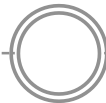




A nyelvtechnológia alapjai



9. Gyakorlat

Gépi fordítás és kiértékelése



Google



Bejelentkezés

Fordító

Azonnali fordítás kikapcsolása



angol magyar német Nyelvfelismerés ▼



magyar angol német ▼

Fordítás

I am a spendthrift at heart.



28/5000

Én egy pazarló a szíve.





Google



Bejelentkezés

Fordító

Azonnali fordítás kikapcsolása



angol magyar német Nyelvfelismerés ▼



magyar angol német ▼

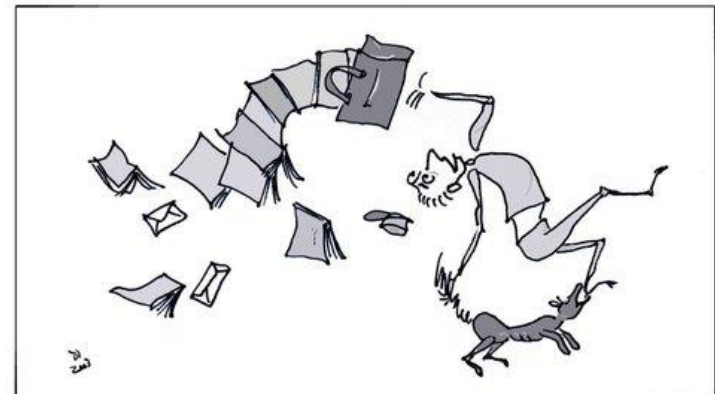
Fordítás

The dog was bitten by the postman.



34/5000

A kutya megharapta a postást.





- Korai ifjúságban az oldalak alakulatában, tapasztalta egy visszautasítás megalázását, amikor próbált miközben nehézségekben volt, hogy pénzt vegyen kölcsön, és azóta ő soha egyszer ugyanabba tette magát pozíció megint.



- Lexikai
 - red vs. vörös, piros
- Morfológiai
 - izoláló vs. poliszintetikus
 - agglutináló vs. flektáló
- Szintatikai
 - Subject-Verb-Object, VSO, SOV
- Egyéb strukturális különbségek
 - YYYY. MM. DD. vs. MM/DD/YY

A gépi fordítás néhány tipikus nehézségéről

- Lexikai többértelműségek
 - *Egyik nap még a nap is süttött.*
- Strukturális többértelműségek
 - *Dani szerzett egy könyvet a piacról.*
- Többszavas kifejezések (terminusok)
 - *A lány elsírta magát a lovagi tornán.*
- Idiómák
 - *Itatja az egereket. Kutyából nem lesz szalonna.*



Gépi fordítás célja

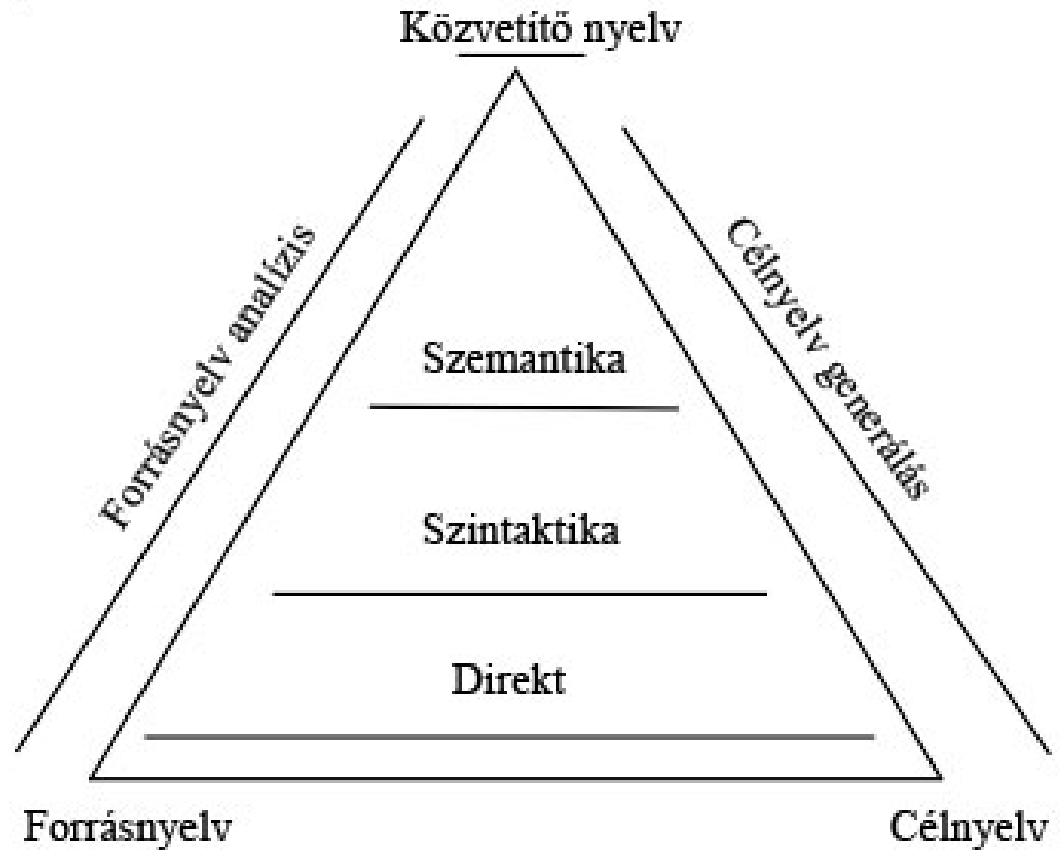
- megértés,
- információkinyerés, nem a tökéletes fordítás



- Példa alapú gépi fordítás
- Szabály alapú gépi fordítás
- Statisztikai gépi fordítás
- Hibrid gépi fordítás
- Emberi közreműködéssel történő gépi fordítás
- Neurális gépi fordítás

- a felnőtt: szabályok megtanulása és szótár

- direkt fordítás
- transzfer
- interlingua



<http://www.webforditas.hu/>

The dog was bitten by the postman.

A kutyát megharapta a postás.



The man saw the girl with the telescope.

Az ember a távcsővel látta a lányt.

Az ember, akinek kék háza, egy kutya, öt piros labdája és hat gyönyörű elvált lánya van, örül.

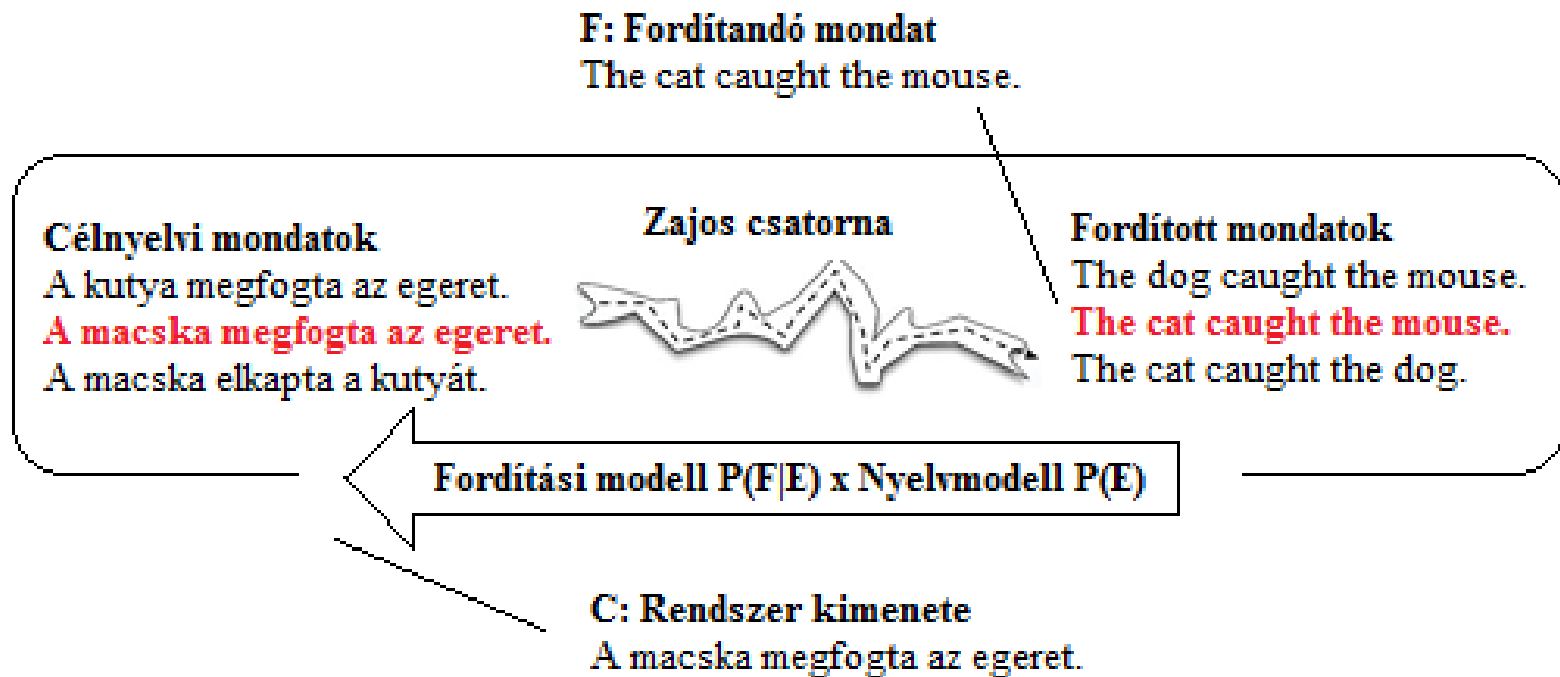
The man, who, a dog of his, five of his red balls and divorced one with six pleasures has blue house, a daughter, is glad.

Tegnap este facebookoztam.

Yesterday in the evening facebookoztam.



- a gyerek: példák megfigyelése
- párhuzamos korpuszból tanul
- emberek által előre lefordított szövegek felhasználásán alapszik
- beszédtechnológiában használt Shannon-féle zajoscsatorna-modelljét vették alapul



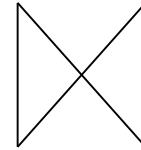
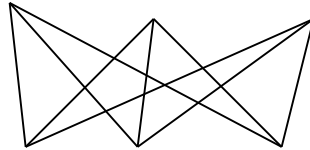
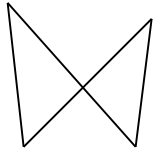


- Fordítási modell
 - Párhuzamos korpusz
- Nyelvmodell / N-gram modell
 - Nem esik messze az alma ...
- Dekódoló algoritmus
 - nyelvmodell + fordítási modell + bemenő mondat
=> fordítás
- Problémák:
 - Tulajdonnevek, számok / dátumok, főnévi csoportok



- IBM model
 - Expectation–maximization algorithm (EM algoritmus)

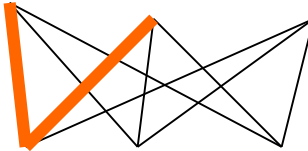
... la maison ... la maison bleue ... la fleur ...



... the house ... the blue house ... the flower ...



- IBM model
 - Expectation–maximization algorithm (EM algoritmus)

... la maison ... la maison bleue ... la fleur ...

... the house ... the blue house ... the flower ...



- IBM model
 - Expectation–maximization algorithm (EM algoritmus)

... la maison ... la maison bleue ... la fleur ...

... the house ... the blue house ... the flower ...



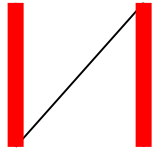
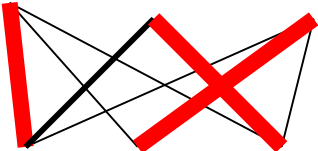
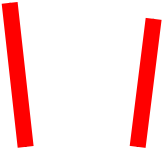
- IBM model
 - Expectation–maximization algorithm (EM algoritmus)

... la maison ... la maison bleue ... la fleur ...

... the house ... the blue house ... the flower ...



- IBM model
 - Expectation–maximization algorithm (EM algoritmus)

... la maison ... la maison bleue ... la fleur ...

... the house ... the blue house ... the flower ...



- IBM model
 - Expectation–maximization algorithm (EM algoritmus)

... la maison ... la maison bleue ... la fleur ...
... the house ... the blue house ... the flower ...

Diagram illustrating word alignment between French and English sentences. Red lines connect words in the French sentence to their corresponding words in the English sentence. The alignment is as follows:

- la (French) connects to the (English)
- maison (French) connects to house (English)
- bleue (French) connects to blue (English)
- fleur (French) connects to flower (English)

The word "bleue" is crossed out with a red X, indicating it is not aligned with "house".

$$p(\text{la} | \text{the}) = 0.453$$

$$p(\text{le} | \text{the}) = 0.334$$

$$p(\text{maison} | \text{house}) = 0.876$$

$$p(\text{bleu} | \text{blue}) = 0.563$$

- Szóalapú n-gram modell

1-gram: Ez egy nem túl hosszú mondat .

2-gram: Ez egy nem túl hosszú mondat .

3-gram: Ez egy nem túl hosszú mondat .

4-gram, 5-gram ...

- Google 2016 előtt

[Bejelentkezés](#)

FordítóAzonnali fordítás kikapcsolása

angol magyar német Nyelvfelismerés

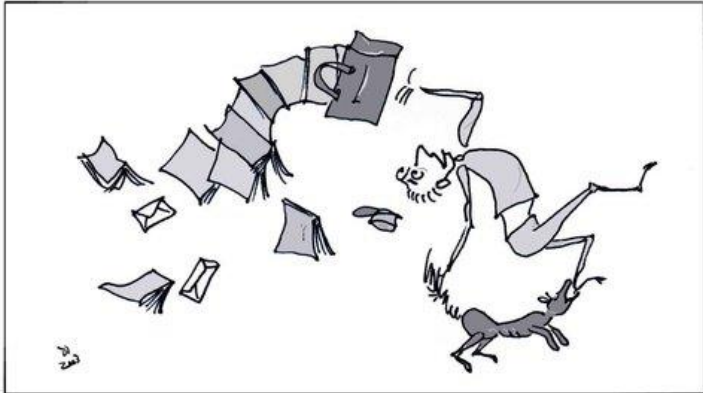
 magyar angol német

Fordítás

The dog was bitten by the postman.
  34/5000



A kutya megharapta a postást.
   





Microsoft Translator



Angol ▾

Then God said, "Let there be light"; and there was light. ✕



Magyar ▾

Akkor Isten azt mondta: "Legyen világosság"; és lőn világosság.

Fordítás

Then God said, "Let there be light"; and there was light.
Akkor Isten azt mondta: "Legyen világosság"; és lőn világosság.

there was light – nem volt könnyű



The man saw the girl with the telescope.

A férfi a lány látta, hogy a távcső.

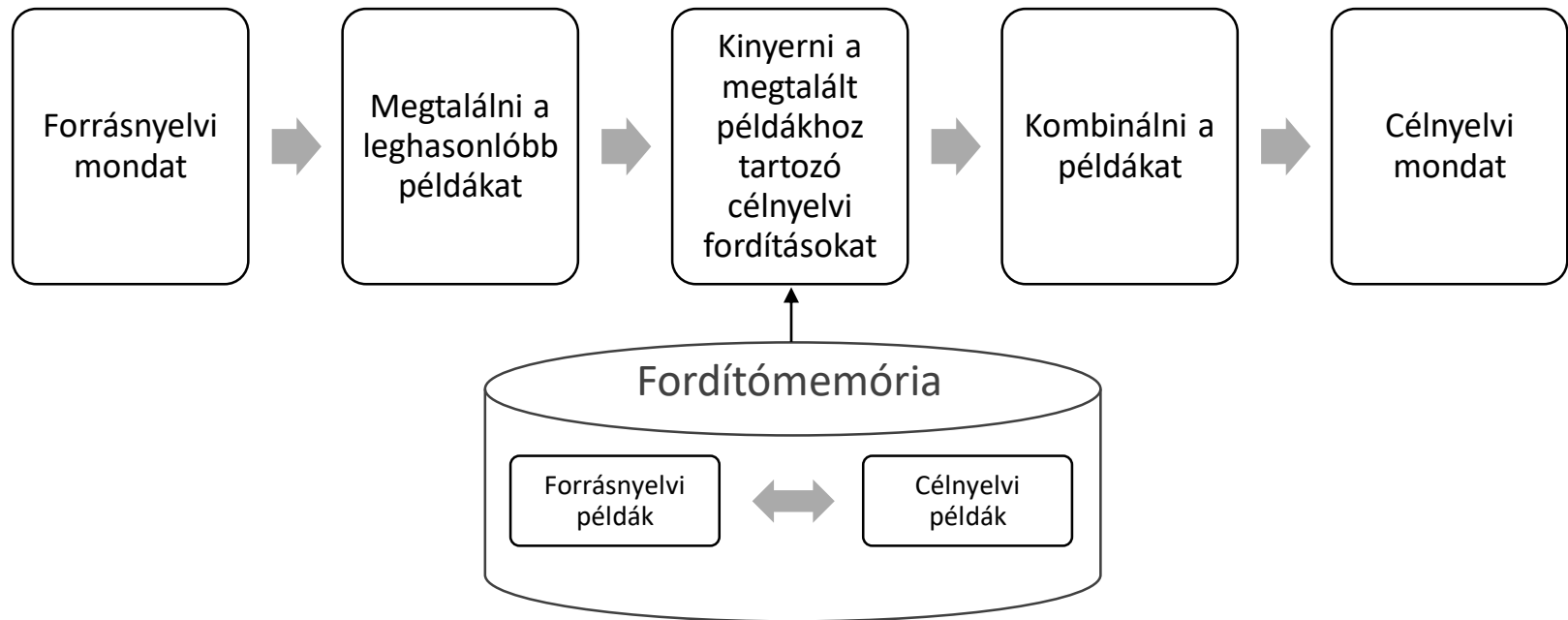
Az ember, akinek kék háza, egy kutya, öt piros labdája és hat gyönyörű elvált lánya van, örül.

The man, whose Blue's House, a dog, five red labdája and six beautiful daughters divorced, happy.

Tegnap este facebookoztam.

Last night facebookoztam.

- Fordítómemória





- Fordítandó szöveg:
A kedves nagymama macskája megfogta az egeret.
- A memóriában található szöveg:
A kedves nagymama kutyája megfogta a csúnya macskát



- Fordítandó szöveg:
A kedves nagymama macskája megfogta az egeret.
- A memóriában található szöveg:
A kedves nagymama kuttyája megfogta a csúnya macskát.



- Fordítandó szöveg:
A kedves nagymama macskája megfogta az egeret.
- A memóriában található szöveg:
A kedves nagymama kutyája megfogta a csúnya macskát.
The sweet grandma's dog caught the nasty cat.



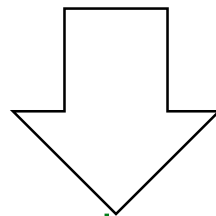
- Fordítandó szöveg:

A kedves nagymama macskája megfogta az egeret.

- A memóriában található szöveg:

A kedves nagymama kutya megfogta a csúnya macskát.

The sweet grandma's dog caught the nasty cat.

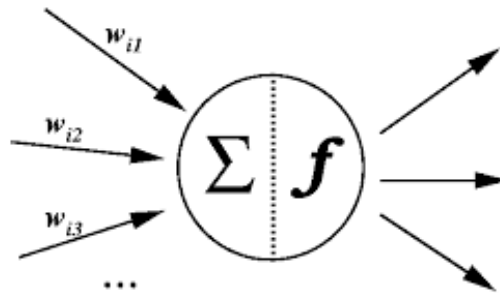
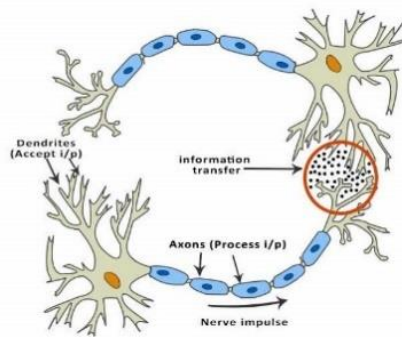


The sweet grandma's cat caught the mouse.

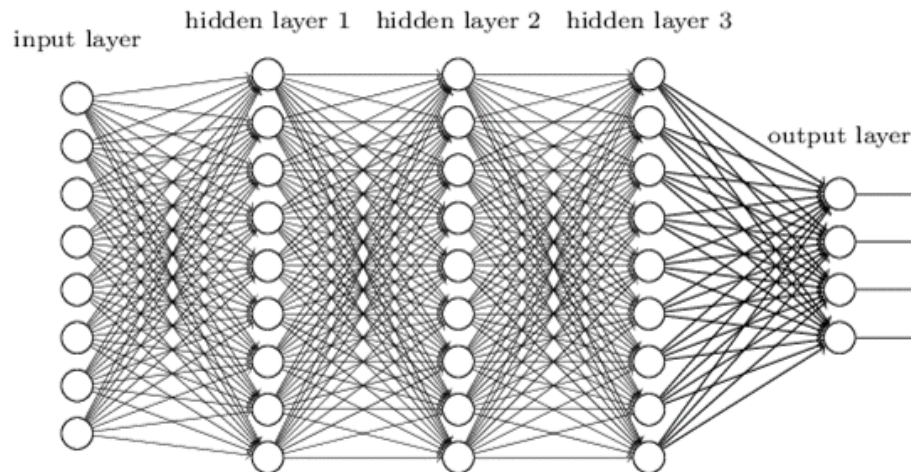
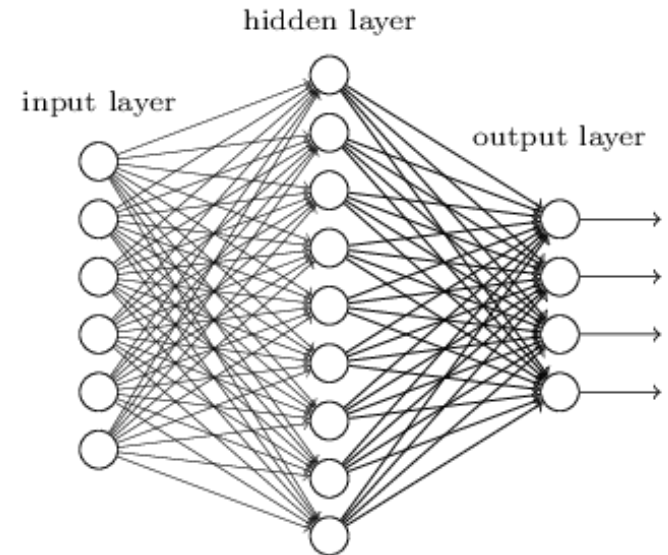


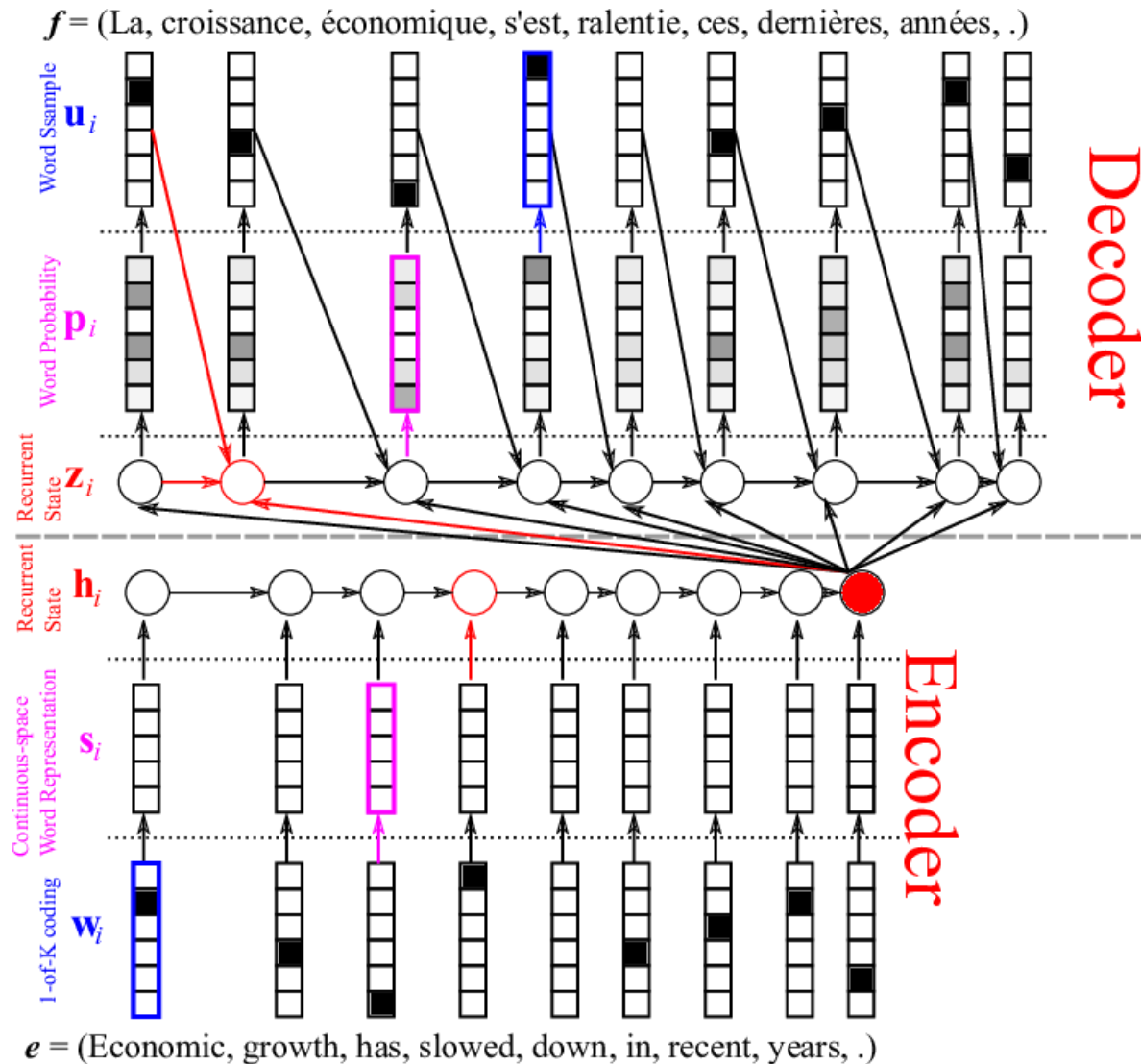
- Szabályalapú gépi fordító + statisztika
- Statisztikai gépi fordító + szabályok
- Szabályalapú gépi fordító + fordítómemória
- Statisztikai gépi fordító + fordítómemória
- Fordítómemória
 - + Statisztikai / Szabályalapú gépi fordító

- Neurális hálózat

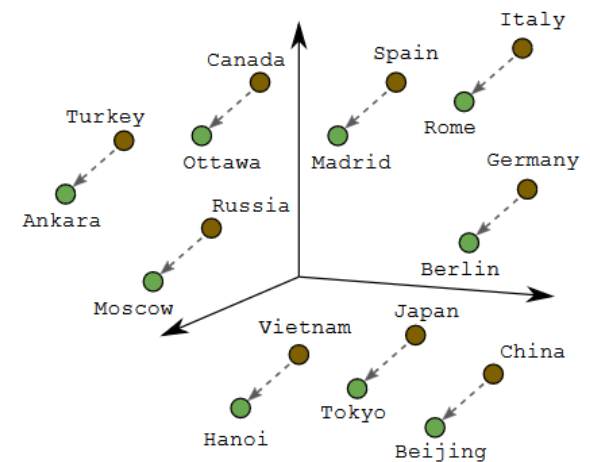
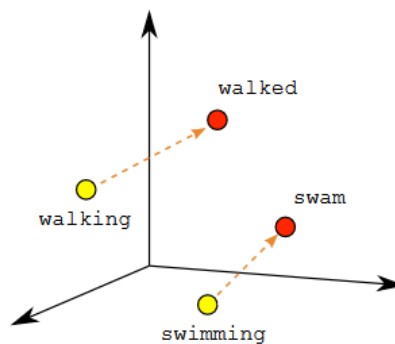
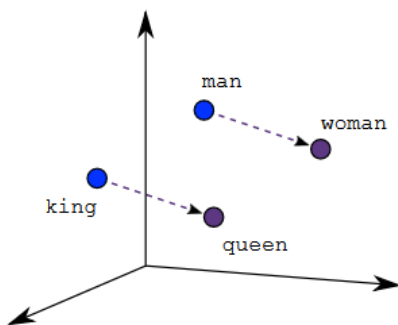


Deep neural network





- Szóbeágyazás (Tomas Mikolov, Google)
- A szavakat a környezetükkel jellemezzük.
 - A **kutya** mancsa barna.
 - A **cica** mancsa fekete.
 - Az **ember** mancsa ...
 - A **szék** mancsa ...

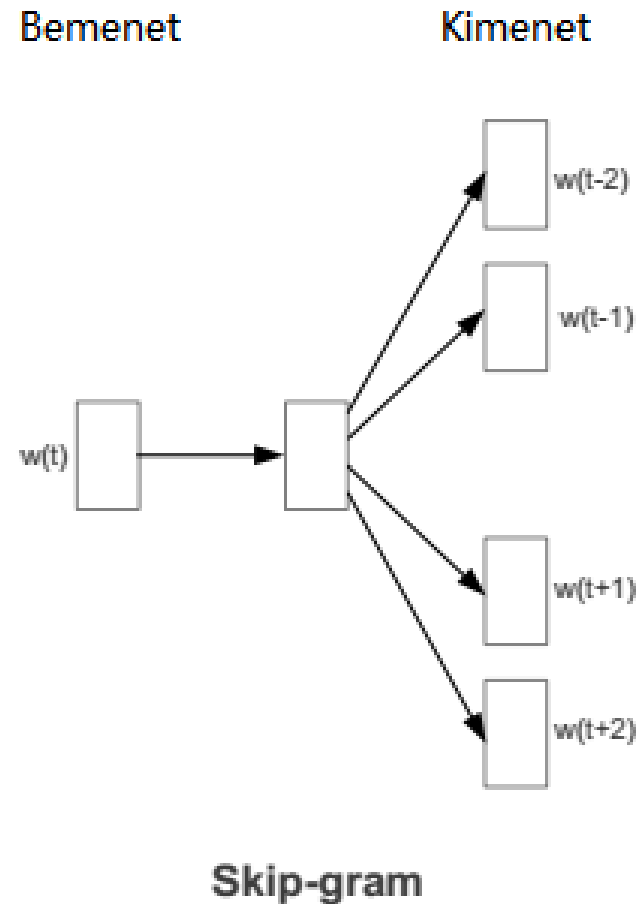
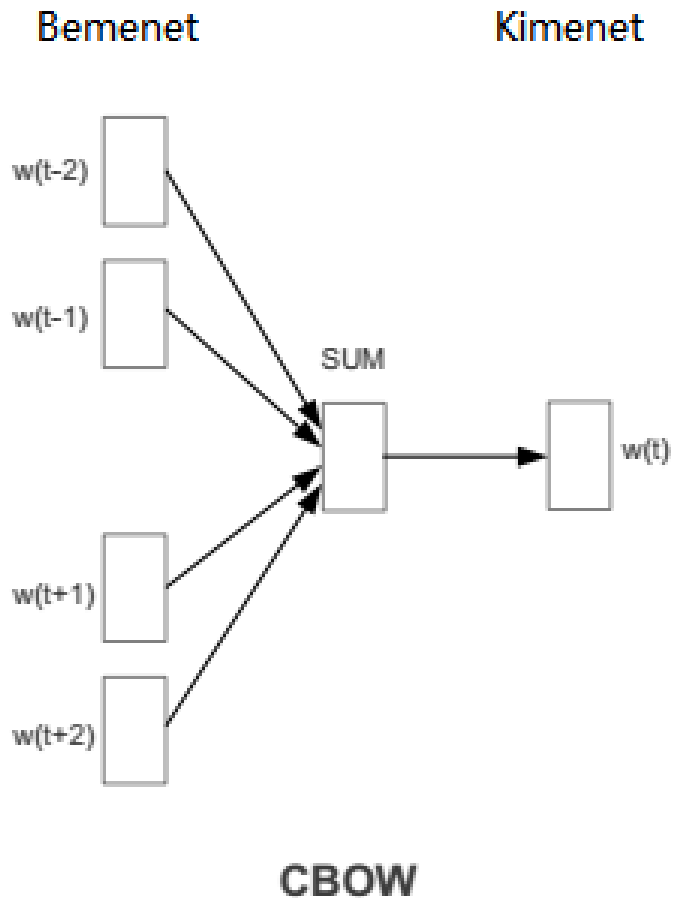




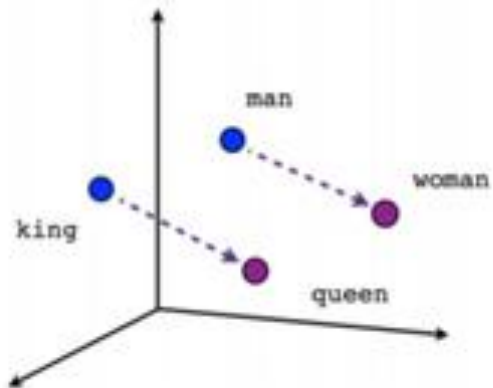
- Szóbeágyazás: nyelvmodellezési és jegytanulási technikák gyűjtőneve, ahol a szavakat egy n -dimenziós valós vektor reprezentálja
- A leképezés neurális hálókkal, a szókörnyezet reprezentációjával történik
- A lexikai elemek egy valós vektortér egyes pontjai, melyek konzisztensen helyezkednek el az adott térben.
- Az egymáshoz szemantikailag és/vagy morfológiailag hasonló szavak az adott vektortérben közel állnak egymáshoz.



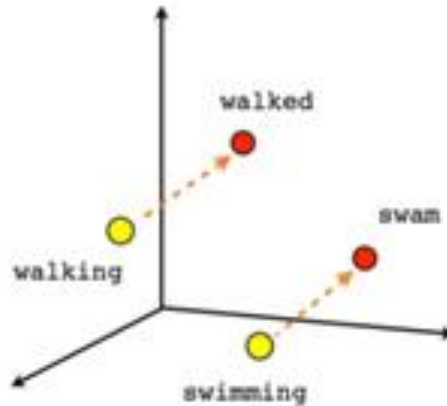
- Mikolov (2013): word2vec egy kétrétegű neurális háló szövegreprezentációs célokra, ahol a bemenet a szövegtörzs és a kimenet a törzs szavait reprezentáló jegyvektorok halmaza
- A word2vec maga nem mélyneurális háló, de a kimenetét mélyneurális hálók is tudják kezelni
- Két lehetséges alapmodell:
 - (1) környezet→szó: folytonos szózsák (Continuous Bag Of Words, CBOW)
 - (2) szó→környezet: skip-gram



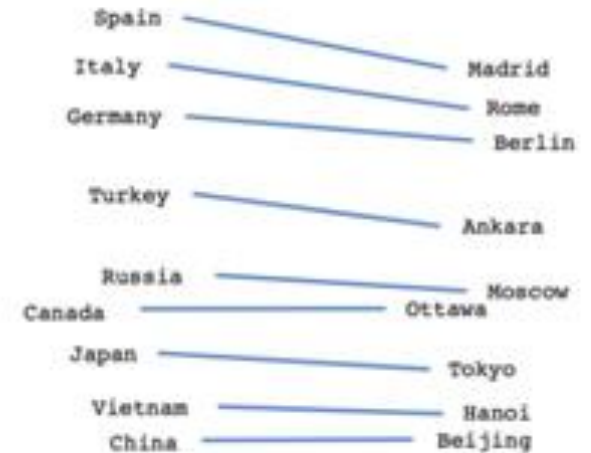
Szóbeágyazás (word embedding)



Male-Female

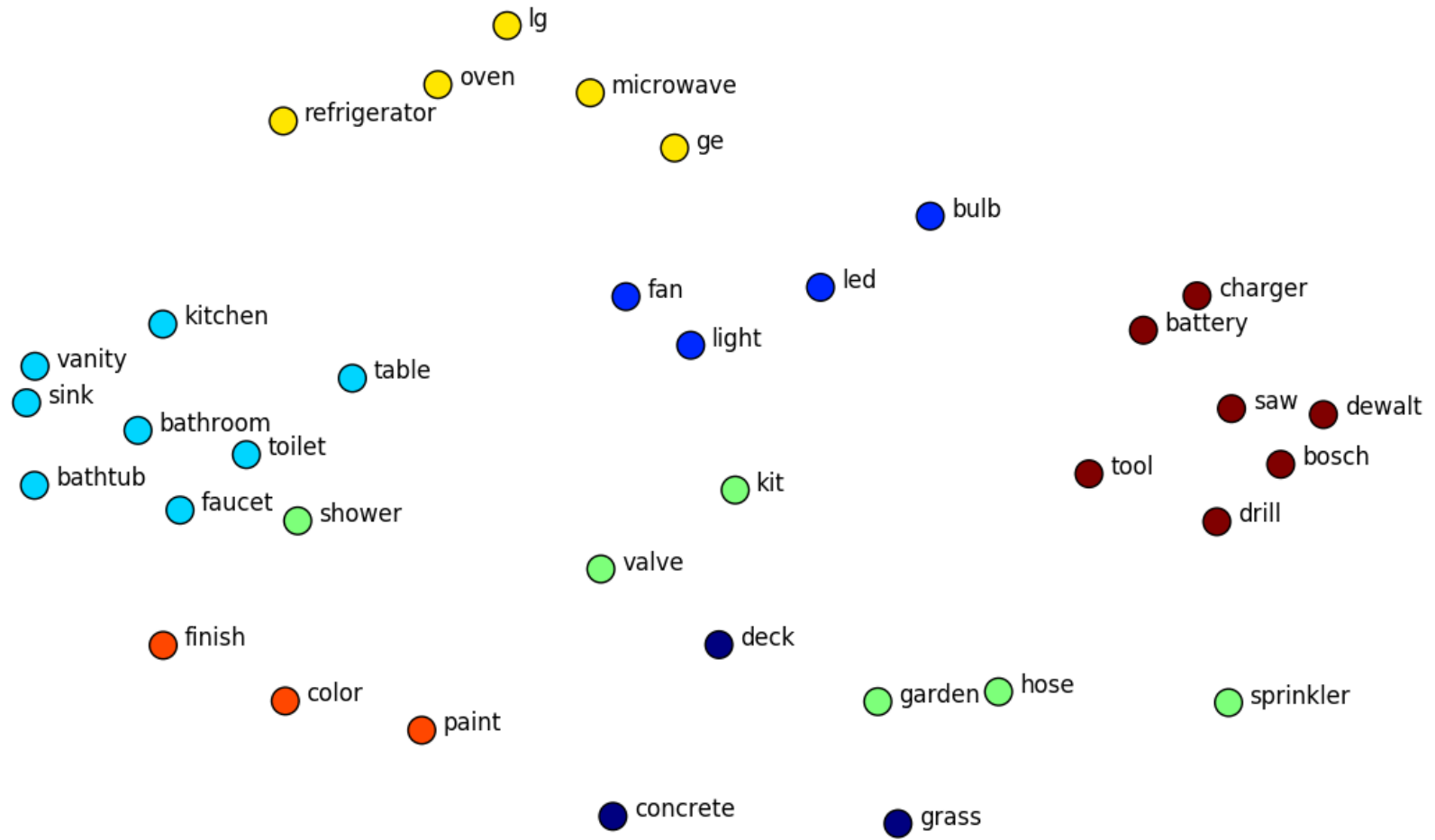


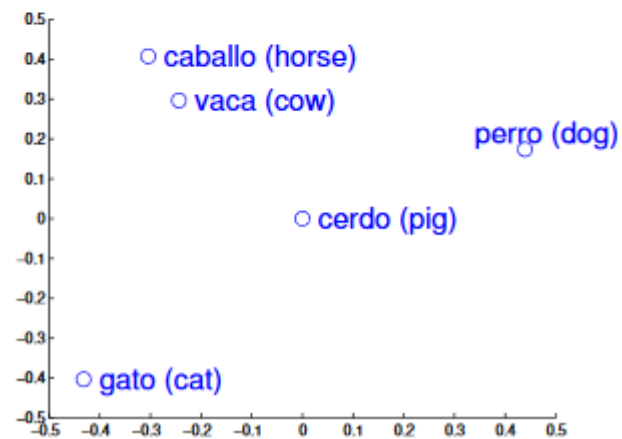
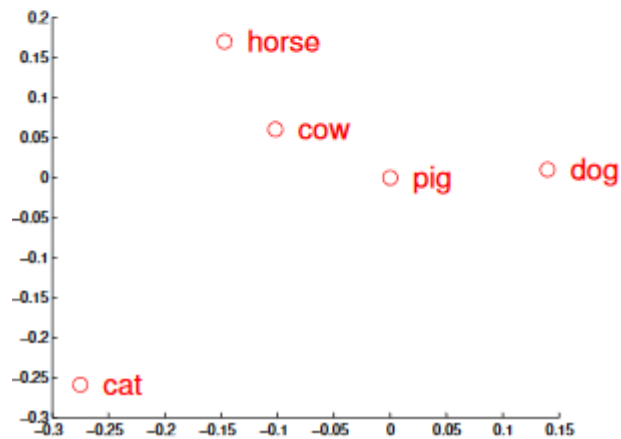
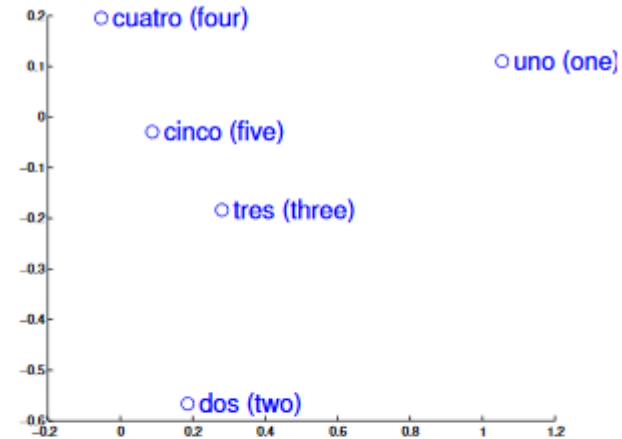
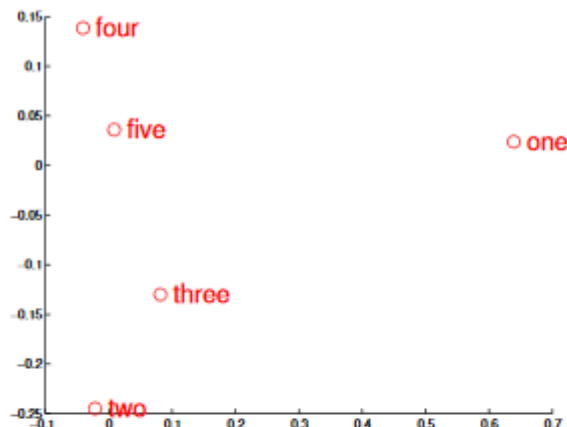
Verb tense



Country-Capital

Szóbeágyazás (word embedding)





Az első kísérletek magyarra: főnevek, igék

0	kenyerek	1	2270	0	sógorom	1	2176	0	iszunk	1	6737
1	zsemlék	0.8105	283	1	sógornőm	0.8524	1855	1	megiszunk	0.8151	907
2	péksütemények	0.8048	997	2	nagynéném	0.8374	3088	2	igyunk	0.8046	5814
3	kekszek	0.7972	1046	3	nagybátyám	0.8314	4394	3	ihatunk	0.7878	875
4	pékárúk	0.7957	771	4	apósom	0.8235	2538	4	eszünk	0.7638	18243
5	tészták	0.7881	2466	5	nővérem	0.8208	10266	5	innánk	0.7497	432
6	lepények	0.7849	202	6	keresztanyám	0.8170	859	6	fogyasztunk	0.7429	7196
7	kiflik	0.7843	349	7	vejem	0.8121	526	7	ittunk	0.7342	8195
8	kalácsok	0.7841	277	8	keresztapám	0.8105	931	8	megisszuk	0.7339	694

0	franciakulcs	1	255	0	csavargó	1	6123	0	migráns	1	3264
1	feszítővas	0.8590	846	1	koldus	0.8026	15793	1	bevándorló	0.8162	23954
2	csavarkulcs	0.8445	473	2	zsivány	0.7691	3497	2	menedékkérő	0.8043	1309
3	csípőfogó	0.8242	345	3	haramia	0.7624	2024	3	vendégmunkás	0.7241	4808
4	pajszer	0.8219	567	4	csirkefogó	0.7570	2019	4	menekült	0.7218	27430
5	hidegvágó	0.8054	156	5	vadember	0.7531	2497	5	kivándorló	0.7083	1893
6	csavarhúzó	0.7984	4369	6	útonálló	0.7472	1942	6	szexmunkás	0.6668	803
7	csőfogó	0.7890	111	7	utcagyerek	0.7401	1653	7	roma	0.6571	162822
8	villáskulcs	0.7890	764	8	gazfickó	0.7366	1651	8	EU-állampolgár	0.6269	316

... és érti a tulajdonneveket?



0	Luther	1	16466
1	Zwingli	0.7782	853
2	Melanchton	0.7558	873
3	Augustinus	0.7487	1749
4	Kálvin	0.7467	19884
5	Melanchthon	0.7313	626
6	Bucer	0.7218	274
7	Servet	0.7128	1202
8	reformátor	0.7068	4649

0	Majka
1	Kozsó
2	Lilu
3	Tilla
4	Dundika
5	Győzike
6	Dopeman
7	Liptai
8	Cristofel
9	Alekosz
10	Ágica

Google



Bejelentkezés

Fordító

Azonnali fordítás kikapcsolása



angol magyar német Nyelvfelismerés ▼



magyar angol német ▼

Fordítás

The dog was bitten by the postman. ✕



35/5000

A kutyát megharapta a posta.

A kutyát megharapta a posta.

A kutya megharapta a postást.

Fordítás javítása

A kutyát megharapta a posta

Alkalmaz

 Módosítási javaslat



The man saw the girl with the telescope.

A férfi látta a lányt a távcsővel.

Az ember, akinek kék háza, egy kutya, öt piros labdája és hat gyönyörű elvált lánya van, örül.

A man with a blue house, a dog, five red balls and six beautiful divorced daughters, is happy.

Tegnap este facebookoztam.

Last night I was back.



- Precision (pontosság) – Recall (fedés)
- feladat: célhalmaz elemeinek azonosítása
- eljárás eredménye: eredményhalmaz
- az eredményhalmaz jóságának mérésére szolgáló két mérték:

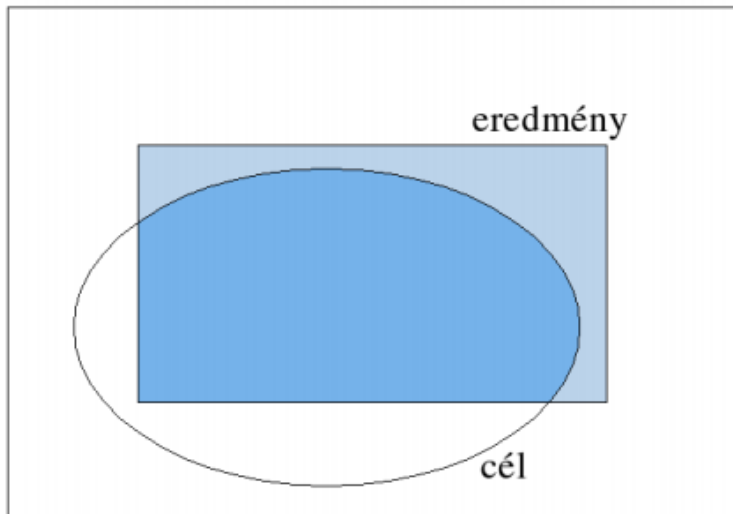


Precision (pontosság)

$$|\text{metszet}| / |\text{eredmény}|$$

„a találatokból hány volt eredetileg jó”

precision

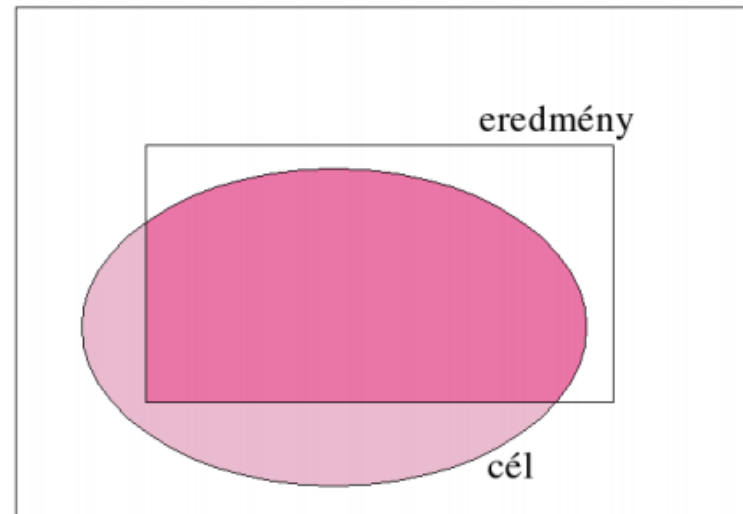


Recall (fedés)

$$|\text{metszet}| / |\text{cél}|$$

„az eredetileg jók közül hányat találtunk meg”

recall



Cél

Pozitív
eredmény

True positive

False positive

Negatív
eredmény

False negative

True negative



$$\text{pontosság} = \frac{|\{\text{releváns adat}\} \cap \{\text{kinyert adat}\}|}{|\{\text{kinyert adat}\}|}$$

$$\text{pontosság} = \frac{\text{true positive}}{\text{true positive} + \text{false positive}}$$

$$\text{fedés} = \frac{|\{\text{releváns adat}\}| \cap |\{\text{kinyert adat}\}|}{|\{\text{releváns adat}\}|}$$

$$\text{fedés} = \frac{\text{true positive}}{\text{true positive} + \text{false negative}}$$

- F-measure, F-score, F1 score
- A pontosság és a fedés harmonikus közepe/átlaga

$$F = 2 * \frac{\textit{pontosság} * \textit{fedés}}{\textit{pontosság} + \textit{fedés}}$$



- Tartalomhűség (adequacy)
 - jelentésben mennyire jó
- Gördülékenység, nyelvhelyesség (fluency)
 - a célnyelvi mondat nyelvtanilag mennyire helyes
- Elfogadhatóság
 - Korai ifjúságban az oldalak alakulatában, tapasztalta egy visszautasítás megalázását, amikor próbált miközben nehézségekben volt, hogy pénzt vegyen kölcsön, és azóta ő soha egyszer ugyanabba tette magát pozíció megint.



- Referenciafordítással történő kiértékelés:
 - Szükséges hozzá emberi fordítás.
 - Drága, nem valós idejű kiértékelés.
- Minőségbecslés (referenciafordítás nélküli):
 - Nem használ referenciafordítást.
 - Valós idejű kiértékelés.
 - Magas korreláció az emberi kiértékeléssel.

- **fordítás:** It is a guide to action which ensures that the military always obeys the commands of the party
- **referencia:** It is a guide to action that ensures that the military will forever heed Party commands

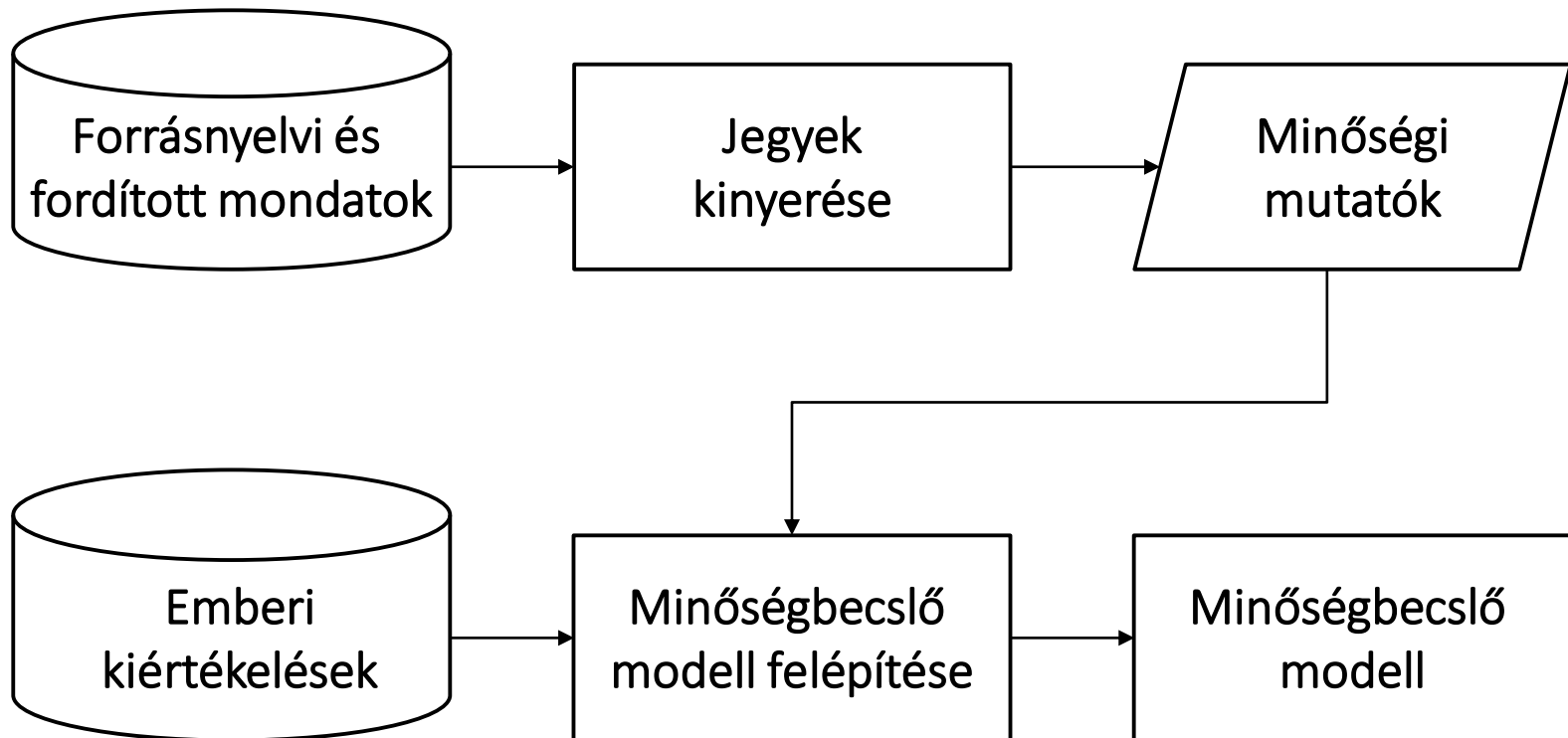
- **fordítás:** It is a guide to action which ensures that the military always obeys the commands of the party
- **referencia:** It is a guide to action that ensures that the military will forever heed Party commands



- Automatikus kiértékelő módszerek.
- Nincsen szükség emberi beavatkozásra.
- Nem kell ismerni a két nyelvet.
- Hasonlóságot és eltérést mér a fordítás és a referenciafordítás között.
- Módszerek:
 - BLEU, NIST, METEOR, LEPOR, ROGUE, RIBES
 - WER, TER, HTER



- Minőségbecslés – Quality Estimation
- Nincsen referenciafordítás
- Minőségi mutatók kinyerése a forrásnyelvi és a célnyelvi mondatokból.
- Gépi tanuláson alapuló módszer.
- Emberi kiértékelésekre betanított modell
 - Magasan korrelál az emberi kiértékeléssel.
- Minőségbecslő keretrendszer
 - QuEst





Forrásnyelvi mondat	Célnyelvi (lefordított) mondat	Minőségi mutató						Becslés	Emberi kiértékelés
She shot herself.	Lelőtte magát.	4.0	3.0	1.33	0.75	0.25	=	?	3.943
Now, as i have never known any ...	Most, ahogy még sohasem ismert ...	26.0	19.0	1.37	0.73	0.26	=	?	2.914
No work, no dinner.	Nincs munka, nincs vacsora.	8.0	8.0	1.00	1.00	0.00	=	?	3.596
But there seemed no place to go ...	De nem volt hová mennem ...	8.0	6.0	1.33	0.75	0.25	=	?	2.500
We tramped through the dim corridors ...	Mi órákon át a sötét folyosókon, kanyargós ...	34.0	29.0	1.17	0.85	0.14	=	?	2.627
Then he jumped like a cat ...	Aztán felugrott, mint egy macska, ...	28.0	26.0	1.08	0.93	0.07	=	?	3.275



Köszönöm a figyelmet



ध्यान देने के लिए आपका धन्यवाद
आप अपना ध्यान के लिए धन्यवाद
थैंक यू फॉर योर अटेंशन

Thank you for your attention
Thank you for your attention
Thank you for the attention

Merci de votre attention
Je vous remercie pour votre attention
Merci pour l'attention

感谢您的关注
谢谢你的关注
谢谢您的注意