5. Bunjatian E.P.: Metodika Socialjnyĥ Rekonstrukcij v Arĥeologii – Kiev, 1985, 228 p.
6. Häusler A.: Zur Problematik der Gräbersoziologie / Moderne Probleme der Archäologie – Berlin, 1976, p. 83–102.
7. Ĥazanov A.M.: Klassobrazovanije: Faktory i Meĥanizmy / Issledovanija po Obŝĉej Etnografii – Moskva, 1979, p. 125–177.
8. Ionesov V.I.: Stanovlenie i Razvitie Ranneklassovyĥ Otnoŝenij v edlozemledeljĉeskom Obŝĉestve Severnoj Baktrii – Autoreferat Dis. Kand. Ist. Nauk, Samarkand, 1990, 25 p.
9. Iordanskij V.B.: Ĥaos i Garmonija – Moskva, 1982, 344 p.
10. Kameneckij I.S.: Kod dla Opisanija Pogrebaljnogo Obrjada / Arĥeologiĉeskie Otkrytija na Novostrojkaĥ, Vyp.1 – Moskva, 1986, p. 136–194.
11. Kapustin N.S.: Osobenosti Evolucii Religii – Moskva, 1984, 222 p.
12. Leonova N.B., Smirnov Ju.A.: Pogrebenija kak Objekt Formaljnogo Analiza, Vyp. 148 – Moskva, 1977, p. 16–23.
13. Losev A.F.: Filosofia, Mitologia, Kultura – Moskva, 1991, 526 p.
14. Masson V.M.: Ekonomika i Socialjnyj Stroj Drevniĥ Obŝĉestv. – Leningrad, 1976, 192 p.
15. Renfrew C.: Approaches to Social Archaeology – Edinburgh, 1984.
16. Tokarev S.A.: Pogrebaljnye Obyĉai, Iĥ Smysl i Proisĥoĵdenie – "Priroda", N 9, 1985, p 82–87.

Adreso de la autoro:

Vladimir I. Ionesov, C. Sc.,
Supera Scienca Kunlaboranto de la Instituto de Arkeologio de Akademio de Sciencoj de Uzbekistano
ul. Akad. Abdullaeva 3
703051 SAMARKAND
UZBEKISTANO.

Pri resinkronigo en kodoj

Reinhard Fössmeier
AIS San-Marino, sekcio 1
Dollmannstr. 19, DE-81541 München
<refo@ixos.de>

1 Enkonduko

Kiam en iu mesaĝo, konsistanta el vico da unuopaj signoj, perdiĝas parto, ekestas la problemo, kie post la breĉo oni povas rekomenci la legadon de la mesaĝo, por ricevi almenaŭ parton de la enhavo. Se la mesaĝeroj estas literoj, oni povas ekzemple rekomenci ĉe la unua interspaco post la breĉo, sciante ke tie komenciĝas nova vorto. Se la mesaĝeroj estas unuopaj bitoj, la situacio estas malpli favora: Ne sciante, kiom da bitoj perdiĝis, oni devas eksperimenti por eble ricevi sencohavan tekston. Se ekzemple la uzata kodo uzas unu bitokon por ĉiu signo, ekzistas ok malsamaj ebloj por resinkronigi la legadon post la breĉo en la vico. Se la mesaĝo estas sufiĉe redundanca, tamen nur unu el la ok eblaj legaĵoj estas sencohava mesaĝo. Bildo 1 montras la komencon de la ĉarto de la Unuiĝintaj Nacioj (*NI, LA POPOLOJ...*), kodita laŭ la (kvinbita) teleksa kodo, kaj la kvin eblajn legaĵojn de tiu mesaĝo, laŭ la kvin eklego-punktoj.

Breĉo en mesaĝo povas okazi pro perturbo en la transsenda kanalo, aŭ pro difekto de storilo, ekzemple bendo aŭ disko. En ambaŭ okazoj ofte estas sciate, kiom da informo perdiĝis, tiel ke la resinkronigo ne estas problemo, se la uzata kodo havas fiksan vorto-longon. En la ekzemplo de bitoka kodo do ne necesas eksperimentado, se almenaŭ estas sciate ke perdiĝis entjera nombro da bitokoj.

Alia afero estas kodoj kun varia vortolongo. Ĝuste por la longtempa storado de datenoj oni ofte uzas tiajn kodojn por malpliigi la bezonon de storo. Ekzemplo estas NKI-arkivoj (NKI = ne-kodita informo), kie dokumentoj estas storataj en faksimila formo, do kiel bildo el nigraj kaj blankaj punktoj, sen konsidero de la enhava informo. Tiajn bildojn oni kutime densigas laŭ la telekopia aŭ faksimila kodo, laŭ la proponoj T4 [2] kaj T6 [3] de CCITT (konsulta komitato internacia pri telefonio kaj telegrafio).

Tiaj kodoj prezentas duoblan problemon ĉe la resinkronigo: Pro la varia vortolongo, estas nesciate kie estas la limoj inter la signoj; pro la malpliigita redundanco, estas malfacile distingi inter la vera mesaĝo kaj la aliaj eblaj legaĵoj de la sama kodo-vico.

2 Aŭtomata resinkronigo

Kodoj kun varia vorto-longo tamen havas grandan avantaĝon rilate al resinkronigo: Ĝuste pro la varia vortolongo, ili tendencas aŭtomate resinkronigi la legadon de mesaĝo, sendepende de la eklego-punkto.

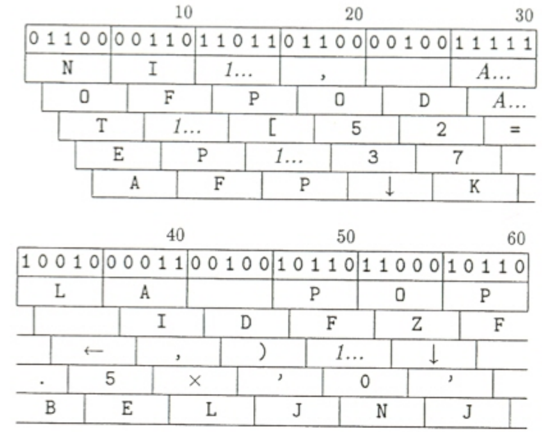
Kiel ekzemplon ni rigardu la kodon de *Huffman* por optimuma kodado de anglalingvaj tekstoj (vidu ekz-e [1]). En bildo 2 ni vidas la komencon de la UN-ĉarto koditan per tiu kodo (sen komo kaj spacoj, kiujn la kodo ne posedas) kaj la diversajn legaĵojn, kiuj rezultas el malsamaj eklego-punktoj.

Ni vidas ke, komencante post la unua, dua kaj tria bitoj, la legado resinkroniĝas post la kvara resp-e la 42-a bito. Por esti tute certa pri la ĝusta resinkroniĝo, sen scio de la ĝusta teksto, ni devus provi dek unu sinsekvajn eklego-punktojn, ĉar la maksimuma longeco de unu signo de la *Huffman*-kodo estas dek unu bitoj. Eklegante post la pozicioj 0 ĝis 10 ni efektive trovas ke ĉiuj legaĵoj kuniĝas ĉe la 42-a bito (vidu tabelon 1); do ek de tie ne estas dubo pri la ĝusta legaĵo.

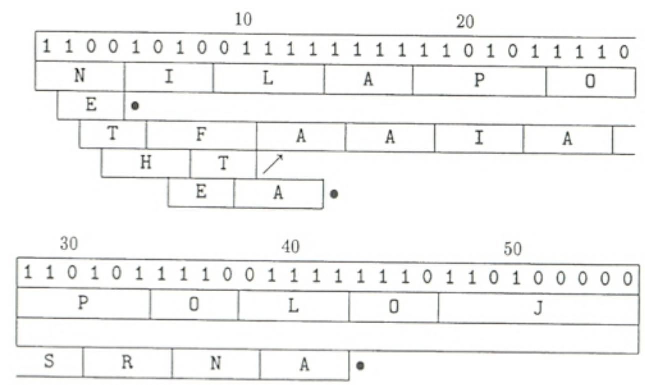
La tendenco al resinkroniĝo tamen ne estas garantio. En la plej multaj kodoj ekzistas ajne longaj vicoj, kie la legado neniam resinkroniĝas; en la *Huffman*-kodo, la vico *10101010...* estas laŭplaĉe dividebla al vico de $I = 1010$ aŭ $H = 0101$. Eĉ kodoj kun tute similaj strukturoj povas havi malsame fortan tendencon al resinkroniĝo: En la unua kodo de bildo 3 ekzistas nur unu perioda vico, kiu estas dividebla laŭ pluraj manieroj; en la dua kodo ekzistas pluraj tiaj vicoj, do la probableco por aŭtomata resinkroniĝo estas malpli granda.

Tabelo 1: Loko de resinkroniĝo en bildo 2, depende de la eklego-punkto

eklego en pozicio:	0	1	2	3	4	5	6	7	8	9	10
sinkroniĝo ĉe:	0	4	42	42	4	42	13	42	8	13	42
nesinkrona vojo:	0	3	40	39	0	37	7	35	0	4	32



Bildo 1: Komenco de la UN-ĉarto, kodita laŭ la teleksa kodo, kaj ĝiaj legaĵoj laŭ diversaj eklego-punktoj (sagoj estas movoj de la ĉareto, *A...* kaj *1...* estas ŝaltoj inter la aroj de la literoj kaj de la aliaj signoj)



Bildo 2: Komenco de la UN-ĉarto, kodita laŭ *Huffman* (• indikas kuniĝon kun la ĝusta legaĵo = sinkroniĝon)

$p_a = \frac{1}{2},\ p_b = \frac{1}{4},\ p_c = p_{\hat{c}} = \frac{1}{8}$	
a = 0 b = 10 c = 110 ĉ = 111	a = 0 b = 11 c = 100 ĉ = 101
111 111 111 = *1* 111 111 11	11 11 11 = *1* 11 11 1 11 0 11 0 11 = *1* 101 101 1 101 11 101 11 = *1* 0 11 11 0 11 1

Bildo 3: Du kodoj por la alfabeto {a, b, c, ĉ}
kun periodaj vicoj, kiuj neniam resinkroniĝas

3 La mezuma nesinkrona vojo

La probableco de vicoj, kiuj ne resinkroniĝas, estas ju pli longa des pli malgranda. Se iu kodo entute havas la kapablon de resinkroniĝo (ni vidis, ke kodoj kun egala vortolongo ne havas tiun kapablon), la probableco por resinkroniĝo kreskas kun la longeco de la kodaĵo. Leviĝas la demando, kiom longa estas mezume la erare legata kodaĵo ĝis la resinkroniĝo. Tiun longecon ni nomu *mezuma nesinkrona vojo* de la kodo.

En la unua ekzemplo de bildo 3 la kodaĵoj de la signoj a, b kaj c resinkroniĝas ĉiam tuj, t. e. ĉe la fino de la signo mem. Nur la signo $\hat{c}$ povas kaŭzi problemojn. Se la signoj havas la oftecojn indikitajn supre de la bildo, la mezuma kodo-longo por unu signo estas $\bar{l} = \frac{7}{4} = 1{,}75$; hazarda eklego do trafas la diversajn lokojn kun la probablecoj indikitaj en tabelo 2. Kun tiuj probablecoj la signoj kontribuas al la mezuma nesinkrona vojo.

Por la signoj a, b kaj c la nesinkrona vojo ĉe la diversaj eklego-punktoj estas rekte kalkulebla, ĉar la legado nepre resinkroniĝas ĉe la fino de la kodvorto. Por $\hat{c}$ tiu vojo dependas de la sekva signo: Se ankaŭ tiu signo estas $\hat{c}$, ni troviĝas en precize la sama situacio; alie restas la nesinkrona vojo de a, b aŭ c, konata el tabelo 2. Nomante $\lambda_{s,\mu}$ la mezuman nesinkronan vojon post eklego ĉe pozicio $\mu \geq 0$ en la signo s, ni povas starigi jenan ekvacion ($|s|$ signifas la longecon de la kodovorto por la signo s):

$$
\begin{aligned}
\lambda_{\hat{c},1} &= p_a\,(|\mathtt{c}| + \lambda_{a,1}) + p_b\,(|\hat{\mathtt{c}}| + \lambda_{b,1}) + p_c\,(|\hat{\mathtt{c}}| + \lambda_{c,1}) + p_{\hat{c}}\,(|\hat{\mathtt{c}}| + \lambda_{\hat{c},1}) && (1)\\
&= \frac{1}{2}\,(3 + \lambda_{a,1}) + \frac{1}{4}\,(3 + \lambda_{b,1}) + \frac{1}{8}\,(3 + \lambda_{c,1}) + \frac{1}{8}\,(3 + \lambda_{\hat{c},1}) && (2)\\
&= \frac{3+0}{2} + \frac{3+1}{4} + \frac{3+2}{8} + \frac{3}{8} + \frac{1}{8}\,\lambda_{\hat{c},1} && (3)\\
&= \frac{7}{2} + \frac{1}{8}\,\lambda_{\hat{c},1} \;=\; 4 && (4)
\end{aligned}
$$

Simile rezultas ke

$$
\begin{aligned}
\lambda_{\hat{c},2} &= \frac{1}{2}\,(2 + \lambda_{a,1}) + \frac{1}{4}\,(3 + \lambda_{b,2}) + \frac{1}{8}\,(3 + \lambda_{c,2}) + \frac{1}{8}\,(3 + \lambda_{\hat{c},2}) && (5)\\
&= \frac{2+0}{2} + \frac{3+0}{4} + \frac{3+1}{8} + \frac{3}{8} + \frac{1}{8}\lambda_{\hat{c},2} && (6)\\
&= \frac{21}{8} \;+\; \frac{1}{8}\,\lambda_{\hat{c},2} \;=\; 3 && (7)
\end{aligned}
$$

Pezigante tiujn mezumajn vojojn por la unuopaj signoj kaj pozicioj per ties oftecoj p_s, ni ricevas

$$
\begin{aligned}
\lambda &= \sum_{s \in S} \sum_{1 \leq \mu < |s|} \frac{p_s}{\bar{l}} \cdot \lambda_{s,\mu}\\
&= \frac{1}{7}\lambda_{b,1} + \frac{1}{14}\lambda_{c,1} + \frac{1}{14}\lambda_{c,2} + \frac{1}{14}\lambda_{\hat{c},1} + \frac{1}{14}\lambda_{\hat{c},2}\\
&= \frac{1}{7} \cdot 1 \;+\; \frac{1}{14}(2 + 1 + 4 + 3)\\
&= \frac{6}{7}
\end{aligned}
$$

La mezuma nesinkrona vojo do estas nur $\frac{6}{7}$, malpli ol unu signo. Tion kaŭzas, ke ofte la hazarda eklego komencas ĉe ĝusta loko, do ĉe signo-limo.

Tabelo 2: Probablecoj por hazarda eklego,
en la unua kodo de bildo 3

signo s	loko μ		probableco $p_s/\bar{l}$	nesinkrona vojo $\lambda_{s,\mu}$
a	0	′0	$\frac{1\cdot 4}{2\cdot 7}$	0
b	0	′10	$\frac{1\cdot 4}{4\cdot 7}$	0
b	1	1′0	$\frac{1\cdot 4}{4\cdot 7}$	1
c	0	′110	$\frac{1\cdot 4}{8\cdot 7}$	0
c	1	1′10	$\frac{1\cdot 4}{8\cdot 7}$	2
c	2	11′0	$\frac{1\cdot 4}{8\cdot 7}$	1
ĉ	0	′111	$\frac{1\cdot 4}{8\cdot 7}$	0
ĉ	1	1′11	$\frac{1\cdot 4}{8\cdot 7}$	4
ĉ	2	11′1	$\frac{1\cdot 4}{8\cdot 7}$	3

La kalkuladon tre simpligis la fakto, ke en la ekzemplo la resinkroniĝo okazas plej ofte ene de unu signo. Nur la signo $\hat{c}$ kaŭzas interdependecon de iuj $\lambda_{s,\mu}$ kiu kondukis al la supraj ekvacioj. En pli ampleksaj kaj komplikaj kodoj, λ estas kalkulebla nur per ampleksa lineara sistemo (komparu kun la ekvacioj (1), (5)):

$$\lambda_{s,\mu} = \sum_{u \in S} \sum_{1 \leq \nu < |u|} (a_{s,\mu,u,\nu} \cdot \lambda_{u,\nu} + b_{s,\mu,u,\nu}) \qquad (8)$$

La valoroj a interligantaj la $\lambda_{s,\mu}$ kaj la konstantoj b dependas de tio, ĉu ekzistas daŭrigo(j) $A \in S^+$ de s tia(j), ke elirante de la μ-a pozicio en $s \in S$ eblas atingi la ν-an pozicion en $u \in S$ per la legado de iu kodvorto $t \in S$:

$$\begin{aligned} A &= a_1 + \cdots + a_k \in S^+, \quad k \geq 1 \\ & \quad a_1 = s, \ a_k = u \\ A &= x + t + y, \quad |x| = \mu, \quad |u| - |y| = \nu \end{aligned}$$

Tio signifas, ke eklegante ĉe la komenco de s oni legas la vicon A, kaj eklegante ĉe la μ-a pozicio de s oni legas la signon t, alvenante ĉe la ν-a pozicio de a_k, la lasta signo de A. Bildo 4 skizas la situacion. Atentu ke A ankaŭ povas konsisti nur el $s = u$, kaj ke por la sama t povas ekzisti pluraj A, diferencaj per siaj a_k.

Se por iuj (s, μ, u, ν) ne ekzistas tia t, la koeficientoj a kaj b nulas por tiuj indicoj. Alie ili havas la sekvajn valorojn:

$$\begin{aligned} a_{s,\mu,u,\nu} &= \sum_{t \in S, A(t)} p(a_2 + \cdots + a_k) \\ b_{s,\mu,u,\nu} &= \sum_{t \in S, A(t)} p(a_2 + \cdots + a_k) \cdot |t| \end{aligned}$$

Tiu lineara sistemo difinas ĉiujn $\lambda_{s,\mu}$. La starigo de la sistemo jam estas kompleksa tasko, por kiu konvenas komputila helpo.

$s = a_1$	$\cdots$	$a_k = u$
x	t	y

Bildo 4: Legaĵoj de kod-vico,
komenciĝantaj ĉe la komenco (nula pozicio) de s, kaj ĉe la μ-a pozicio

Apliko de tiu algoritmo al la dua kodo de bildo 3 rezultigas $\lambda \approx 2,29$. Ne mirigas, ke tiu kodo havas pli longan mezuman nesinkronan vojon ol la unua ekzemplo. Por kalkuli la mezuman nesinkronan vojon de la *Huffman*-kodo ni supozu, ke la relativa ofteco de ĉiu signo respondas al la logaritmo de la longeco de ĝia kodo. Tiam rezultas valoro de 19,35. En la ekzemplo de bildo 2 ni vidis nesinkronajn vojojn de longecoj ĝis 40; la mezumo de la 11 kalkulitaj valoroj estas 17,09.

Por kodoj sen resinkroniga kapablo la lineara sistemo (8) ne estas solvebla; ni metas tiuokaze $\lambda = \infty$.

4 Konkludo

Per la algoritmo skizita laŭ la ekzempla kodo el bildo 3 estas eble kalkuli la mezuman nesinkronan vojon λ de ĉiu kodo. Tiu nombro λ diras, kiom da kodaĵo necesas mezume legi por resinkronigi legadon, kiu komenciĝas ĉe iu arbitra punkto. Por certiĝi ke la legado sinkroniĝis kun la vera mesaĝo, necesas prove eklegi ĉe tiom da sinsekvaj punktoj, kiom estas la maksimuma vortolongo de la kodo. Tie, kie ĉiuj legoprovoj konverĝas, certe okazis resinkroniĝo.

Resinkroniĝo ebligas la partan legadon de kodaĵoj, kiuj parte fariĝis nelegeblaj, ekzemple per fizika difekto de la storilo. Tio tamen validas nur, se la kodado estas nedependa de la antaŭa mesaĝo. Ofte ekzistas tia dependeco, ekzemple ĉe multaj algoritmoj por densigo de datenoj. Tiam aldone al resinkroniĝo necesas teknikoj, kiuj similas kriptologiajn atakojn al ĉifritaj mesaĝoj.

Literaturo

[1] Bauer, F. L., Goos, G.: Informatik 1. Berlin: Springer-Verlag, 1991.

[2] CCITT: Standardization of group 3 facsimile apparatus for document transmission. Rekomendo T.4. Ĝenevo: 1980; amendita en Malaga-Torremolinos: 1984.

[3] CCITT: Facsimile coding schemes and coding control functions for group 4 facsimile apparatus. Rekomendo T.6. Malaga-Torremolinos: 1984.