

A MAGYAR BRAILLE-RÖVIDÍRÁS MEGÚJÍTÁSA FÉLAUTOMATIKUS MÓDSZERREL

Sass Bálint

sass.balint@nytud.mta.hu

MTA Nyelvtudományi Intézet, Budapest

MSZNY2014

Szeged, 2014. január 16-17.

BEVEZETŐ



Braille - írás

- *Braille-rövidírás*

nyelvspecifikus rövidítési, tömörítési szabályok, pl.: $r(\text{hogya})=h$

→ használatával:

- gyorsul az írás-jegyzetelés
- gyorsul az olvasás-felismerés
- kinyomtatva kisebb helyigény

- kezdeményezés az MVGYOSZ Braille-bizottsága részéről:

bővítsük a magyar rövidírást új szabályokkal annak érdekében,
hogy a rövidítési képessége jelentősen növekedjen.

- 1 A MAGYAR BRAILLE-ÍRÁS
- 2 MÓDSZER: KORPUSZVEZÉRELT + FÉLAUTOMATIKUS
- 3 A RÖVIDÍTENDŐ ELEMÉKET MEGADÓ ALGORITMUS
- 4 A RÖVIDÍTÉSJELEK MANUÁLIS MEGVÁLASZTÁSA
- 5 ÚJ MAGYAR RÖVIDÍRÁS: A JAVASLAT
- 6 A JAVASLAT KIÉRTÉKELÉSE
- 7 ÖSSZEFOGLALÁS

- 1 A MAGYAR BRAILLE-ÍRÁS
- 2 MÓDSZER: KORPUSZVEZÉRELT + FÉLAUTOMATIKUS
- 3 A RÖVIDÍTENDŐ ELEMÉKET MEGADÓ ALGORITMUS
- 4 A RÖVIDÍTÉSJELEK MANUÁLIS MEGVÁLASZTÁSA
- 5 ÚJ MAGYAR RÖVIDÍRÁS: A JAVASLAT
- 6 A JAVASLAT KIÉRTÉKELÉSE
- 7 ÖSSZEFOGLALÁS

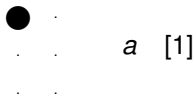
BRAILLE-ÍRÁS

- A vakok által világszerte használt tapintáson alapuló írásmód.
- 6 pont különböző elrendezéseiből → 64 féle karakter

1	4
2	5
3	6

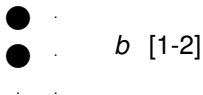
BRAILLE-ÍRÁS

- A vakok által világszerte használt tapintáson alapuló írásmód.
- 6 pont különböző elrendezéseiből → 64 féle karakter



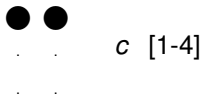
BRAILLE-ÍRÁS

- A vakok által világszerte használt tapintáson alapuló írásmód.
- 6 pont különböző elrendezéseiből → 64 féle karakter



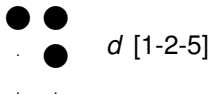
BRAILLE-ÍRÁS

- A vakok által világszerte használt tapintáson alapuló írásmód.
- 6 pont különböző elrendezéseiből → 64 féle karakter



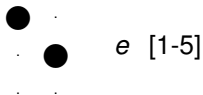
BRAILLE-ÍRÁS

- A vakok által világszerte használt tapintáson alapuló írásmód.
- 6 pont különböző elrendezéseiből → 64 féle karakter



BRAILLE-ÍRÁS

- A vakok által világszerte használt tapintáson alapuló írásmód.
- 6 pont különböző elrendezéseiből → 64 féle karakter



A MAGYAR BRAILLE-ÍRÁS

- önálló jelek az *ékezetes* magánhangzóknak

⠁	⠃	⠅	⠇	⠉	⠋	⠌	⠎	⠐	⠒	⠔	⠖	⠘	⠚	⠜
a	á	e	é	i	í	o	ó	ő	ó	u	ú	ü	ű	

- önálló jelek a *kettős* mássalhangzóknak

⠉	⠏	⠒	⠔	⠖	⠘	⠎	⠌	⠐	⠒	⠔	⠖	⠘	⠚	⠜
c	cs	g	gy	l	ly	n	ny	s	sz	t	ty	z	zs	

- „tükrözés” – kivéve: *e, i, c, g*
- erős jel → önmagában azonosítható

- ha Braille-írást sima szöveggént ábrázolunk,
minden jelet egy karakterrel jelölünk,
a kettős mássalhangzókat az első karakter nagybetűs
változatával: C, G, L, N, S, T, Z – pl. *maGar*

A MAGYAR BRAILLE-RÖVIDÍRÁS

Magyarországon használatos sztenderd rövidírás: ún. „kis” rövidírás

Elemi:	rk()
➊ nagybetűjel :: törlése	2,3%
➋ vessző utáni szóköz törlése	1,4%
➌ határozott névelő utáni szóköz törlése, az rövidítése	1,9%
➍ 44 szabály:	
<ul style="list-style-type: none"> 7 db szóvégi rövidítés: ban/ben → b, ból/ból → bl, hoz/hez/höz → hz ... 	1,4%
<ul style="list-style-type: none"> 16 db egyjelű szórövidítés: Cak → C, de → d, és → é ... 	1,8%
<ul style="list-style-type: none"> 21 db kétjelű szórövidítés: aNNi → ai, boldog → bg, gond → gd ... 	1,1%

→ a kis rövidírás mért rövidítési képessége: **9,9%**

A MAGYAR BRAILLE-RÖVIDÍRÁS

Magyarországon használatos sztenderd rövidírás: ún. „kis” rövidírás

Elemi:	rk()
➊ nagybetűjel :: törlése	2,3%
➋ vessző utáni szóköz törlése	1,4%
➌ határozott névelő utáni szóköz törlése, az rövidítése	1,9%
➍ 44 szabály:	
<ul style="list-style-type: none"> 7 db szóvégi rövidítés: ban/ben → b, ból/ból → bl, hoz/hez/höz → hz ... 	0,20% 1,4%
<ul style="list-style-type: none"> 16 db egyjelű szórövidítés: Cak → C, de → d, és → é ... 	0,12% 1,8%
<ul style="list-style-type: none"> 21 db kétjelű szórövidítés: aNNi → ai, boldog → bg, gond → gd ... 	0,05% 1,1%
	szabálynként

→ a kis rövidírás mért rövidítési képessége: **9,9%**

KETTŐS CÉL

Bővítsük a kis rövidírást új szabályokkal annak érdekében, hogy a rövidítési képessége jelentősen növekedjen.

Valamint. . .

Feleljen meg az alábbi használhatósági követelményeknek:

- az ismert kis rövidírást egészítse ki;
- *jó olvashatóság*: a rövidítések emlékeztessenek az eredetire;
- *jó felismerhetőség*: tapintás útján könnyen felismerhető jelek;
- *könnyű megtanulhatóság*: kevés, egyszerű szabály.

Kompromisszumos javaslat, mely. . .

- törekszik a maximális rövidítési képesség elérésére,
- miközben megfelel a használhatósági feltételeknek is.

- 1 A MAGYAR BRAILLE-ÍRÁS
- 2 **MÓDSZER: KORPUSZVEZÉRELT + FÉLAUTOMATIKUS**
- 3 A RÖVIDÍTENDŐ ELEMÉKET MEGADÓ ALGORITMUS
- 4 A RÖVIDÍTÉSJELEK MANUÁLIS MEGVÁLASZTÁSA
- 5 ÚJ MAGYAR RÖVIDÍRÁS: A JAVASLAT
- 6 A JAVASLAT KIÉRTÉKELÉSE
- 7 ÖSSZEFOGLALÁS

MÓDSZER: KORPUSZVEZÉRELT + FÉLAUTOMATIKUS

Mit rövidítsünk?

Alapötlet: a maximális rövidítési képességgel bíró ideális rövidírás a magyar nyelv korpuszgyakorisági adatai alapján korpuszvezérelt módon, automatikusan kialakítható.

Elv: a lehető leggyakoribb betűkapcsolatokat kell a lehető legrövidebbre rövidíteni.

Pl.: $ty \leftrightarrow et \sim 20\times$ -os különbség a gyakoriságban

→ automatikus algoritmus: *iteratív, mohó* eljárás

Mire rövidítsünk?

Szakértői döntések a gyakorisági adatok és a használhatósági megfontolások alapján. Akár az automatikus eredményt felülbíráva.

→ manuális munka

- 1 A MAGYAR BRAILLE-ÍRÁS
- 2 MÓDSZER: KORPUSZVEZÉRELT + FÉLAUTOMATIKUS
- 3 A RÖVIDÍTENDŐ ELEMÉKET MEGADÓ ALGORITMUS**
- 4 A RÖVIDÍTÉSJELEK MANUÁLIS MEGVÁLASZTÁSA
- 5 ÚJ MAGYAR RÖVIDÍRÁS: A JAVASLAT
- 6 A JAVASLAT KIÉRTÉKELÉSE
- 7 ÖSSZEFOGLALÁS

AZ ALGORITMUS MŰKÖDÉSE

- 1 Korpusz alapján: szavak gyakorisági listája.
- 2 Számba vesszük az összes elvben rövidíthető nyelvi elemet: szavak gyaklistája → karakter- n -gramok gyaklistája.
- 3 Kiszámoljuk az elemek rövidítési képességét a következő módon:
$$rk(w, r(w)) = [l(w) - l(r(w))] \cdot fq(w)$$
 w az eredeti rövidítendő karaktersorozat, $r(w)$ a rövidítés, $l()$ a hossz (karakterszám), $fq()$ a gyakoriság.
Pl.: a *sem* és a *maGar* gyakorisága nagyjából azonos
→ érdekesebb az utóbbit két karakterre rövidíteni.
- 4 Rendezzük a gyakorisági listát rövidítési képesség szerint.
- 5 A rendezett lista *első* bejegyzése adja a maximális rövidítési képességet, vesszük ezt az elemet, és hozzárendelünk egy megfelelő rövidítést.
- 6 Az így kialakított rövidítő szabályt *alkalmazzuk a szavak gyakorisági listájára.*
- 7 → 2

MŰKÖDÉSI PÉLDA: AZ ÚJRARENDEZÉS SZEREPE

2 karakterre rövidítünk

A gyakorisági lista eleje:

w	$\text{fq}(w)$	$l(w) - l(r(w))$	$\text{rk}(w)$
<i>Ser</i>	10901	$\times 1 =$	10901
<i>Serint</i>	2515	$\times 4 =$	10060
<i>maGar</i>	3326	$\times 3 =$	9978

Az első elemet rövidítjük: *Ser* \rightarrow *Sr*

MŰKÖDÉSI PÉLDA: AZ ÚJRARENDEZÉS SZEREPE

2 karakterre rövidítünk

A gyakorisági lista eleje:

w	$\text{fq}(w)$	$l(w) - l(r(w))$	$\text{rk}(w)$
<i>Ser</i>	10901	$\times 1 =$	10901
<i>Serint</i>	2515	$\times 4 =$	10060
<i>maGar</i>	3326	$\times 3 =$	9978

Az első elemet rövidítjük: *Ser* \rightarrow *Sr*

w	$\text{fq}(w)$	$l(w) - l(r(w))$	$\text{rk}(w)$
<i>maGar</i>	3326	$\times 3 =$	9978
...			
<i>Srint</i>	2515	$\times 3 =$	7545
...			
<i>Sr</i>	10901	$\times 0 =$	0

MŰKÖDÉSI PÉLDA: AZ ÚJRARENDEZÉS SZEREPE

2 karakterre rövidítünk

A gyakorisági lista eleje:

w	$\text{fq}(w)$	$l(w) - l(r(w))$	$\text{rk}(w)$
<i>Ser</i>	10901	$\times 1 =$	10901
<i>Serint</i>	2515	$\times 4 =$	10060
<i>maGar</i>	3326	$\times 3 =$	9978

A következő rövidítendő elem a *maGar* lesz.

Az első elemet rövidítjük: *Ser* \rightarrow *Sr*

w	$\text{fq}(w)$	$l(w) - l(r(w))$	$\text{rk}(w)$
<i>maGar</i>	3326	$\times 3 =$	9978
...			
<i>Srint</i>	2515	$\times 3 =$	7545
...			
<i>Sr</i>	10901	$\times 0 =$	0

MŰKÖDÉSI PÉLDA: AZ ÚJRARENDEZÉS SZEREPE

2 karakterre rövidítünk

A gyakorisági lista eleje:

w	$\text{fq}(w)$	$l(w) - l(r(w))$	$\text{rk}(w)$
<i>Ser</i>	10901	$\times 1 =$	10901
<i>Serint</i>	2515	$\times 4 =$	10060
<i>maGar</i>	3326	$\times 3 =$	9978

Az első elemet rövidítjük: *Ser* \rightarrow *Sr*

w	$\text{fq}(w)$	$l(w) - l(r(w))$	$\text{rk}(w)$
<i>maGar</i>	3326	$\times 3 =$	9978
...			
<i>Srint</i>	2515	$\times 3 =$	7545
...			
<i>Sr</i>	10901	$\times 0 =$	0

A következő rövidítendő elem a *maGar* lesz.

A *Sr* kétjelű rövidítéssel nyilván nem rövidíthető tovább.

MŰKÖDÉSI PÉLDA: AZ ÚJRARENDEZÉS SZEREPE

2 karakterre rövidítünk

A gyakorisági lista eleje:

w	$\text{fq}(w)$	$l(w) - l(r(w))$	$\text{rk}(w)$
<i>Ser</i>	10901	$\times 1 =$	10901
<i>Serint</i>	2515	$\times 4 =$	10060
<i>maGar</i>	3326	$\times 3 =$	9978

Az első elemet rövidítjük: *Ser* \rightarrow *Sr*

w	$\text{fq}(w)$	$l(w) - l(r(w))$	$\text{rk}(w)$
<i>maGar</i>	3326	$\times 3 =$	9978
...			
<i>Srint</i>	2515	$\times 3 =$	7545
...			
<i>Sr</i>	10901	$\times 0 =$	0

A következő rövidítendő elem a *maGar* lesz.

A *Serint* az új szabály miatt sokkal lejjebb csúszott. Ha mégis sorra kerül, továbbrövidíthetjük.

A *Sr* kétjelű rövidítéssel nyilván nem rövidíthető tovább.

MŰKÖDÉSI PÉLDA: AZ ÚJRARENDEZÉS SZEREPE

2 karakterre rövidítünk

A gyakorisági lista eleje:

w	$\text{fq}(w)$	$l(w) - l(r(w))$	$\text{rk}(w)$
<i>Ser</i>	10901	$\times 1 =$	10901
<i>Serint</i>	2515	$\times 4 =$	10060
<i>maGar</i>	3326	$\times 3 =$	9978

Az első elemet rövidítjük: *Ser* \rightarrow *Sr*

w	$\text{fq}(w)$	$l(w) - l(r(w))$	$\text{rk}(w)$
<i>maGar</i>	3326	$\times 3 =$	9978
...			
<i>Srint</i>	2515	$\times 3 =$	7545
...			
<i>Sr</i>	10901	$\times 0 =$	0

A következő rövidítendő elem a *maGar* lesz.

A *Serint* az új szabály miatt sokkal lejjebb csúszott. Ha mégis sorra kerül, továbbrövidíthetjük. Így is lesz: $r(\text{Srint}) = \text{St}$

A *Sr* kétjelű rövidítéssel nyilván nem rövidíthető tovább.

MŰKÖDÉSI PÉLDA: AZ ÚJRARENDEZÉS SZEREPE

2 karakterre rövidítünk

A gyakorisági lista eleje:

w	$\text{fq}(w)$	$l(w) - l(r(w))$	$\text{rk}(w)$
<i>Ser</i>	10901	$\times 1 =$	10901
<i>Serint</i>	2515	$\times 4 =$	10060
<i>maGar</i>	3326	$\times 3 =$	9978

Az első elemet rövidítjük: *Ser* \rightarrow *Sr*

w	$\text{fq}(w)$	$l(w) - l(r(w))$	$\text{rk}(w)$
<i>maGar</i>	3326	$\times 3 =$	9978
...			
<i>Srint</i>	2515	$\times 3 =$	7545
...			
<i>Sr</i>	10901	$\times 0 =$	0

A következő rövidítendő elem a *maGar* lesz.

A *Serint* az új szabály miatt sokkal lejjebb csúszott. Ha mégis sorra kerül, továbbrövidíthetjük.

Így is lesz: $r(\text{Srint}) = \text{St}$

Kifejtve: $r(\text{Serint}) = \text{St}$

A *Sr* kétjelű rövidítéssel nyilván nem rövidíthető tovább.

- 1 A MAGYAR BRAILLE-ÍRÁS
- 2 MÓDSZER: KORPUSZVEZÉRELT + FÉLAUTOMATIKUS
- 3 A RÖVIDÍTENDŐ ELEMÉKET MEGADÓ ALGORITMUS
- 4 A RÖVIDÍTÉSJELEK MANUÁLIS MEGVÁLASZTÁSA**
- 5 ÚJ MAGYAR RÖVIDÍRÁS: A JAVASLAT
- 6 A JAVASLAT KIÉRTÉKELÉSE
- 7 ÖSSZEFOGLALÁS

HASZNÁLHATÓSÁGI KÖVETELMÉNYEK

1 jó olvashatóság:

a rövidítések emlékeztessenek az eredetire

- ideális: szó kezdő és záró mássalhangzója
r(*rövid*)=*rd* r(*boldog*)=*bg*
- kiegészítő lehetőség: a jel *kinézete* emlékeztet a jelentésre

Elv: „értelmest értelmesre”

= (lehetőleg) értelmezhető szóelemeket – morfémákat – rövidítünk,
és olvasható rövidítésjelet rendelünk hozzájuk

2 jó felismerhetőség:

tapintás útján könnyen felismerhető jelek alkalmazása

→ törekszünk rá, hogy minden rövidítés tartalmazzon erős jelet

3 könnyű megtanulhatóság:

kevés, egyszerű szabály

→ a kis rövidírásban lévőkhöz hasonló szabályok

- 1 A MAGYAR BRAILLE-ÍRÁS
- 2 MÓDSZER: KORPUSZVEZÉRELT + FÉLAUTOMATIKUS
- 3 A RÖVIDÍTENDŐ ELEMÉKET MEGADÓ ALGORITMUS
- 4 A RÖVIDÍTÉSJELEK MANUÁLIS MEGVÁLASZTÁSA
- 5 ÚJ MAGYAR RÖVIDÍRÁS: A JAVASLAT**
- 6 A JAVASLAT KIÉRTÉKELÉSE
- 7 ÖSSZEFOGLALÁS

A JAVASLAT 1/4

	rövidítendő	rövidítés	megjegyzés
1.	⠠⠍⠑⠎	⠠⠍⠎	szó elején (98%)
2.	⠠⠑⠞⠞	⠠⠑⠞	= ott ⠠~⠞
3.	⠠⠎⠑⠗	⠠⠎⠗	
4.	⠠⠑⠞⠞	⠠⠑⠞	= ett ⠠~⠞
5.	⠠⠍⠁⠎⠗	⠠⠍⠎	
6.	⠠⠎⠑⠞	⠠⠎	
7.	⠠⠎⠑⠎	⠠⠎	= ság
8.	⠠⠞⠑⠞	⠠⠞	szó elején (97%) <i>kompl!</i>

A JAVASLAT 2/4

	rövidítendő	rövidítés	megjegyzés
9.	⠠⠠⠠⠠⠠[vezet]	⠠⠠[vz]	
10.	⠠⠠⠠[nek]	⠠⠠[nx]	= nak ⠠~⠠ szó végén (83%)
11.	⠠⠠⠠[köz]	⠠⠠[kz]	
12.	⠠⠠⠠[nak]	⠠⠠[nx]	= nek ⠠~⠠ szó végén (96%)
13.	⠠⠠⠠[fel]	⠠⠠[fl]	szó elején (86%)
14.	⠠⠠⠠⠠⠠⠠[Serint]	⠠⠠[St]	szó elején (95%)
15.	⠠⠠⠠⠠[rend]	⠠⠠[rH]	csúsztatott <i>d</i> (<i>rd</i> foglalt, <i>rn</i> gyak)
16.	⠠⠠⠠⠠⠠[ember]	⠠⠠[wr]	jel kinézete

A JAVASLAT 2/4

	rövidítendő	rövidítés	megjegyzés
9.	⠠⠠⠠⠠⠠⠠[vezet]	⠠⠠⠠[vz]	
10.	⠠⠠⠠⠠[nek]	⠠⠠⠠[nx]	= nak ⠠⠠~⠠ szó végén (83%)
11.	⠠⠠⠠⠠[köz]	⠠⠠⠠[kz]	
12.	⠠⠠⠠⠠[nak]	⠠⠠⠠[nx]	= nek ⠠⠠~⠠ szó végén (96%)
13.	⠠⠠⠠⠠[fel]	⠠⠠⠠[fl]	szó elején (86%)
14.	⠠⠠⠠⠠⠠⠠⠠[Serint]	⠠⠠⠠[St]	szó elején (95%)
15.	⠠⠠⠠⠠⠠[rend]	⠠⠠⠠[rH]	csúsztatott d (<i>rd</i> foglalt, <i>rn</i> gyak)
16.	⠠⠠⠠⠠⠠⠠[ember]	⠠⠠⠠[wr]	jel kinézete

A JAVASLAT 3/4

	rövidítendő	rövidítés	megjegyzés
17.	⠠⠠⠠⠠⠠⠠⠠[kormán]	⠠⠠[kN]	
18.	⠠⠠⠠⠠⠠[ellen]	⠠⠠[oó]	jel kinézete
19.	⠠⠠⠠⠠⠠[elnök]	⠠⠠[eö]	
20.	⠠⠠⠠⠠[elő]	⠠⠠[eő]	
21.	⠠⠠⠠⠠⠠[tart]	⠠⠠[tt]	szó elején (78%) <i>kompl!</i>
22.	⠠⠠⠠⠠[áll]	⠠⠠[LI]	jel kinézete
23.	⠠⠠⠠⠠⠠[mond]	⠠⠠[mH]	csúsztatott <i>d</i> (<i>md</i> foglalt)
24.	⠠⠠⠠⠠⠠⠠⠠[támogat]	⠠⠠[tg]	

- 1 A MAGYAR BRAILLE-ÍRÁS
- 2 MÓDSZER: KORPUSZVEZÉRELT + FÉLAUTOMATIKUS
- 3 A RÖVIDÍTENDŐ ELEMÉKET MEGADÓ ALGORITMUS
- 4 A RÖVIDÍTÉSJELEK MANUÁLIS MEGVÁLASZTÁSA
- 5 ÚJ MAGYAR RÖVIDÍRÁS: A JAVASLAT
- 6 A JAVASLAT KIÉRTÉKELÉSE**
- 7 ÖSSZEFOGLALÁS

A JAVASLAT KIÉRTÉKELÉSE

3 tesztfájl:

megnevezés	méret	
egy zenei híreket tartalmazó fájl az MVGYOSZ-ból	11000	szó
a <i>Vakok Világa</i> 31 számának anyaga	180000	szó
Mikszáth Kálmán: <i>Szent Péter esernyője</i> c. regénye	53000	szó

Mérés módszere:

karaktorszám „egy jel egy karakter” alakban + átlagolás

EREDMÉNY

tesztanyag	kis rövidírás rk()	új rövidírás rk()	Δ rk()
zenei hírek	9,5%	12,5%	+3,0%
Vakok Világa	9,5%	13,9%	+4,4%
Szent Péter...	10,7%	13,5%	+2,8%
átlag	9,9%	13,3%	+3,4%

Kevés szabály, jelentős növekedés.

A rövidítési képesség: **szabályonként 0,1%**

EREDMÉNY

tesztanyag	kis rövidírás rk()	új rövidírás rk()	Δ rk()
zenei hírek	9,5%	12,5%	+3,0%
Vakok Világa	9,5%	13,9%	+4,4%
Szent Péter. . .	10,7%	13,5%	+2,8%
átlag	9,9%	13,3%	+3,4%

Kevés szabály, jelentős növekedés.

A rövidítési képesség: **szabályonként 0,1%**

Új magyar rövidírás: 13,3% \leftrightarrow Unified English Braille: \approx 20%

Az angol rendszer kb. 200 szabályt tartalmaz,
számos nehezen olvasható/tanulható
rövidítési lehetőséget kiterjedten alkalmaz.

- 1 A MAGYAR BRAILLE-ÍRÁS
- 2 MÓDSZER: KORPUSZVEZÉRELT + FÉLAUTOMATIKUS
- 3 A RÖVIDÍTENDŐ ELEMÉKET MEGADÓ ALGORITMUS
- 4 A RÖVIDÍTÉSJELEK MANUÁLIS MEGVÁLASZTÁSA
- 5 ÚJ MAGYAR RÖVIDÍRÁS: A JAVASLAT
- 6 A JAVASLAT KIÉRTÉKELÉSE
- 7 ÖSSZEFOGLALÁS**

ÖSSZEFOGLALÁS

A közel maximális rövidítési képességgel bíró, ugyanakkor kényelmesen használható új magyar rövidírás szabálykészlete félautomatikus módon, a rövidítendők korpuszvezérelt meghatározásával és a rövidítések kézi finomításával előállítható.

Összesített rövidítési képesség: $13,3\% \ll$ elvi maximum.

A bevezetés megkönnyítése érdekében a változtatás mértékével óvatosak voltunk, nagy hangsúlyt fektettünk a használhatóságra.

A szabályrendszer a jövőben könnyen bővíthető.

KONKLÚZIÓ

(két korábbi dia. . .)

A MAGYAR BRAILLE-RÖVIDÍRÁS

Magyarországon használatos sztenderd rövidírás: ún. „kis” rövidírás

Elemi:	rk()
➊ nagybetűjel :: törlése	2,3%
➋ vessző utáni szóköz törlése	1,4%
➌ határozott névelő utáni szóköz törlése, az rövidítése	1,9%
➍ 44 szabály:	
<ul style="list-style-type: none"> 7 db szóvégi rövidítés: ban/ben → b, ból/ból → bl, hoz/hez/höz → hz ... 	0,20% 1,4%
<ul style="list-style-type: none"> 16 db egyjelű szórövidítés: Cak → C, de → d, és → é ... 	0,12% 1,8%
<ul style="list-style-type: none"> 21 db kétjelű szórövidítés: aNNi → ai, boldog → bg, gond → gd ... 	0,05% 1,1%
	szabálynként

→ a kis rövidírás mért rövidítési képessége: **9,9%**

EREDMÉNY

tesztanyag	kis rövidírás rk()	új rövidírás rk()	Δ rk()
zenei hírek	9,5%	12,5%	+3,0%
Vakok Világa	9,5%	13,9%	+4,4%
Szent Péter...	10,7%	13,5%	+2,8%
átlag	9,9%	13,3%	+3,4%

Kevés szabály, jelentős növekedés.

A rövidítési képesség: **szabályonként 0,1%**

KONKLÚZIÓ

Kis rövidírás kétjelű rövidítések $rk() = 0,05\%$ szabályonként,
új kétjelű rövidítések $rk() = 0,10\%$ szabályonként.

KONKLÚZIÓ

Kis rövidírás kétjelű rövidítések $rk() = 0,05\%$ szabályonként,
új kétjelű rövidítések $rk() = 0,10\%$ szabályonként.

⇒

A korpuszvezérelt módon létrehozott rendszer még úgy is *kétszeres* teljesítményre képes az intuíció illetve hagyomány talaján álló rendszerrel szemben, hogy már eleve jelentősen rövidített szövegen kell dolgoznia.

KONKLÚZIÓ

Kis rövidírás kétjelű rövidítések $rk() = 0,05\%$ szabályonként,
új kétjelű rövidítések $rk() = 0,10\%$ szabályonként.

⇒

A korpuszvezérelt módon létrehozott rendszer még úgy is *kétszeres* teljesítményre képes az intuíció illetve hagyomány talaján álló rendszerrel szemben, hogy már eleve jelentősen rövidített szövegen kell dolgoznia.

⇒

Ha valamit meglévő (gyakorisági) adatokból automatikusan származtatni tudunk, akkor nem érdemes intuitív megközelítést alkalmazni.

KONKLÚZIÓ

Kis rövidírás kétjelű rövidítések rk() = 0,05% szabályonként,
 új kétjelű rövidítések rk() = 0,10% szabályonként.

⇒

A korpuszvezérelt módon létrehozott rendszer még úgy is *kétszeres* teljesítményre képes az intuíció illetve hagyomány talaján álló rendszerrel szemben, hogy már eleve jelentősen rövidített szövegen kell dolgoznia.

⇒

Ha valamit meglévő (gyakorisági) adatokból automatikusan származtatni tudunk, akkor nem érdemes intuitív megközelítést alkalmazni.



ö = ö