

IGEI VONZATKERETEK AZ MNSZ TAGMONDATAIBAN

Sass Bálint

joker@nytud.hu

MTA Nyelvtudományi Intézet, Nyelvtechnológiai Osztály
PPKE, Információs Technológiai Kar, MMT Doktori Iskola

MSZNY2006

Szeged, 2006. december 7–8.

- 1 BEVEZETÉS
- 2 TAGMONDATRA BONTÁS
- 3 FELÜLET
- 4 FEJLESZTÉSI LEHETŐSÉGEK, ALKALMAZÁS

- 1 **BEVEZETÉS**
- 2 TAGMONDATRA BONTÁS
- 3 FELÜLET
- 4 FEJLESZTÉSI LEHETŐSÉGEK, ALKALMAZÁS

BEVEZETÉS

- *Cél:* a magyar igei vonzatkeret-rendszer feltérképezése
- korpusznyelvészeti eszköztár
- előzmények
 - ismert keretek azonosítása egyszerű mondatokban
 - új keretek keresése, azonosítása:
igék és a velük egy tagmondatban lévő NP-k statisztikai feldolgozása + *idiómasági mérés*
- most
 - tagmondatra bontás
 - vonzatkeretek online prezentálása, lekérdezése

ÚJ FELDOLGOZÁSI LÉPÉSEK

- főnévi igenév – mindig *kerethordozó*
- igekötőragasztás (*heurisztika*)
- egyértelműsítetlen igetövek részleges kezelése (*heurisztika*)
- képzők levágása: *-hAt*

- 1 BEVEZETÉS
- 2 TAGMONDATRA BONTÁS**
- 3 FELÜLET
- 4 FEJLESZTÉSI LEHETŐSÉGEK, ALKALMAZÁS

MOTIVÁCIÓ

Szükségesek: *egyvonzatkeretes* egységek

- egyszerű mondatok

MOTIVÁCIÓ

Szükségesek: *egyvonzatkeretes* egységek

- egyszerű mondatok – „nyilvánvalóan nem reprezentálják a valós nyelvhasználatot” (?)

MOTIVÁCIÓ

Szükségesek: egyvonzatkeretes egységek

- egyszerű mondatok – „nyilvánvalóan nem reprezentálják a valós nyelvhasználatot” (?)
- helyettük: *tagmondatok*

MOTIVÁCIÓ

Szükségesek: egyvonzatkeretes egységek

- egyszerű mondatok – „nyilvánvalóan nem reprezentálják a valós nyelvhasználatot” (?)
- helyettük: *tagmondatok*
→ tetszőleges szöveg feldolgozhatóvá válik

KORÁBBI EREDMÉNYEK

- magyar nyelvre alkalmazható, részletesen ismertetett, reprodukálható tagmondatra bontó módszerről nem tudok
- két kézirat
 - szabályalapú módszer
 - tagmondatok és koordináció

ÁLTALÁNOS ELV

A ragozott (finit) ige vonzatai az igét tartalmazó tagmondaton belül vannak.

KORÁBBI EREDMÉNYEK

- magyar nyelvre alkalmazható, részletesen ismertetett, reprodukálható tagmondatra bontó módszerről nem tudok
- két kézirat
 - szabályalapú módszer
 - tagmondatok és koordináció

ÁLTALÁNOS ELV

A ragozott (finit) ige vonzatai az ige tartalmazó tagmondaton belül vannak.

- ragozott ige → ragozott ige vagy főnévi igenév

SZABÁLYALAPÚ MÓDSZER

- különféle reguláris szabályok – kötőszavakra és központosásra építve bizonyos pozíciókon tagmondathatárt helyez el
- *végül*: lehetséges tagmondathatárként megjelöli az összes kötőszót, mely két olyan ige között helyezkedik el, melyek között még nincs tagmondathatár

TAGMONDATOK ÉS KOORDINÁCIÓ

- a szöveg *kötőszavairól* dönt:
szerkezeteket koordinálnak vagy tagmondatokat kötnek össze
- eljárás
 - az első előtti és az utolsó utáni kötőszó → koordináció
 - két ragozott ige között *egyetlen* kötőszó → TMH
 - majd: a kötőszó melletti két frázis összehasonlítása, ha „egyeznek” → koordináció; ha „különbözőek” → TMH
- problémák
 - névszói állítmány
 - kötőszó nélküli tagmondatok

A MÓDSZER

(1/2)

A fent ismertetett első módszerre építék.

- 1 [,-] @ [kötőszó|határozószó]? [vonatkozó névmás]
- 2 [,-] @ [bármilyenNP|AdjP]?
[pedig|akár|azonban|viszont|ellenben|mihelyt|tehát|ugyanis]
- 3 [,-] @ [határozószó]? [nehogy|mintha]
- 4 [,] @ [kötőszó, kivéve: de|illetve|illetőleg|mintegy]
- 5 [,-] @ [múlt idejű, egyes szám harmadik személyű ige]
- 6 [,] @ [kérdőszó]
- 7 [,] @ [az szótóként]? [határozói igenév]

A MÓDSZER

(2/2)

- a *kettőspont* és a *pontosvessző* mindig tagmondathatár
- *meg* kötőszó és *ki* kérdőszó
→ az ezekre épülő szabályokat elhagytam
- frázisok helyett: 1, 2, 3 tetszőleges szó
- *csak* az ige vonzatai vannak az ige tagmondatában
→ ige-koordináció nincs, két ige között mindig van TMH.
→ ha két ige között *egyetlen* kötőszó van: TMH
- kiegészítés
 - *központosítás* ugyanolyan jó, mint a kötőszó
 - ha több van, akkor az utolsót választom (*heurisztika*)

KIÉRTÉKELÉS

- Tesztkorpusz: 200 mondat
- „Útmutató”
 - 1 Jelöljük be a szövegben a tagmondatokat.
 - 2 Minden ige külön tagmondatba kerüljön.
 - 3 A tagmondatvégi központosítás minden esetben a megelőző tagmondathoz tartozzon.
- 171 TMH: 148-at megtalált, 23-at kihagyott, 29 téves

pontosság = **83,6%** lefedés = **86,5%**

- bonyolult jogi szöveg,
sok helyen rossz a régi mondatra bontás
- tipikus hibahelyek
 - nincs kötőszó
 - névszói állítmány


- 1 BEVEZETÉS
- 2 TAGMONDATRA BONTÁS
- 3 FELÜLET**
- 4 FEJLESZTÉSI LEHETŐSÉGEK, ALKALMAZÁS

„MAZSOLA”

<http://corpus.nytud.hu/mazsola>

Közzétett korpuszok:

- 3-10 szavas mondatok (8mió szó)
- Magyar Nemzet (12mió szó)
- Index fórum (18mió szó)



Korpusz:

Igető:

Nem: Eset/névutó: Nem: Lemma:

Nem: Eset/névutó: Nem: Lemma:

Nem: String:

Igető-stat:

Lemma-stat:

Lemma-stat:

Nincs stat:

Készítette, javaslat, megjegyzés: [Sass Bálint](#)
[MTA Nyelvtudományi Intézet](#), 2006.

„MAZSOLA”

<http://corpus.nytud.hu/mazsola>

Közzétett korpuszok:

- 3-10 szavas mondatok (8mió szó)
- Magyar Nemzet (12mió szó)
- Index fórum (18mió szó)



Korpusz:



Igető:

Nem: Eset/névutó:

Nem: Lemma:

Nem: Eset/névutó:

Nem: Lemma:

Nem: String:

Igető-stat:

Lemma-stat:

Lemma-stat:

Nincs stat:

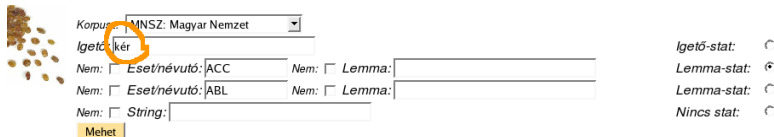
Készítette, javaslat, megjegyzés: [Sass Bálint](#)
[MTA Nyelvtudományi Intézet, 2006.](#)

„MAZSOLA”

<http://corpus.nytud.hu/mazsola>

Közzétett korpuszok:

- 3-10 szavas mondatok (8mió szó)
- Magyar Nemzet (12mió szó)
- Index fórum (18mió szó)



Korpus:

Igető:kér

Nem: Eset/névutó: ACC Nem: Lemma:

Nem: Eset/névutó: ABL Nem: Lemma:

Nem: String:

Igető-stat:

Lemma-stat:

Lemma-stat:

Nincs stat:

Készítette, javaslat, megjegyzés: [Sass Bálint](#)
[MTA Nyelvtudományi Intézet, 2006.](#)

„MAZSOLA”

<http://corpus.nytud.hu/mazsola>

Közzétett korpuszok:

- 3-10 szavas mondatok (8mió szó)
- Magyar Nemzet (12mió szó)
- Index fórum (18mió szó)



Korpusz:

Igető:

Nem: Eset/névu:

Nem: Lemma:

Nem: Eset/névu:

Nem: Lemma:

Nem: String:

Igető-stat:

Lemma-stat:

Lemma-stat:

Nincs stat:

Készítette, javaslat, megjegyzés: [Sass Bálint](#)
[MTA Nyelvtudományi Intézet, 2006.](#)

„MAZSOLA”

<http://corpus.nytud.hu/mazsola>

Közzétett korpuszok:

- 3-10 szavas mondatok (8mió szó)
- Magyar Nemzet (12mió szó)
- Index fórum (18mió szó)



Korpusz:

Igétő:

Nem: Eset/névu:

Nem: Lemma:

Nem: Eset/névu:

Nem: Lemma:

Nem: String:

Igétő-stat: 

Lemma-stat:

Lemma-stat:

Nincs stat:

Készítette, javaslat, megjegyzés: [Sass Bálint](#)
[MTA Nyelvtudományi Intézet, 2006.](#)

„MAZSOLA”

<http://corpus.nytud.hu/mazsola>

Közzétett korpuszok:

- 3-10 szavas mondatok (8mió szó)
- Magyar Nemzet (12mió szó)
- Index fórum (18mió szó)

Korpusz:

Igető:

Nem: Eset/névutó: Nem: Lemma:

Nem: Eset/névutó: Nem: Lemma:

Nem: String:

Igető-stat:

Lemma-stat:

Lemma-stat:

Nincs stat:

Készítette, javaslat, megjegyzés: [Sass Bálint](#)
[MTA Nyelvtudományi Intézet, 2006.](#)

mazsola - Mozilla

File Edit View Go Bookmarks Tools Window Help

Back Forward Reload Stop http://camel.cgi-bin/mazsola/mazsola.pl

Home Bookmarks szótár Norvég Zene EEC ProPapa nyitüd PhD MOKK SZTAKI DYNA Programozás LaTeX Utazás AzBest Olvas

mazsola mazsola mazsola mazsola

Körpuz:

Igétő:

Nem: Eset/hévívót: ACC Nem: Lemma:

Nem: Eset/hévívót: ABL Nem: Lemma:

Nem: String:

Igétő-stat:

Lemma-stat:

Lemma-stat:

Nincs stat:

1010 találat. Stat: [bocsánat](#) [51] [segítség](#) [53] [elnézés](#) [32] [az](#) [136] [engedély](#) [32] [tájékoztatás](#) [21] [támogatás](#) [25] [péruz](#) [20] [felmentés](#) [12] [válasz](#) [16] [tanács](#) [13]

[forint](#) [16] [magyarokat](#) [9] [igazolás](#) [8] [állásfoglalás](#) [8] [kiadás](#) [7] [normakontrol](#) [6] [információ](#) [9] [teféigazsáts](#) [7] [ksháptat](#) [6] [megőképzés](#) [7] [símek](#) [6] [tanács](#) [6] [tekévizsgálat](#) [6] [sz](#) [12] [szólat](#) [7] [vádkam](#) [6] [az](#) [7] [6]

adat

Az információk ellenőrzésére persze a körháztól kér adatokat. Feladatai teljesítéséhez adatokat kérhet a bíróságtól, az ügyészségtől, az üzvétszabónsági szolgálatoktól, a társadalombiztosítási igazgatási szervektől. ha maximum öt főről kér adatot a hivataltól. hogy a rendőrség az adóhatóságtól, telefonszolgálatoktól, bankoktól ügyészi jóváhagyás nélkül kérjen adatokat, hogy kérje tőlük a művelődési tárca a bérfejlesztéshez szükséges adatokat.

adatgyűjtés

hanen Orbán Viktor és Deutsch Tanás ellen is adatgyűjtést kértek az ügyben érintett nagyműozomótól.


adóigazolás

asely szerint a gépkocsi átíratásakor adóigazolást kértek a polgároktól.

adókedvezmény

Korábban Budapest és a vidéki nagyvárosok különféle adókedvezményeket is kértek az előző kabinettől,

adószás

Ha a bérbesző magánszemély, akkor adószámat kell kérnie az APEH-től,  ha a magánszemély adószámat kér az APEH-től

aggregátor

A hadseregtől kértek aggregátort,

ajánlat

A helyreállításal megbízott Szabolcs-Szatmár-Bereg Megyei Közérkező Kht. tíz vállalkozótól másfél milliárd forintos felújítási munkára kért a közelszűltban kivitelezési ajánlat az eredménytelen pályázatok után a bizottság új ajánlatot kért a pályázóktól. és cserébe a pályázóktól - meghatározott műszaki paraméterekkel rendelkező - új kocskira kért ajánlatot. hogy az MKK vezetői már ajánlatokat kértek néhány légitársaságtól a norvégiai repülőútall kapcsolatban.

aláírás

A MÁV nyilatkozat aláírását kérte a szakszervezetektől, hogy az aláírásokat lehetőleg olyan személyektől kell kérni,

alkotmányértelmezés

Az MDF alkotmányértelmezést kér a legfőbb ügyésztől

Az Országgyűlés alkotmányügyi bizottsága absztrakt alkotmányértelmezést kér az Alkotmánybíróságtól az ellenékek által kezdeményezett népszavazás ügyében. így alkotmányértelmezést sen kérhet az Alkotmánybíróságtól.

állásfoglalás

A bírósági döntés előtt a TEBÉSZ állásfoglalást kért a Pénzügyminisztériumtól (PM) is az ügyben.

adatgyűjtés

hanem Orbán Viktor és Deutsch Tamás ellen is adatgyűjtést kértek az ügyben érintet

adóigazolás

amely szerint a gépkocsi átíratásakor adóigazolást kértek a polgároktól.

adókedvezmény

Korábban Budapest és a vidéki nagyvárosok különféle adókedvezményeket is kértek az

adószám

Ha a bérbeadó magánszemély, akkor adószámot kell kérnie az APEH-től,

ha a magánszemély adószámot kér az APEH-től

aggregátor

A hadseregtől kértek aggregátort,

ajánlat

A helyreállítással megbízott Szabolcs-Szatmár-Bereg Megyei Közútkezelő Kht. tíz vá

Az eredménytelen pályázatok után a bizottság új ajánlatot kért a pályázóktól.

és cserébe a pályázóktól - meghatározott műszaki paraméterekkel rendelkező - új ko

hogy az MTK vezetői már ajánlatokat kértek néhány légitársaságtól a norvégiai repü

aláírás

A MÁV nyilatkozat aláírását kérte a szakszervezetektől,

hogy az aláírásokat lehetőleg olyan személyektől kell kérni,

alkotmányértelmezés

Az MDF alkotmányértelmezést kér a legfőbb ügyésztől

Az Országgyűlés alkotmányügyi bizottsága absztrakt alkotmányértelmezést kér az Alk

így alkotmányértelmezést sem kérhet az Alkotmánybíróságtól.

állásfoglalás

A bírósági döntés előtt a TEBÉSZ állásfoglalást kért a Pénzügyminisztériumtól (PM



ESETTANULMÁNY

(1/2)

PÉLDA

'ad ACC' keret
a Magyar Nemzetben és az Index fórumban

8674 találat. Stat: [hang](#) [477] [lehetőség](#) [412] [válasz](#) [339] [otthon](#) [281] [tájékoztatás](#) [170] [hír](#) [209] [ok](#) [180] [áldás](#) [85] [esély](#) [114] [lendület](#) [74] [hely](#) [150] [pénz](#) [129] [magyarázat](#) [87] [szám](#) [128] [alkalom](#) [101] [tanács](#) [85] [koncert](#) [71] [munka](#) [116] [utasítás](#) [61] [felvilágosítás](#) [52] [hangverseny](#) [49] [támogatás](#) [93] [kép](#) [85] [igaz](#) [82] [lökés](#) [41] [százalék](#) [92] [segítség](#) [70] [engedély](#) [59] [megbízás](#) [50] [forint](#) [86] [információ](#) [59] [felhatalmazás](#) [39] [voks](#) [38] [ízeltő](#) [29] [alap](#) [69] [mód](#) [59] [jel](#) [48] [nyomaték](#) [27] [tér](#) [46]

7649 találat. Stat: [igaz](#) [320] [válasz](#) [275] [tanács](#) [172] [hang](#) [148] [tipp](#) [67] [ami](#) [212] [lehetőség](#) [155] [pénz](#) [151] [magyarázat](#) [88] [hála](#) [67] [ok](#) [116] [név](#) [109] [esély](#) [83] [ötlet](#) [68] [kegyelem](#) [49] [mi](#) [119] [hitel](#) [61] [vér](#) [44] [link](#) [34] [bizonyosság](#) [31] [hely](#) [74] [élet](#) [66] [cím](#) [54] [felmentés](#) [32] [több](#) [75] [valami](#) [59] [fegyver](#) [41] [áldás](#) [26] [az](#) [138] [maga](#) [62] [alap](#) [53] [semmi](#) [46] [alkalom](#) [40] [fej](#) [39] [információ](#) [37] [jel](#) [35] [interjú](#) [32] [parancs](#) [28] [megbízás](#) [28] [puszi](#) [18] [erő](#) [40] [hír](#) [35] [engedély](#) [33] [hit](#) [30] [kedvezmény](#) [27] [felvilágosítás](#) [22] [pofon](#) [20] [ez](#) [74] [sok](#) [40] [tér](#) [34] [kép](#) [30]

ESETTANULMÁNY

(1/2)

PÉLDA

'ad ACC' keret
a Magyar Nemzetben és az Index fórumban

8674 találat. Stat: hang [477] lehetőség [412] válasz [339] otthon [281] tájékoztatás [170] hír [209] ok [180] áldás [85] esély [114] lendület [74] hely [150] pénz [129] magyarázat [87] szám [128] alkalom [101] tanács [85] koncert [71] munka [116] utasítás [61] felvilágosítás [52] hangverseny [49] támogatás [93] kép [85] igaz [82] lökés [41] százalék [92] segítség [70] engedély [59] megbízás [50] forint [86] információ [59] felhatalmazás [39] voks [38] ízeltő [29] alap [69] mód [59] jel [48] nyomaték [27] tér [46]

közös elemek

7649 találat. Stat: igaz [320] válasz [275] tanács [172] hang [148] tipp [67] ami [212] lehetőség [155] pénz [151] magyarázat [88] hála [67] ok [116] név [109] esély [83] ötlet [68] kegyelem [49] mi [119] hitel [61] vér [44] link [34] bizonyosság [31] hely [74] élet [66] cím [54] felmentés [32] több [75] valami [59] fegyver [41] áldás [26] az [138] magá [62] alap [53] semmi [46] alkalom [40] fej [39] információ [37] jel [35] interjú [32] parancs [28] megbízás [28] puszi [18] erő [40] hír [35] engedély [33] hit [30] kedvezmény [27] felvilágosítás [22] pofon [20] ez [74] sok [40] tér [34] kép [30]

ESETTANULMÁNY

(1/2)

PÉLDA

'ad ACC' keret
a Magyar Nemzetben és az Index fórumban

8674 találat. Stat: [hang](#) [477] [lehetőség](#) [412] [válasz](#) [336] [otthon](#) [281] [tájékoztatás](#) [170] [hír](#) [309] [ok](#) [180] [áldás](#) [85] [esély](#) [114] [lendület](#) [74] [hely](#) [150] [pénz](#) [129] [magyarázat](#) [87] [szám](#) [128] [alkalom](#) [101] [tanács](#) [85] [koncert](#) [71] [munka](#) [116] [utasítás](#) [61] [felvilágosítás](#) [52] [hangverseny](#) [49] [támogatás](#) [93] [kép](#) [85] [igaz](#) [82] [lökés](#) [41] [százalék](#) [92] [segítség](#) [70] [engedély](#) [59] [megbízás](#) [50] [forint](#) [86] [információ](#) [59] [felhatalmazás](#) [39] [voks](#) [38] [ízeltő](#) [29] [alap](#) [69] [mód](#) [59] [jel](#) [48] [nyomaték](#) [27] [tér](#) [46]

eltérések

7649 találat. Stat: [igaz](#) [20] [válasz](#) [275] [tanács](#) [172] [hang](#) [143] [tipp](#) [61] [ami](#) [212] [lehetőség](#) [155] [pénz](#) [151] [magyarázat](#) [88] [hála](#) [67] [ok](#) [116] [név](#) [109] [esély](#) [83] [ötlet](#) [68] [kegyelem](#) [49] [mi](#) [119] [hitel](#) [61] [vér](#) [44] [link](#) [34] [bizonyosság](#) [31] [hely](#) [74] [élet](#) [66] [cím](#) [54] [felmentés](#) [32] [több](#) [75] [valami](#) [59] [fegyver](#) [41] [áldás](#) [26] [az](#) [138] [maga](#) [62] [alap](#) [53] [semmi](#) [46] [alkalom](#) [40] [fej](#) [39] [információ](#) [37] [jel](#) [35] [interjú](#) [32] [parancs](#) [28] [megbízás](#) [28] [puszi](#) [18] [erő](#) [40] [hír](#) [35] [engedély](#) [33] [hit](#) [30] [kedvezmény](#) [27] [felvilágosítás](#) [22] [pofon](#) [20] [ez](#) [74] [sok](#) [40] [tér](#) [34] [kép](#) [30]

ESETTANULMÁNY

(2/2)

PÉLDA

'ad **hang**ACC DAT'

a Magyar Nemzetben és az Index fórumban

467 találat. Stat: [meggyőződés](#) [61] [remény](#) [57] [aggodalom](#) [44][vélemény](#) [65] [értetlenség](#) [15] [csalódottság](#) [10] [aggály](#) [12] [megdöbbenés](#) [10][elégedetlenség](#) [8] [sajnálkozás](#) [7] [elégedettség](#) [7] [sajnálát](#) [7] [öröm](#) [8] [félelem](#) [7] [nézet](#) [6] = [10] = [6]116 találat. Stat: [vélemény](#) [28] [felháborodás](#) [6] [aggodalom](#) [6] = [9] = [8] & [6]

ESETTANULMÁNY

(2/2)

PÉLDA

'ad **hang**ACC **DAT**'

a Magyar Nemzetben és az Index fórumban

467 találat. Stat: [meggyőződés](#) [61] [remény](#) [57] [aggodalom](#) [44]
[vélemény](#) [65] [értetlenség](#) [15] [csalódottság](#) [10] [aggály](#) [12] [megdőbbenés](#) [10]
[elégedetlenség](#) [8] [sajnálkozás](#) [7] [elégedettség](#) [7] [sajnálát](#) [7] [öröm](#) [8] [félelem](#) [7] [nézet](#) [6] = [10] = [6]

közös elemek

116 találat. Stat: [vélemény](#) [28] [felháborodás](#) [6] [aggodalom](#) [6] = [9] = [8] & [6]

ESETTANULMÁNY

(2/2)

PÉLDA

'ad hangACC DAT'

a Magyar Nemzetben és az Index fórumban

467 találat. Stat: meggyőződés [61] remény [57] aggodalom [44]
vélemény [65] értetlenség [15] csalódottság [10] aggály [12] megdöbbenés [10]
elégedetlenség [8] sajnálkozás [7] elégedettség [7] sajnálát [7] öröm [8] télelem [7] nézet [6] = [10] = [6]

116 találat. Stat: vélemény [28] felháborodás [6] aggodalom [6] = [9] = [8] & [6]

- 1 BEVEZETÉS
- 2 TAGMONDATRA BONTÁS
- 3 FELÜLET
- 4 FEJLESZTÉSI LEHETŐSÉGEK, ALKALMAZÁS**

FEJLESZTÉSI LEHETŐSÉGEK, ALKALMAZÁS

- szintaktikai elemzést követően *is* lehessen alkalmazni a tagmondatra bontást
- névszói állítmányok kezelése – „*prédikátum-azonosító*”
- adott ige összes *lehetséges* keretére irányuló lekérdezés
- párhuzamos eloszlás *több* bővítményre

→ Korpuszalapú német–magyar valenciaszótár terve

→ Akadémiai Nagyszótár

HOZZÁFÉRÉS

```
http://corpus.nytud.hu/mazsola
```

Igényeljen jelszót a `joker@nytud.hu` címen!

HOZZÁFÉRÉS

```
http://corpus.nytud.hu/mazsola
```

Igényeljen jelszót a `joker@nytud.hu` címen!

Köszönöm a figyelmet!