય ભાષાઓ માટેનું(બહુભાષીય) વર્ડ પ્રોસેશીંગ સૉફ્ટવેર જેવું કે iLeab પણ C-D AC દ્વારા વિકસાવેલ છે તે પણ તે જ સિદ્ધાંત ઉપર કાર્ય કરે છે. અને લિપિયાંતર શક્ય બનાવે છે.

	Hex	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
Hex	Dec	0	16	32	48	64	80	96	112	128	144	160	176	192	208	224	240
0	0	NUL	DLE	SP	0	@	P		p				ओ	ढ			EXT
1	1	SOH	DC1	!	1	A	Q	a	q				औ	ण	ल		०
2	2	STX	DC2	"	2	B	R	b	r				ओ	त	क		१
3	3	ETX	DC3	#	3	C	S	c	s				क	च	क		२
4	4	EOT	DC4	$	4	D	T	d	t			अ	ख	द	व		३
5	5	ENQ	NAK	%	5	E	U	c	u			आ	व	ध	श		४
6	6	ACK	SYN	&	6	F	V	f	v			इ	घ	न	व		५
7	7	BEL	ETB	'	7	G	W	g	w			ई	ऊ	न	स		६
8	8	BS	CAN	(	8	H	X	h	x			उ	थ	प	ह		७
9	9	HT	EM	)	9	1	Y	i	y			ऊ	छ	फ	INV		८
A	10	LF	SUB	=	:	J	Z	j	z			ऋ	अ	थ			९
B	11	VT	ESC	+	;	K	(	k	{			ऐ	श	म			
C	12	FF	FS	.	<	L	\	|	|			ए	ज	म			
D	13	CR	GS	-	=	M	]	m	}			ऐ	ट	थ			
E	14	SO	RS	-	>	N	^	n	~			ए	ठ	य			
F	15	SI	US	/	?	O	_	o	DEL			ओ	ड	र	व		ATR

13.4.2 વેબ પ્રયોજિતતાઓ અને ગીસ્ટ(Web Applicetions and GIST.)

ગીસ્ટ વેબ આધારિત ઉકેલ મેળવવામાં મદદ કરે છે. iPlugin નામે ઓળખાતું સૉફ્ટવૅર. ભારતીય ભાષાઓમાં વેબસાઈટ વિકસાવવામાં સગવડ પૂરી પાડે છે. તે કોઈ પણ ભારતીય ભાષામાં વાતચીત ખંડ(Chat Room) તથા ઈ-મેઈલની રચનામાં સુવિધા પૂરી પાડે છે. કે iplugin સોફ્ટવેરનો સૌથી મોટો લાભ એ છે કે હવે કોઈ પણ વ્યક્તિએ પોતાના સ્થાનિક મશીન(કૉમ્પ્યુટર) ઉપર ભાષાના ફોન્ટ ડાઉનલોડ કરવા પડતા નથી. જ્યારે વેબસાઈટની પ્રાપ્તિ કરવામાં આવે ત્યારે જરૂરિયાતો iPlugin સીધી જ મશીન ઉપર મેળવી આપે છે. જેને ગતિશીલ ફોન્ટ તરીકે પણ ઓળખવામાં આવે છે. ભારતીય ભાષાઓમાં ઘણી વેબસાઈટ ગીસ્ટને ઉપયોગ કરી રહી છે.

13.4.3 ISCIIના અવરોધો(Constrains With ISCII)

ISCIIના ઉપયોગનો મુખ્ય ગેરલાભ એ છે કે એક કરતાં વધુ ભારતીય લિપિઓની એક જ પડદા ઉપર રજૂઆત કરવી અશક્ય છે કારણ કે ISCII એક જ સરખા સ્થળ નામ ઉપયોગ કરે છે એટલે કે 160-255 સ્થાન તમામ ભારતીય ભાષાઓ માટે. ISCIIના ઉપયોગમાં કોઈ પણ વ્યક્તિ અંગ્રેજી ભાષામાં વિષયવસ્તુની સાથે બીજી ભારતીય ભાષા દર્શાવી શકે છે પરંતુ બે કે તેથી વધુ ભારતીય ભાષાઓ એક જ સમયે દર્શાવી શકાતી નથી. ISCIIનો મુખ્ય પ્રશ્ન તેના અક્ષર સાંકેતીકરણનો છે. તે 8 બીટ અક્ષરોનો ઉપયોગ કરે છે. અંગ્રેજીમાંથી કોઈ પણ ભારતીય ભાષામાં રૂપાતંર કરવું હોય તો પણ ઘણા વધુ શ્રમની જરૂરિયાત પડે છે અને અન્ય ભાષાઓમાંથી કોઈ પણ ભારતીય ભાષામાં તે કાર્ય ઘણું મુશ્કેલ છે. બે અથવા વધુ ભારતીય ભાષાઓમાં ભાષાકીય અદલાબદલી(Translingual) પ્રક્રિયા ઈસ્કીના ઉપયોગ દ્વારા શક્ય નથી.

◆ **તમારી પ્રગતિ ચકાસો.**

(3) ભારતીય લિપિમાં અક્ષરોના સાંકેતિકરણ માટે ભારતીય માનાંક શું છે? સમજાવો.

નોંધ :- (1) નીચે આપેલી જગ્યામાં તમારો ઉત્તર લખો.

(2) એકમને અંતે આપેલ ઉત્તરો સાથે તમારો ઉત્તર ચકાસો.

....

....

....

....

....

....

13.5 યુનિકોડ(UNICODE)

યુનિકોડની રચના સૉફ્ટવેર વિકસાવનાર સંગઠનો જેવા કે IBM, માઈક્રોસૉફ્ટ, ઝેરોક્ષ ઓરેકલ વગેરેનું યુનિકોડ સહકારી સંગઠન દ્વારા કરવામાં આવી હતી. તેઓએ 16 બીટ્સ ધરાવતો કોડ તૈયાર કર્યો જે વિશ્વની તમામ લિપિઓનો સમાવેશ કરવા માટે પ્રતિજ્ઞાબદ્ધ હતો. યુનિકોડનું પ્રતિજ્ઞાબદ્ધ લક્ષણ એ છે કે તે 65,536 અક્ષરોનું પ્રતિનિધિત્વ કરે છે. યુનિકોડનું પ્રથમ સંસ્કરણ 1991માં બહાર પડ્યું. ત્યારબાદ બે વધુ સંસ્કરણ બહાર આવ્યા અને સંગઠન સમયનું 3.0 સંસ્કરણ છે.

યુનિકોડ એ ISO I0646 BMP સાથે એકરૂપતા ધરાવે છે, જે લેટિન -1(અને 32 બીટ્સ સંસ્કરણની 2,080,000,000 અક્ષરોની ક્ષમતા છે) 256 અક્ષરોના સાંકેતિકરણન ક્ષમતાને બદલે 16 બીટ્સ ક્ષમતા જે 65536 અક્ષરની ક્ષમતા ઉપર આધારિત છે. BMP બહુભાષીય સમતલ આધારની ભૂમિકા પૂરી પાડે છે. જે કેટલીકવાર શૂન્ય સમતલ આધાર તરીકે ઓળખવામાં આવે છે. જ્યારે કેટલીક લેટિન-1, 1 થી 256 અક્ષરોને મૂલ્ય પ્રદાન કરે છે, યુનિકોડ U^{+}nnnn મૂલ્ય પ્રદાન કરે છે જ્યાં nnnn એ છ દશાંશ અંકન મૂલ્યવાળી સંજ્ઞામાં ચાર અંક સ્થાન ધરાવતી સંખ્યા છે. આ મૂલ્ય કોડ પ્રીન્ટ સંદર્ભ આપે છે. (Jonnson, Dave)

યુનિકોડ 7 બીટ્સ, 8 બીટ્સ અને 16 બીટ્સ માનાંકોના માળખાનો ઉપયોગ કરી ને રજૂ કરી શકાય છે.UTF -7(Universal Charatter Set Transformation Format-7) એ એક માળખું છે જે યુનિકોડ કોડબિંદુનું 7 બીટ મૂલ્યમાં વિભાજીત કરે છે જેને ઈમેલ(જ 7 બીટ્સ આસ્કી સાંકેતીકરણનો ઉપયોગ કરે છે) દ્વારા અને ઇન્ટરનેટ ઉપર મોકલી શકાય છે આ માળખું મુશ્કેલીઓ રજૂ કરે છે કારણ કે કેટલાંક મૂલ્યો ગૂંચવણ ભર્યા છે. પરંતુ હજુ પણ તેનો ઉપયોગ કરી શકાય છે. UTF -8(Universal Character Set Transformation Format-8) માળખું યુનિકોડ મૂલ્યને 8 બીટ શ્રેણીમાં વિભાજિત કરે છે જે વેબ ઉપર સારી રીતે કાર્ય કરે છે. UCS -2(Uinversal Character set-2) દરેક અક્ષરને 16 બીટ્ મૂલ્ય તરીકે સંગ્રહ કરે છે અને દરેક મૂલ્ય યુનિકોડ માનાંકમાં તેના મૂલ્યને અનુરૂપ હોય છે.

યુનિકોડ ભાષાને બદલે લિપિ દ્વારા પાઠનું સાંકેતિકરણ કરે છે. આ બાબત અક્ષરોનું થતું પુનરાવર્તન દૂર રાખે છે. ઉદાહરણ તરીકે લેટિન ‘A’ ઉપયોગ કેટેલન અંગ્રેજી, ઈંડોનેશિયન, સ્વીડીશ અથવા સ્વાહિલીમાં ભેદ પાડ્યા સિવાય પાઠ માટે ઉપયોગ કરે છે, તે જ રીતે દેવનાગરી લિપિ સંસ્કૃત હિન્દી અથવા નેપાળીમાં ઉપયોગ કરી શકાય છે.અક્ષરોના સમૂહનો માનાંક પાઠની અદલાબદલી કરવા સાથે સંબંધ ધરાવે છે. તેઓ ભાષા, રજૂઆતનું માળખું, રંગ અથવા અક્ષરના આકારને સંબંધિત લક્ષણોનો સમાવેશ કરતા નથી. યુનિકોડના અક્ષરોને સ્પષ્ટ રીતે અલગ તારવી શકાય તેવી ચિત્રણ પ્રક્રિયા જે અક્ષરોનું (ડેટા તબદીલી જે સામાન્ય રીતે ખાસ સંકેતોને સ્પષ્ટ કરવા એકમોનો ઉપયોગ થાય છે) બીજા સંકેતોમાં(આપેલ અક્ષરોના ખાસ આકાર જે પ્રદર્શિત થાય છે)(મોટા ટાઈપ ઈમેજ કોર્પોરેશન) આલેખન કરે છે તે દ્વારા દરમ્યાન બનાવી શકાય છે. ઉદાહરણ તરીકે અંગ્રેજીમાં નાનો મૂળાક્ષર ‘C’ નાના મૂળાક્ષરો માટે મોકલી આપે છે. જ્યારે સંકેત પ્રાપ્ત કરવામાં આવે છે ત્યારે તે બીજા સંકેત બીબાંમાં ખાસ આકાર ધરાવતો અક્ષર જે પડદા ઉપર પ્રદર્શિત થશે અથવા પ્રિન્ટર ઉપર મોકલવામાં આવે તેમાં તબદીલ કરે છે.

યુનિકોડને આસ્કી સાથે સરખાવી શકાય છે. એટલે કે પ્રથમ 128 અક્ષરો આસ્કીના જ છે. આ રીતે જે સૉફ્ટવેર આસ્કીનો ઉપયોગ કરી રહ્યાં છે. તે કોઈ પણ ફેરફાર કર્યા સિવાય યુનિકોડની કોઈ પણ પ્રકારના ફેરફાર સિવાય સાથે સામ્યતા ધરાવે છે. યુનિકોડે જુદી જુદી ભાષાઓના વિભાગો તેમની આનુસંગિક લિપિઓ માટે વ્યાખ્યાયિત કર્યા છે. દાત. દેવનાગરી લિપિનો આરંભ 2304 થી થાય છે. અને અંત 2431 એ આવે છે આ જ રીતે વિશ્વની તમામ લિપિઓનો સમાવેશ કરે છે.

13.5.1 યુનિકોડની રચનાના સિદ્ધાંતો(Design Principles Of Unicode)

યુનિકોડની જેના ઉપર રચના કરવામાં આવી છે તે સામાન્ય સિદ્ધાંતો આ પ્રમાણે છે.

→ તે 16 બીટ કોડ છે.

→ ભવિષ્યના વિસ્તરણ માટે પ્રતિનિધિરૂપ વિસ્તાર વ્યાખ્યાયિત કરવામાં આવ્યો છે, જે બે લાઈટ્સના પેકેટનું સંયોજન કરે છે.

→ યુનિકોડ અક્ષરો સાથે સંબંધ ધરાવે છે, બીબાં સાથે નહી.

→ તે લખાણ માટેનું તાર્કિક વર્ણન કરે છે, દા.ત. ડાબી બાજુથી જમણી બાજુ અને હિબ્રુ, ઉર્દૂ વગેરેમાં જમણી બાજૂથી ડાબી બાજુ.

→ બે અક્ષરોનું સંયોજન પણ તાર્કિક ક્રમને અનુસરે છે અક્ષરોનું સંયોજન એ જગ્યા રાખ્યા સિવાયનું સંયોજન છે જે નવા અક્ષરની રજૂઆત માટે સંયોજનની રચના કરે છે.

13.5.2 યુનિકોડનું માળખું (Structure of UNICODE)

યુનિકોડ એ પ્રારંભમાં સૉફ્ટવૅર સંગઠનો માટે રસનો વિષય હતો. શરૂઆતમાં જુદી જુદી ભાષાઓના સૉફ્ટવૅર ઉત્પાદન કરવા તે જ ભાષામાં આસ્કી કોડનો ઉપયોગ કરવો જરૂરી હતો, પરંતુ એકવાર યુનિકોડ આવી ગયો ત્યાર પછી જુદી જુદી ભાષામાં સૉફ્ટવૅર લખવાની આવશ્યકતા રહેતી નથી. તેથી તે સૉફ્ટવૅર વિકસાવનાર સંસ્થાઓ માટે પાયાની વસ્તુ બની ગઈ છે. બહુભાષીય પ્રત્યાયન માટ આધાર પૂરો પાડે છે તેથી તે લોકપ્રિય બનવા પામેલું છે.

યુનિકોડ એ સંપૂર્ણ રીતે ISO/10646 માનાંક સમૂહ જે ઈન્ટરનેશનલ ઓર્ગેનાઈઝેશન ફૉર સ્ટાન્ડડાઝેશન અને ઈન્ટરનેશનલ ઈલેક્ટ્રોની કલ કમીશન-1993 એ વિકસાવેલ તેને સક્ષમ છે. બીજું એક રસપ્રદ સાહસ છે. જે પણ 16 બીટની ખાત્રી આપે છે. પણ તે 16 બીટની મર્યાદા મુકતું નથી પરંતુ જો અક્ષરોની રજૂઆત માટે વાસ્તવિક જરૂરિયાત હોય તો 32 બીટ પૂરા પાડવાની પ્રતિજ્ઞાબદ્ધ છે. તે આઠના ગુણાંકમાં ઉપયોગ કરે છે, કારણ કે પદ્ધતિમાં 8 બીટ ગુણાંક પદ્ધતિઓ ઉપયોગ કરવામાં આવ્યો છે. આ કોડને યુનિવર્સલ ઓકટેટ કોડ(UCS) તરીકે ઓળખવામાં આવે છે. જ્યારે આ અક્ષરકોડ પદ્ધતિ 16 બીટર અને 32 બીટ્સનો ઉપયોગ કરે છે. ત્યારે તેને અનુક્રમે UCS -2 અને UCS -4 તરીકે ઓળખવામાં આવે છે.

13.5.3 યુનિકોડ અને તેના લક્ષણો(UNICODE AND ITS FEATURES).

અત્રે તે સુવિદિત છે કે યુનિકોડ 16 બીટ અક્ષરકોડ છે અને વિશ્વની અસ્તિત્વ ધરાવતી તમામ લિપિઓને તે પાળવા પ્રયત્ન કરે છે. પરંતુ તેની સાથે થોડાક પડકારો પણ તારવવામાં આવ્યા છે.

→ તેને અસ્તિત્વ ધરાવતી આસ્કી પદ્ધતિની કાળજી રાખવી પડે છે.(ભારતીય લિપિઓમાં સંદર્ભમાં)

→ કેટલાંક અક્ષરો સંયોગીકરણનો ઉપયોગ ધરાવતા હોય છે. તેથી તેને સંયોગીકરણની પણ કરવી પડે છે.(ખાસ ચોક્સાઈથી કહીએ આ પ્રશ્નનો ઉકેલ હજુ પણ આવ્યો નથી કારણ કે તે અક્ષરોના સ્વરૂપનો ગુણધર્મો છે જેનો યુનિકોડ અક્ષરો લખવામાં ઉપયોગ કરે છે.)

13.5.3.1 સોળ બીટ અક્ષરકોડ(Sixteen Bit Charater Code).

તે સુવિદિત છે કે, યુનિકોડ એ 16 બીટ અક્ષરકોડ છે. તે 65536 અક્ષરોને સુવિધા પૂરી પાડે છે; પરંતુ, સહયોગી સંગઠનો તમામ અક્ષર સ્થાન વિસ્તારેલ નથી. 2048 અક્ષર સ્થાન ભવિષ્યના અક્ષર સ્થાન માટે બાકી રાખ્યા છે. યુનિકોડનો આ વિસ્તાર પ્રતિનિધિરૂપ વિસ્તાર(Surrogate Area) તરીકે ઓળખાય છે. જે 55296-57343 અક્ષર સ્થાન આવરી લે છે. બે પ્રકારના પ્રતિનિધિરૂપ વિસ્તાર છે.

(અ) ઉચ્ચસ્તરીય પ્રતિનિધિરૂપ વિસ્તાર(High Leval Surrogat area) તે 55296 અક્ષર સ્થાનથી શરૂ થાય છે અને 56319માં અંત આવે છે. અર્થાત્ કુલ 1024 અક્ષર સ્થાને છે.

(બ) નિમ્ન સ્તરીય પ્રતિનિધિરૂપ વિસ્તાર(Low Leval Surrogate) તેની શરૂઆત 55320 અક્ષર સ્થાનથી થાય છે. અને અંત 57343માં અંત આવે છે. અર્થાત્ 1024 અક્ષર સ્થાન ધરાવે છે.

ઉચ્ચસ્તરીય પ્રતિનિધિ વિસ્તાર અને નિમ્ન સ્તરીય પ્રતિનિધિ વિસ્તાર જોડકું બનાવે છે અને અક્ષર પ્રતિનિધિત્વ કરે છે. તેનો અર્થ એ થાય કે આ અક્ષર 4 બાઈટ્સ અક્ષરો છે અને તે રીતે કુલ 10,48,576 અક્ષરની રજૂઆત થઈ શકે છે જે ભવિષ્ય માટે પણ ઉપયોગી થઈ શકે છે. આ ઉચ્ચસ્તરીય પ્રતિનિધિ અને નિમ્ન સ્તરીય પ્રતિનિધિ વિસ્તારનું સંયોગીકરણ પ્રતિનિધિનું જોડકું બનાવે છે.

13.5.3.2 કાર્યક્ષમતા(Efficiency)

યુનિકોડ એ વધુ કાર્યક્ષમ કોડ છે તે નિશંક બાબત છે. સામ્યતા ધરાવતા અક્ષરોનું પુનરાવર્તન કરવામાં આવતું નથી. તેથી તેમને સરળતાથી તેઓ અન્ય સમૂહમાં સ્થાન ધરાવતા હોય ત્યાંથી ઉછીના લેવામાં આવે છે. દા.ત હિન્દી અક્ષરો 2304 થી 2431 સુધી બંગાળી અક્ષરો 2432થી 2559 સુધી અને તેલુગુ અક્ષરો 3072 થી 3199 સુધી અને તેવા બીજા.

છાપી ન શકાય તેવા અક્ષરો(Carriage Return(CR)) માટે કોડ મૂલ્ય 14 છે.(જૂના આસ્કીનું મૂલ્ય) અને તે અંગ્રેજી મૂળાક્ષર સાથે જાળવી રાખવામાં આવે છે જ્યારે અન્ય લિપિઓનું સંકેતિકરણ કરવામાં આવે ત્યારે જ્યાં જરૂર હોય ત્યાં યુનિકોડ 14 મૂલ્યનો હજુ પણ ઉપયોગ કરે છે. એકંદરે કન્નડ ભાષામાં છપાતો ન હોય તેવો અક્ષર અંગ્રેજીમા છપાતા અક્ષર કરતાં જુદો નથી. આ બાબત એક સૂત્રતાથી આસ્કીના તમામ વિશિષ્ટ અક્ષરોને લાગુ પડે છે આ રીતે યુનિકોડ પુનરાવર્તન અટકાવે છે અને સાંકેતિકરણમાં કાર્યક્ષમતા વધારે છે.

13.5.3.3 પ્રતિકો નહીં અક્ષરો(Characters not Glyph)

વાક્યમાં સૌથી નાનો એકમ અક્ષર છે જે લખી શકાય છે અને તેને અર્થમૂલ્ય હોય છે. બીબાં સ્વરૂપે પ્રતિકોએ બીજું કંઈ નથી પરંતુ આલેખન સ્વરૂપમાં અક્ષરોને ખુલ્લા કરાય છે. અર્થાત્ અર્થાત્ અક્ષરો પડદા ઉપર કેવી રીતે પ્રદર્શિત થાય છે. દા.ત. ઘાટા અક્ષર, ત્રાંસા મરોડ વાળા અક્ષર(Italic) અથવા બીજી કોઈ રીતે હોય તે બીબાં પ્રતિકો અક્ષરોના સ્વરૂપ દ્વારા નક્કી થાય છે, અને યુનિકોડ અક્ષરોનું મૂલ્ય એક જ સરખું રહે છે. અક્ષરોને બીબાં પ્રતિક સ્વરૂપે પ્રદર્શિત કરવા, અક્ષરોના સ્વરૂપને બીબાં સંકેત સ્વરૂપ સ્ત્રોતમાં હોવું જરૂરી છે જે શા માટે તે સંદર્ભમાં કહેવામાં આવે છે કે તે લખવા માટે અક્ષરોના સ્વરૂપનો ગુણવત્તા ગુણધર્મ છે.

કોઈ એક વ્યક્તિ શબ્દમાં JAVA™ લખવા પ્રયત્ન કરે તો તેનો દેખાવ JAVA ના ઉપર બીબાં સંકેતના કારણે હોય છે. જે અધિલેખ(Supper Script) બનાવે છે અને તેને અક્ષરના સ્વરૂપનો ટેકો હોય છે. પરંતુ T નું મૂલ્ય 84 હશે અને M નું મૂલ્ય 77 હશે. આમ શા માટે ? તે માટે કહેવામાં આવે છે કે યુનિકોડએ અક્ષરો માટેનો કોડ છે. બીબાં સંકેતો માટે નથી.

13.5.3.4 સપાટપાઠ(Plain Text)

ઉપર નોંધ્યું છે તે પ્રમાણે યુનિકોડને બીબાં સંકેત સાથે કઈ કરવાનું રહેતું નથી અર્થાત્ તે ફક્ત સપાટ પાઠ સાથે કાર્ય કરે છે. દા.ત જો એક અક્ષર 65 લખે તો તે 'A' તરીકે રજૂઆત થાય છે.(જેનું દ્વિઅંકી પદ્ધતિમાં મૂલ્ય 0000000001000001 હોય છે.) અક્ષર સાથે સ્વરૂપ ઉમેરાય છે. કે નહી તેની સાથે સંબંધ નથી. જેને સમૃદ્ધ અક્ષર(Rich Character) તરીકે ઓળખવામાં આવે છે. સમૃદ્ધ પાઠ રજૂ કરવા માટે કોઈ પણ માર્કઅપ ભાષા જેવી કે HTML અથવા XML અથવા MS Office જેવા સૉફ્ટવેરનો પણ ઉપયોગ થઈ શકે છે.

13.5.3.5 તાર્કિક ક્રમ(Logical Order)

યુનિકોડ અક્ષરો તાર્કિક ક્રમમાં સંગ્રહ કરે છે તાર્કિક ક્રમ અર્થાત્ સ્મૃતિમાં જે ક્રમમાં અક્ષરો સંગ્રહ કરવામાં આવે છે. તે જ ક્રમમાં અક્ષરો કીબોર્ડ ઉપર ઉકેલ માટે ઉપયોગ કરવામાં આવે છે. કેટલીક જગ્યાએ તેનો ભંગ કરવામાં આવે છે. કારણ કે પાઠની રજૂઆત ઉલટા ક્રમમાં હોય છે. દા.ત ઉર્દૂ ભાષામાં જમણી બાજૂથી ડાબી બાજુ લખવામાં આવે છે તેવી જ રીતે હિબ્રુમાં પણ તેથી જ્યારે અક્ષરો ટાઈપ કરવામાં આવે ત્યારે રજૂઆતનું સ્વરૂપ પણ જમણી બાજૂથી ડાબી બાજુનું હોય છે. પરંતુ જે ક્રમમાં અક્ષરો સ્મૃતિમાં રજૂ થાય છે. તે એક સરખો હોય છે અને યુનિકોડ તે આપમેળે જ કરે છે. યુનિકોડનું આ લક્ષણ દિશામુખી ક્રમ તરીકે ઓળખાય છે.

13.5.3.6 એકસૂત્રીકરણ(Unification)

યુનિકોડ જુદી જુદી ભાષાના એક જ સરખા અક્ષરોને ઓળખવા માટે અને યુનિકોડ ટેબલ ઉપર એક જ જગ્યાએ મૂકવાની કાળજી રાખે છે. ઉદાહરણ, તરીકે, વિરામ ચિહ્નો સામાન્ય અક્ષરો વગેરે.

પૂર્વ એશિયાઈ લિપિઓના અક્ષરોને Han table તરીકે ઓળખાતા અલગ ટેબલમાં રાખવામાં આવ્યા છે. જે ચીની, જાપાની, કોરિયન(CJK)ને આવરી લે છે,

કારણ કે આ લિપિઓમાં એક જ સરખા અક્ષરો ઘણા છે. ઘણા એક જ સમયે તે જ અક્ષર બધી ભાષાઓમાં ઉપયોગ થાય છે. તે જ રીતે યુનિકોડને સંયોગીકરણ ટેબલ છે. જે જગ્યા ન રાખવાના અક્ષરો(સંયુક્તાક્ષર) આવરી લે છે. બીજા અક્ષરો સાથે સંયોજાય છે.

13.5.3.7 સક્ષમતાવાળા અક્ષરો(Compatible Character)

સક્ષમતાવાળા અક્ષરો એવા અક્ષરો છે જેવું સાંકેતિકરણ થયું નથી; કારણ કે, તેઓ કેટલાંક અક્ષરોના જુદા જુદા સ્વરૂપના છે જેનું સાંકેતિકરણ થઈ ગયું છે.

ઉદાહરણ તરીકે

વિસર્ગચિહ્ન(:) અક્ષર સ્થાન 58ને સક્ષમતાવાળા બે સ્થાન છે એક અક્ષર સ્થાન 847 અર્થાત્ ગુણોત્તર(Ratio) અને બીજું 1417 છે અર્થાત્ બીજું પૂર્ણવિરામ. વિસર્ગચિહ્ન(:) અને ગુણોત્તરચિહ્ન(:) વચ્ચે માત્ર રચના કરવા અંગેનો તફાવત છે. ગુણોત્તર ચિહ્નમાં બિંદુઓ વચ્ચેનું અંતર વધુ છે. આ અક્ષરો સક્ષમતા વિસ્તારમાં સ્થાન પામે છે. જે કોઈપણ વ્યક્તિ અક્ષરોના સ્થાનની નિર્દેશિકામાં જોઈ શકે છે. પરંતુ તે સાચું હોવું જરૂરી નથી કારણ કે બધા જ અક્ષરો આ વિસ્તારમાં સ્થાન પામે છે ઉદાહરણ તરીકે લેટિન અક્ષર Aની જોડણી(રૂપભેદ)(અક્ષર સ્થાન 65) ને પૂર્ણ લંબાઈવાળી લેટિનમાં Aની બીજી જોડણી છે જેનું યુનિકોડ ટેબલમાં છેલ્લે સ્થાન આપવામાં આવ્યું છે.

13.5.3.8 ગતિશીલ અક્ષરોની રચના(Daynamic Composition)

ગતિશીલ અક્ષરોની રચનાનો ઉપયોગ વધુ ભારવાળા સ્વરૂપના સર્જન માટે થાય છે. યુનિકોડ અક્ષર આલેખનમાં એક વિભાગ ઘટક હોય છે. જેને 'અક્ષરોની જોડાણ પ્રક્રિયા વિભાગ' તરીકે ઓળખવામાં આવે છે. 'A'નું સર્જન કરવા અક્ષરો Aનો ઉપયોગ થાય છે.

A → A +'

અલબત્ત બીબાં સંકેત 'A' અક્ષર સ્વરૂપ(આકાર) સમૂહ ઉપયોગો કરી મંજૂરી આપે છે. ભારતીય ભાષાઓમાં સ્વર અને વ્યંજન બંને જોડાણ પ્રક્રિયા અક્ષરો તરીકે(બંને વચ્ચે જગ્યા રાખ્યા સિવાય) ઉપયોગો થઈ શકે છે.

13.5.3.9 સમકક્ષ શ્રેણી(Equivalent Sequence)

યુનિકોડ અક્ષર સમૂહમાં પૂર્વ નિર્ધારિત અક્ષર કોડ પ્રાપ્ય છે, પરંતુ તેજ અક્ષરોનું સર્જન અક્ષરોની ગોઠવણી પ્રક્રિયા ના ઉપયોગ દ્વારા થઈ શકે છે. દા.ત,

A	→	A	+	’
192		65		768

તેથી આ ઉદાહરણમાં 192 એ 65 અને 768 ને સમકક્ષ છે. તેનો અર્થ 65+768 એ નિયમાનુસાર 192 છે. તે પદ્ધતિની જવાબદારી છે કે કોઈ પણ એક સ્વરૂપમાં રૂપાંતર કરવું અર્થાત્ 192 અથવા 65+768 આ બાબત સામાન્યીકરણ તરીકે ઓળખાય છે.

13.5.3.10 રૂપાંતરિતા(Convertibility).

યુનિકોડ, કોઈપણ માનાંકિત કોડમાં રૂપાંતર કરવા માટે પ્રતિજ્ઞાબદ્ધ છે, કારણ કે દરેક અક્ષરને વિભિન્ન નંબર આપવામાં આવ્યાં છે. યુનિકોડનું આ ખાસ લક્ષણ 1993ના કોડના સંસ્કરણમાં બહાર આવ્યું છે.

◆ **તમારી પ્રગતિ ચકાસો**

(4) યુનિકોડ માનાંકમાં અગત્યના લક્ષણો કયા કયા છે ?

નોંધ :- (1) નીચે આપેલી જગ્યામાં તમારો ઉત્તર લખો.

(2) એકમને અંતે આપેલ ઉત્તરો સાથે તમારો ઉત્તર ચકાસો.

....

....

....

....

....

....

13.6 યુનિકોડ દ્વારા વેબ આધારિત વિષય વસ્તુ વિકાસ(WEB CONTENT DEVELOPMENT THROUGH UNICODE).

યુનિકોડનો સૌથી જાણીતો ઉપયોગ એ ઇન્ટરનેટ ઉપર બહુભાષીય વિષયવસ્તુ વિકાસનો છે. આદર્શ બાબત એ છે કે દરેક કોઈ પણ ભાષામાં લખવા સક્ષમ હોવું જોઈએ. સાંપ્રત સમયમાં વિન્ડો 2000 અને ત્યારબાદ વિન્ડો એક્સપી (Windo xp) યુનિકોડને સક્ષમ છે. હજુ નક્ષત્રના આધાર ઉપર ઘણું કરવાનું બાકી છે. વિન્ડો 2000 ફક્ત બે ભારતીય ભાષાઓને આધાર આપે છે. અર્થાત્ દેવનાગરી અને તમિળ બ્રાઉઝરમાં અક્ષરોની રજૂઆત બે રીતે થઈ શકે છે. પ્રથમ રીતમાં ભાષાની રીત બદલવી અને બીજી રીતમાં અક્ષરોના કોડના ઉપયોગ દ્વારા કારણ કે HTML યુનિકોડ ને આધાર પૂરો પાડે છે. નીચેનું ઉદાહરણ બતાવે છે કે પૃષ્ઠ બંને રીતે લખી શકાય છે.

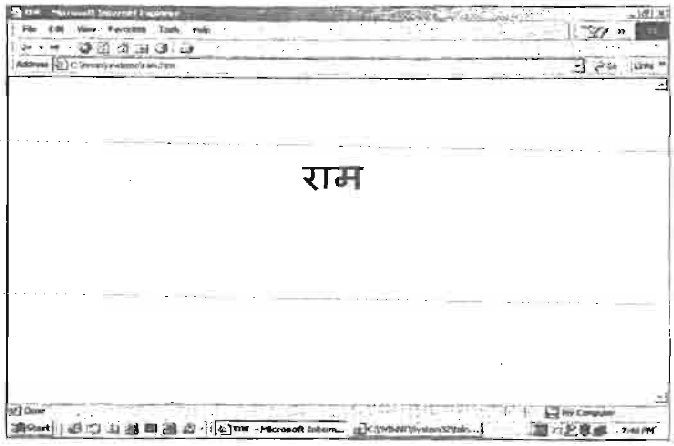

Fig. 13.2: Content Representation in Devanagari

Code Using Devanagari Script in Windows 2000

```
<html>
<title>राम</title>
<body>राम<center><forn size="7">
राम </font></center>
<body>
</html>
```

HTML page Using the UNICODE values in Windows 2000:

```
<html>
<title>&#2352&#2366&#2350 </title>
<body> <center> <font size = "7">
&#2352 &#2366 &#2350 </font> </center>
</body>
</html>
```

બંને અભિગમ સારા છે. પરંતુ વિન્ડોઝ 2000 દેવનાગરી લિપિના ફોન્ટ ટાઈપ કરવામાં આભાસી ચાવીઓ(Virtual keys)ની સુવિધા પૂરી પાડે છે તે સરળ બાબત છે. એક જ પ્રલેખમાં એક કરતાં વધુ લિપિમાં લખવામાં યુનિકોડના ઉપયોગથી સુવિધા પ્રાપ્ત થાય છે.

13.7 યુનિકોડની પ્રયોજિતતાએ(APPLICATIONS OF UNICODE)

યુનિકોડમાં લુપ્ત થઈ ગયેલ સહિત દુનિયાની તમામ લિપિઓનો સમાવેશ કરવા પ્રતિજ્ઞાબદ્ધ છે. તે ભવિષ્યમાં ઉમેરો થનાર લિપિઓના પડકારને પણ ઝીલી લે છે.

13.7.1 બહુભાષીય સાંકેતિકરણ(Multilingua Encoding)

બહુભાષીય સંપાદન એ યુનિકોડનો સૌથી લોકપ્રિય ઉપયોગ છે. કોઈ પણ વ્યક્તિ જો તે / તેણીની ભાષા અને સંકેતો જાણીતી હોય તો તે ગમે તે ભાષામાં લખી શકે છે પરંતુ આ પ્રકારના લખાણ માટે પ્રશ્ન તમામ અક્ષરો માટેના સંકેતો જાણવાનો છે. તેથી કેટલાંક બહુભાષીય સંપાદક જરૂરી છે. જેના સાથે આપણે ગમે તે ભાષામાં લખી શકીએ. તેનો લાભ એ છે કે દરેક લિપિ કોઈ પણ સક્ષમ યુનિકોડ સંપાદકમાં વાંચી શકાય.

13.7.2 બહુભાષીય લિપિયાંતર(Multilingual Transliteration)

એકવાર અક્ષરોના સંકેતોની જાણકારી મેળવેલ હોય, લિપિયાંતરનો અભિગમ સ્વીકારી શકાય. આ લક્ષણ ભારતીય ભાષાઓ સંદર્ભમાં ખાસ લાભપ્રદ છે કારણ કે તમામ ભારતીય ભાષાઓ વધતે ઓછે અંશે એક જ સ્વર અને વ્યંજનનો ઉપયોગ કરે છે. જોકે કેટલીક ભાષાઓમાં કેટલાંક વધારાના અક્ષરો છે પરંતુ યુનિકોડમાં આ એવી રીતે ગોઠવે છે કે જેથી જુદી જુદી લિપિઓ ધરાવતા એક જ સરખો ઉચ્ચારવાળા અક્ષરનું મૂલ્યમાં નક્કી કરેલ તફાવત રહેવા પામે છે. દાખલા તરીકે તેલુગુ અને દેવનાગરીના તમામ અક્ષરોનો તફાવત 768 છે.

13.7.3 બહુભાષીય કાર્ય(Multilingual Works).

યુનિકોડમાં બહુભાષીય કાર્ય જેવું કે બહુભાષીય શબ્દકોશનું સંપાદન જુદી જુદી ભાષામાં પાઠ સાથે ઓનલાઈન શોધ થઈ શકે.

13.8 ગ્રંથાલયોમાં યુનિકોડ પ્રયોજિતતા(APPLYING UNICODE TO THE LIBRARIES)

ભારતએ વૈવિધ્યતાવાળો દેશ છે. ઘણી ભાષાઓ બોલાય છે. અને ઘણી લિપિઓનો ઉપયોગ થાય છે. એક સર્વે પ્રમાણે અંગ્રેજી ભાષા 10% કરતા પણ ઓછા લોકો દ્વારા બોલાય છે. ભારત અનેક ગ્રંથોમાં સમાય તેટલી માહિતી ઘણી સ્થાનિક ભાષાઓમાં ધરાવે છે. ગ્રંથાલયો ખાસ કરીને વિજાણું ગ્રંથાલયો સ્ત્રોત પ્રલેખની ભાષા અને લિપિમાં માહિતીની નોંધ કરી શકે છે. પ્રલેખની ભાષામાં વાઙ્મય સૂચિગત માહિતી પણ પ્રદર્શિત કરી શકાય છે. એક વાર યુનિકોડ સ્વીકારવામાં આવે તો કોઈ પણ લિપિયાંતરનો લાભો પ્રાપ્ત કરી શકે છે. તેથી જ્યારે અને જ્યાં ડેટાની જરૂરિયાત હોય ત્યારે લિપિયાંતર કરી શકાય છે. ભારતની ઘણી લિપિઓમાં સંસ્કૃત સાહિત્ય પ્રાપ્ય છે. તે સામાન્ય બાબત છે. આ પ્રકારના નમૂનાઓમાં જો સંસ્કૃત પાઠ દેવનાગરીનો ઉપયોગ કરી ઉકેલ મેળવવામાં આવેતો કોઈ પણ વ્યક્તિ તે પાઠ પોતાની લિપિમાં વાંચી શકે છે(ઉદાહરણ તરીકે તેલુગુ) બીજા શબ્દોમાં કહીએ તો તેલુગુમાં પાઠ સંગ્રહ કરવાની જરૂર નથી. દેવનાગરી લિપિનું લિપિયાંતર ઉપર વાંચી શકાય છે.

મેટાડેટાના વાઙ્મય સૂચિગત માહિતી ઉદાહરણમાં XML(Extensible mark up language) નો ઉપયોગ કરીને માહિતીનું અર્થમૂલ્ય આપી શકાય છે. શરત એટલી જ છે કે કાર્યક્રમમાં ભાગ લેનાર

ગ્રંથાલયો પાસે ચોક્કસ બંધકોના સમૂહનો સર્વ સામાન્ય અભિપ્રાય હોવો જોઈએ. આ રીતે એક ગ્રંથાલયની સૂચિ બીજી ભાષામાં જોઈ શકાય છે અને વાંચી શકાય છે.

◆ **તમારી પ્રગતિ ચકાસો.**

(5) યુનિકોડની શક્ય પ્રયોજિતતાઓની ચર્ચા કરો.

નોંધ :- (1) નીચે આપેલી જગ્યામાં તમારો ઉત્તર લખો.

(2) એકમને અંતે આપેલ ઉત્તરો સાથે તમારો જવાબ ચકાસો.

....

....

....

....

....

....

13.9 યુનિકોડ સાથે સંકળાયેલા પ્રશ્નો(PROBLEMS ASSOCIATED WITH UNICODE)

યુનિકોડ એ 16 બીટ્સવાળો કોડ છે જે વિશ્વની તમામ લિપિને પ્રદર્શન કરે છે. તેમ છતાં ચીની લિપિ સંદર્ભમાં 65,536નો આંકડો અપુરતો છે. ચીનીએ ભાવ લિપિ છે જેને 40,000 કરતાં વધુ અક્ષરો છે. પરંતુ ફક્ત 6000 અક્ષરો દૈનિક વપરાશમાં ઉપયોગ થાય છે. જાપાનીમાં, ચીની કરતાં થોડાકે ભેદ-તફાવત છે અને રાજકીય રીતે તેઓએ અલગ વિભાગ માટે માંગણી કરેલી છે. જેને કોડ સાથે રાજકીય દ્વિધાઓનું સર્જન કર્યું છે. [Goundry, 2001]

યુનિકોડ વિન્ડો ચાલક પદ્ધતિ(Operating System) માં ઉપયોગ થાય છે. લિનક્સ પર્યાવરણનો વિકાસ થયો છે. પરંતુ હજુ બીબાં સંકેતમાં પ્રશ્નો અસ્તિત્વમાં છે. જો કે વિન્ડો 9x, Me, Xp વગેરેને USP. 10. DLL સાથે લાદી લઈ જઈ શકાય છે. જેને અનન્ય લહિયા તરીકે ઓળખવામાં આવે છે. પરંતુ આ કાર્યક્રમનાં કેટલાંક લક્ષણો, જટિલ લિપિઓની પ્રદર્શિતતા માટે કાર્યો કરતા નથી. યુનિક્સમાં પાઠ સંપાદકો છે જે યુનિકોડને સહાય કરે છે. યુડિટ(Yudit) આવો સંપાદક છે જે યુરોપિયન ભાષાઓને સહાય કરે છે યુનિકોડ ને સક્ષમ ઈમેક્સ(Emacs) ડાઉનલોડ માટે પ્રાપ્ય છે. આ ઉપરાંત ભારતીય ભાષાઓમાં લિનક્સ બનાવવા માટે પ્રયત્નો કર્યા છે, પરંતુ હજુ પણ આ યોજનાઓ મહદ્અંશે કાર્યો જુદા પાડવાનું પ્રાથમિક કાર્ય કરે છે.

બ્રાઉઝર ચાલક પદ્ધતિની સહાય સિવાય એકલું બહુભાષીય પાઠ પ્રદર્શિત કરી શકતું નથી. વિન્ડોઝ 9x અને Me સાથે યંત્રમાં(કૉમ્પ્યુટરમાં) કોઈ પણ વ્યક્તિ ભારતીય ભાષાઓ માટે સંયુક્ત(જોડાણવાળા) અક્ષરો મેળવી શકાતો નથી. આ માટે બીબાં સ્વરૂપને અનુરૂપ સંકેતો નથી એ પ્રાથમિક કારણ નથી. ચાલક પદ્ધતિઓ પાસે ભારતીય ભાષાઓ માટે અક્ષરોનું વર્ણન આપતાં યંત્રો નથી તે કારણ છે.

ઈસ્કીના ઉપયોગ સાથે એક ફાયદો એ છે કે લિપિયાંતર ઊડીને(ઝડપી) માત્ર સાદી રીત બદલીને પ્રાપ્ત કરી શકાય છે. કારણ કે ઈસ્કી એક જ સરખા નામની જગ્યાનો ઉપયોગ કરે છે પરંતુ તે મોંઘી પુરવાર થાય છે અને એક જ પડદા ઉપર એક કરતાં વધુ ભારતીય ભાષાઓ પ્રદર્શિત કરવાનું તે અશક્ય બનાવે છે. લિપિયાંતર યુનિકોડમાં પણ પ્રાપ્ત કરી શકાય છે પરંતુ તે માટે બે જુદી જુદી ભાષાઓના સંકેતોના મૂલ્યના તફાવતોની ગણતરી કરવા માટેના કાર્યક્રમ(Progarmme) આવશ્યક છે.આ વ્યાપાર સંબંધી છે.

13.10 સારાંશ(SUMMARY)

સાંપ્રત સમયમાં જુદી જુદી ભાષાઓમાં વેબ ઉપર માહિતી આપવાનું વલણ બદલાઈ રહ્યું છે. ઇન્ટરનેટ માહિતીનું વૈશ્વિકરણ લાવી રહ્યું છે. જ્યારે વિશ્વમાં સંસ્કૃતિમાં ભાષાઓ અને લિપિઓની વૈવિધ્યતા સંદર્ભમાં ઘણું કરવાનું રહે છે. યુનિકોડ માહિતી ટેક્નોલૉજી અને આભાસી વિશ્વ(ઇન્ટરનેટ)માં વિશ્વની લિપિઓનું વૈશ્વિકરણ કરવાનો પડકાર ઝીલ્યો છે. તેના વિકાસમાં કેટલાંક તકનીકીને લગતા તથા રાજકીય અવરોધો છે. સમયાંતરે તેમાં ફેરફાર આવી રહ્યો છે. જે વધારાની બાબતો સમાવેશ કરે છે. અને નિયમો સુધારે છે. જો કે આ માનાંક સોફ્ટવેર કંપનીઓના પ્રયત્નોનું છે. પરંતુ સાંપ્રત વિશ્વમાં તેની વિવિધ પ્રયોજિતતાઓ રહેલી છે અને ગ્રંથાલયો માહિતી સ્રોતની ભાગીદારીમાં ઉભા રહેવાની સારી તક રહેલી છે.

13.11 તમારી પ્રગતિ ચકાસોના ઉત્તરો(ANSWER TO SELFCHECK EXERCISES)

(1) કમ્પ્યૂટરમાં દરેક અક્ષરને એકજ નંબરનો દરજ્જો અપાય છે. કારણ કે કમ્પ્યૂટર ફક્ત વિદ્યુત પ્રવાહ સ્વરૂપમાં જ ફક્ત નંબરને સમજી શકે છે પરંતુ વાસ્તવિક પ્રક્રિયા તે કરતાં ઘણી જટિલ છે. કારણ કે કમ્પ્યૂટર એ વિજાણું પ્રયુક્તિ છે અને તે પણ ફક્ત વિદ્યુત દબાણ ‘ON’ અને ‘OFF’ સમજે છે વિદ્યુત દબાણ ON અને OFF રજૂઆત તરીકે અનુક્રમે 1 અને 0 સંજ્ઞા વિકસાવવા પાછળનું કારણ આ એક કારણ છે સંજ્ઞાની આ ખાસ પદ્ધતિને દ્વિઅંકી પદ્ધતિ તરીકે ઓળખવામાં આવે છે અને આવી પદ્ધતિનો ઉપયોગ કરીને અક્ષરોનું સંકેતિકરણ કરવાની પદ્ધતિને ‘અક્ષર સંકેતિકરણ’ તરીકે ઓળખવામાં આવે છે. આસ્કી(ASCII) માં જ્યારે ‘A’ સૂચક ચાવી દબાવવામાં આવે ત્યારે કમ્પ્યૂટર તેને 01000001 તરીકે(અથવા દશાંશ 65) સમજે છે અને પડદા ઉપર ‘A’ છાપે છે.

(2) આસ્કીની રચના IBM, AT & T અને તેના પેટા કંપની Teletype દ્વાર રચાયેલ કાર્યશી જૂથ જેને x 3.4 તરીકે ઓળખવામાં આવે છે. તેના દ્વારા કરવામા આવેલ મૂળભૂત રીતે આ કોડ 8 બીટ અક્ષર સાંકેતીકરણ માટે હતો અને તેની ગોઠવણી તાર્કિક હતી અર્થાત્ કોઈ પણ વ્યક્તિ જુદા પાડવા અંગેનો કાર્યક્રમ લખવાન ઈચ્છા રાખે તો તે આસ્કીનો ઉપયોગ કરીન તેનું સંચાલન કરી શકે છે પરંતુ કોડનું ખાસ આ સંસ્કરણ ફક્ત મોટા અક્ષરો(કેપીટલ) સ્વીકારતું હતુ અને વચ્ચે ઘણી ખાલી જગ્યાઓ એવી ધારણા સાથે છોડી દેવામાં આવેલા છે તેનો ઉપયોગ રોમન નાના અક્ષરો માટે કરી શકાશે.(વધુ ઉંડાણ માટે વિભાગ-3 વાંચો)

(3) ISCII(ISCll- Indian Standard Code for Information Interchange) એ ASCIIનું વિસ્તારેલ સ્વરૂપ છે. તે છેલ્લાં 128 અક્ષરોના સ્થાન ભારતીય લિપિઓમાં રજૂઆત માટે 0-255 ના વિસ્તારમાં ઉપયોગ કરે છે. તે આસ્કીના જેવા જ પ્રકારનું કીબોર્ડ ઉપર તેની સાથે અન્ય 10 ભારતીય ભાષાઓના અભિલેખોનો ઉપયોગ કરે છે. તે નિયંત્રક અક્ષર SO(Shift Out) અને SI(Shift In) નો અનુક્રમે ASCII અને ISCII માટે ઉપયોગ કરે છે. ઈસ્કીમાં આસ્કીના અક્ષરો 8 બીટકોડના ટેબલના નીચેના અડધા ભાગમાં(0-127) સ્થાન આપેલ છે. જ્યારે ભારતીય લિપિઓના અક્ષરો ઉપરના અડધા ભાગ(160-225)માં સ્થાન આપેલ છે. ઈસ્કી નીચેની 10 ભારતીય લિપિઓને સેવા આપે છે. દેવનાગરી, ગુજરાતી, પંજાબી, બંગાળી, અસામીઝ, ઉરિયા, તેલુગુ, તમિળ, મલયાલમ કન્નડ, ઈસ્કી કોડ ટેબલ એ ઉપર દર્શાવેલ લિપિઓ માટે જરૂરી તમામ અક્ષરોનો સુયોગ્ય સમૂહ છે. પ્રથમ સંસ્કરણ 1983માં પ્રકાશિત થયેલ ત્યાર બાદ 1986 અને 1988માં સુધારા થયા બાદ બ્યુરો ઑફ ઈન્ડિયન સ્ટાન્ડર્ડ્ઝ(BIS) એ તેનો સ્વીકાર કર્યો.(વાચન માટે વિભાગ-4)

(4) યુનિકોડ અનાવયક અક્ષરો દૂર કરી અક્ષરોના સાંકેતિકરણ માટેનો કાર્યક્ષમ કોડ છે તે 16 બીટ કોડ(65,536 અક્ષર માટે અનુમતિ આપે છે) છે અને નીચેના લક્ષણો ધરાવે છે.

સપાટ પાઠ

ગતિશીલ અક્ષરોની રચના

તાર્કિક ક્રમ

એકસૂત્રીકરણ

સક્ષમતા વાળા અક્ષરો

સમકક્ષ શ્રેણી

રૂપાંતરિતતા

(5) યુનિકોડના મૂળભૂત પ્રયોજિતતાઓ પૈકી એક કમ્પ્યૂટર આધારિત અંગ્રેજી સિવાયની ભાષાઓમાં સેવાઓ પૂરી પાડવાની ક્ષમતા છે. તેની કેટલીક અન્ય પ્રયોજિતતાઓ આ પ્રમાણે છે.

→ બહુભાષીય સંકેતિકરણ, માહિતીની જુદી જુદી પ્રાદેશિક ભાષાઓમાં રજૂઆત શક્ય બનાવે છે.

→ બહુભાષીય શોધ

→ બહુભાષીય કાર્યો જેવા કે બહુભાષીય શબ્દકોષ

→ લિપિયાંતર

→ આડ ભાષાઓમાં માહિતી પુનઃપ્રાપ્તિ(CLIR)

13.12 ચાવીરૂપ શબ્દો(KEYWORDS)

આસ્કી (ASCII) : અમેરિકન સ્ટાન્ડર્ડ કોડ ફોર ઈન્ફોરમેશન ઈન્ટરચેઈન્જ.
: (ઇન્ફોરમેશનને બદલે કેરકટર પણ આવી શકે)

બીટ (BIT) : જ્યારે માહિતીનું વિજાણુ સ્વરૂપમાં રૂપાતંર કરવામાં આવે છે. ત્યારે તે એક અને શૂન્યમાં બદલાવવામાં આવે છે તેથી વિજાણું સ્વરૂપમાં તમામ માહિતી બીટ્સની બનેલી હોય છે.

બાઈટ (BYTE) : ડેટા બીટ્સ નો સમૂહ કે જેની પ્રક્રિયા એક સાથે કરવામાં આવે છે. જટિલ રીતે એક બાઈટ, 8 બીટ્સ ધરાવે છે(1 બાઈટ = 8 બીટ્સ) બાઈટ્સના મોટા એકમો કિલો બાઈટ્સ, મેગા બાઈટ્સ, ગીગા બાઈટ્સ ટેરાબાઈટ્સ વગેરે છે.

1 બાઈટ = 8 બીટ્સ, 1કિલો બાઈટ = લગભગ 1000 બાઈટ્સ

1 મેગા બાઈટ 1MB= લગભગ 1,00,000 બાઈટ્સ

1 ગીગા બાઈટ 1GB = 1,000,000,000 બાઈટ્સ = 1000,000 મેગા બાઈટ્સ MB

1 ટેરા બાઈટ્સ TB = 1,000,000,000, 000 બાઈટ્સ = 1000,000 ગીગા બાઈટ્સ GB

સી-ડેક (C- DAC) : સેન્ટર ફોર ડેવલપમેન્ટ ઑફ એડવાન્સ કૉમ્પ્યુટીંગનું ટીગ- ઈન્ડીયા

કોડ : કંઈક રજૂ કરવા માટેના પ્રતિકોનો ગણ દા.ત મોટા ભાગના કમ્પ્યૂટર અક્ષર રજૂ
(Code) કરવા આસ્કી કોડનો ઉપયોગ કરે છે.

ફાઈલ : માહિતીનું સાંકેતિકરણ કરવા માટે જે માળખું છે તેને ફાઈલ કહે છે. જુદી જુદી
(File) ફાઈલોના પ્રકારો પ્રમાણે જુદા જુદા માળખાં હોય છે. ફાઈલનું માળખું પ્રથમ સ્પષ્ટ કરે છે. કે ફાઈલ દ્વિઅંકીમાં કે આસ્કીમાં છે અને બીજા ક્રમે માહિતી કેવી રીતે સંગઠીત કરવામાં આવી છે.

ગીસ્ટ : ગ્રાફીક્સ એન્ડ ઈન્ટેલીજન્સી બેઈઝ્ડ સ્ક્રીપ્ટીંગ ટેક્નોલોજી(GIST)નો વિકાસ
(GIST) સી-ડેક ભારત દ્વારા કરવામાં આવ્યો છે. તે ભારતની મોટાભાગની ભાષાને સહાય આપે છે.

અક્ષર સ્વરૂપમાં ધરાવતાં : આપેલ અક્ષરોનો ખાસ આધાર જે પ્રદર્શિત કરવામાં આવતો હોય છે/
(Glyphs) પડદા ઉપર વર્ણન આપતો હોય છે.

હેક્ઝાડેસીમલ : અંકો પ્રદાન કરવાની પદ્ધતિ જે 16 અંકોના આધારનો ઉપયોગ કરે છે. પ્રથમ 10
(Hexadecimal) અંકો 0-9 અને ત્યાર બાદ બાકીના 6 A થી F હોય છે હેક્ઝાડેસીમલનો ઉપયોગ રંગીન વેબપૃષ્ઠો માટે થાય છે. દા.ત સફેદ રંગ માટે હેક્ઝાડેસીમલ અંક # FFFFFF જેટલો હોય છે.

ઈસ્કી(ISCII) : ઈન્ડીયન સ્ટાન્ડર્ડ કોડ ફોર ઈન્ફોરમેશન ઈન્ટરચેઈન્જ(ઈન્ફોરમેશન ને બદલે કેરેક્ટર પણ વપરાય છે)

લિનક્સ : યુનિક્સ(Unix)પ્રકારની મુક્ત ચાલક પદ્ધતિ જે મૂળભૂત રીતે લિનૂસ ટોરવાલ્ડે
(LINUX) સમગ્ર વિશ્વમાંથી ચાલક પદ્ધતિ વિકસાવનારાઓની સહાયથી વિકસાવેલ જેનો વિકાસ GNU General Public License દ્વારા કરવામાં આવ્યો લિનક્સ માટેના આધાર સંકેતો(Source Code) દરેકને મુક્ત રીતે પ્રાપ્ત છે.

બહુભાષીય ડેટા

યુનિકોડ

Multilingual Date : એક જ ફાઈલ કે પ્રલેખમાં એક કરતાં વધુભાષામાં ડેટા દુનિયાની લિપિઓની કમ્પ્યૂટરમાં રજૂઆત કરવા માટેનો સાર્વત્રિક કોડ.

13.13 સંદર્ભો અને વિશેષ વાચન(REFERENCES AND FURTHER READING)

Ancientscrpts.com Origin of writing. <http:www. ancicentscrpts. com/ws_ orginis,html>.

Bhatt, shashank.(2001). charachter encoding standard for Indian scripts-a reaport. <http:www.cicc.or.jp/english/hyoujyunka/milta4/7-3 India/India.htm>.

Bureau of Indian Standards.(1999). Indian standard Code for Information Interchange(ISCII)(IS13194:1991). <http:/varamozhi. sourceforge.net/ iscii91.pdf>.

Czyborra, Roman.(1997). The ISO 8859 alphabet soup.<http://wwwwbs.cs.tu-berlin.de/~ czyborra/charsets>.

Encyclopedia Britannica.(1982) Developent of script of india.

<http:/www.geocities.com/Athens/Parthenon/2104/scripts.ht,l>.

Goundry, Nourman.(2001) Why Unicode wan't work on the Internet: Linguistic< political, and technical limitations. <http://www.hastingsresearch.com/net04-unicode-limitations.shtml>.

Internet Server Providers Association of India.(2005).Internet services industry - and overview. <http:/www.ispai/com/overview.html>.

Johnson, Dave, concepts off C/Unix internationalization: Unicode.<http://cns-web.bu.edu/pub/djohson/web_files/i18n/unicode.htmal>.

Malaiya Yashwant K.Language and scripts of India. <http:/www.cs.colostate.edu/~malaiya/scripts.html>.

Monotype Imaging Inc. A Unicode Font soulution: a white paper from Manotype Imaging Inc. <http://www.monotypeimaging.com/isv/uni_whitepaper.asp>.

Raje, Chandrashekhar.(2001). Text processing of national languages(india).

<http://www/cicc.or.jp/engilish/hyoujunka/af08/8-05.html>

Searle, Steve j.(2004). A brief history of character codes in North America Europe, and East Asia, <http://tronweb. super-nova.co.jp/characcodehist.html>.

Technology Development for Indian Languages.(2005). Standars-ISCII.

<http://tdil.mit.gov.in/standards.htm/basic_q/html>.

Unicode.(2005) Freqently asked Questions:basic questions.

<http://www.unicode.org/faq/basic_q.html>.

Unicode.(2005). Frequently asked questions: India scripts and languages.

<http://www/unicode/org/faq/indic_q.htmal>

Unicode.(2005). Versions of the Unicode stsndard. <http://www/unicode.org/versions>.

Wodd, Alan.(2005). Alan Wood's Unicode resoures: Unicode and multilingual support in HTML, Fonts.web brosers and other applications. <http://www.alanwood.net/unicode/>.