UNIUNEA EUROPEANĂ

GUVERNUL ROMÂNIEI

Instrumente Structurale
2007-2013

POLITEHNICA DIN BUCUREȘTI
UNIVERSITATEA 1818

Platformă de e-learning și curriculă e-content pentru învățământul superior tehnic

Transmisia datelor multimedia in retele de calculatoare

15. Codificarea bazata pe digrame

Introducere

- Pana acuma am considerat ca toate simbolurile de intrare sunt independente
  - Acest lucru nu este adevarat pentru cele mai uzuale tipuri de date
  - Ex: text, imagini, cod
- Ideea de baza
  - Se vor identifica pattern-urile (modele) simbolurilor
  - Se for codifica aceste modele mai eficient
  - Restul de simboluri vor fi codificate folosind un algoritmul de baza (mai putin eficient)
  - In cele mai multe cazuri se va obtine o performanta mult mai buna
- Nota
  - Aceasta tehnica poate fi folosite in cazul tipurilor de date precum textul
  - Evident, nu va avea succes in cadrul datelor (aproape) intamplatoare - *random*

Codificarea bazata pe digrame

- Dictionar
  - Toate literele din alfabet +
  - Cat mai multe digrame (perechi de litere)
- Exemplu:
  - Marimea dictionarului = 256 intrari
  - Alfabet: caracterele ASCII printabile = 95
  - Digrame: cele mai comune 161 de perechi de caractere
- Alt exemplu: A = {*a*, *b*, *c*, *d*, *r*},
  - Dictionar:

Code	Entry	Code	Entry
000	*a*	100	*r*
001	*b*	101	*ab*
010	*c*	110	*ac*
011	*d*	111	*ad*

Exemplu

Input: abracadabra

Code	Entry	Code	Entry
000	*a*	100	*r*
001	*b*	101	*ab*
010	*c*	110	*ac*
011	*d*	111	*ad*

Output: **101**

Exemplu

Input: --racadabra

Code	Entry	Code	Entry
000	*a*	100	*r*
001	*b*	101	*ab*
010	*c*	110	*ac*
011	*d*	111	*ad*

X

Output: 101

Exemplu

Input: --racadabra

Code	Entry	Code	Entry
000	*a*	100	*r*
001	*b*	101	*ab*
010	*c*	110	*ac*
011	*d*	111	*ad*

Output: 101**100**

Exemplu

Input: ---**ac**adabra

Code	Entry	Code	Entry
000	*a*	100	*r*
001	*b*	101	*ab*
010	*c*	110	*ac*
011	*d*	111	*ad*

Output: 101100**110**

Exemplu

Input: -----**ad**adabra

Code	Entry	Code	Entry
000	*a*	100	*r*
001	*b*	101	*ab*
010	*c*	110	*ac*
011	*d*	111	*ad*

Output: 10110011**0111**

Exemplu

Input: -------**ab**ra

Code	Entry	Code	Entry
000	*a*	100	*r*
001	*b*	101	*ab*
010	*c*	110	*ac*
011	*d*	111	*ad*

Output: 101100110111**101**

Exemplu

Input: ---------**ra**

Code	Entry	Code	Entry
000	*a*	100	*r*
001	*b*	101	*ab*
010	*c*	110	*ac*
011	*d*	111	*ad*

X

Output: 101100110111101

Exemplu

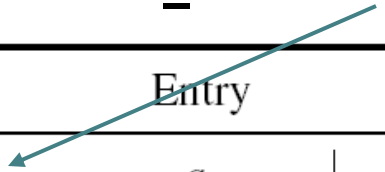
Input: ---------**r**a

Code	Entry	Code	Entry
000	*a*	100	*r*
001	*b*	101	*ab*
010	*c*	110	*ac*
011	*d*	111	*ad*

Output: 101100110111101**100**

Exemplu

Input: ----------**a**



Code	Entry	Code	Entry
000	*a*	100	*r*
001	*b*	101	*ab*
010	*c*	110	*ac*
011	*d*	111	*ad*

Output: 101100110111101100**000**

Problema: Care digrame?

Pair	Count	Pair	Count
eb	1128	*ar*	314
bt	838	*at*	313
bb	823	*bw*	309
th	817	*te*	296
he	712	*bs*	295
in	512	*db*	272
sb	494	*bo*	266
er	433	*io*	257
ba	425	*co*	256
tb	401	*re*	247
en	392	*b$*	246
on	385	*rb*	239
nb	353	*di*	230
ti	322	*ic*	229
bi	317	*ct*	226

Source #1: Textbook chapter (LaTeX)

Pair	Count	Pair	Count
bb	5728	*st*	442
nlb	1471	*le*	440
;nl	1133	*ut*	440
in	985	*f(*	416
nt	739	*ar*	381
=b	687	*or*	374
bi	662	*rb*	373
tb	615	*en*	371
b=	612	*er*	358
);	558	*ri*	357
,b	554	*at*	352
nlnl	506	*pr*	351
bf	505	*te*	349
eb	500	*an*	348
*b**	444	*lo*	347

Source #2: C code