

Szemantikai elemzés gráf-transzformációkkal

Kovács Ádám

`adaam.ko@gmail.com`

2019.01.14.

Kontribúció

Kontribúció

- Gráf alapú szemantikai modellek definiálása

Kontribúcióm

- Gráf alapú szemantikai modellek definiálása
- 4lang funkcionalitásának szolgáltatásokba való csomagolása

Kontribúció

- Gráf alapú szemantikai modellek definiálása
- 41ang funkcionalitásának szolgáltatásokba való csomagolása
- Modelleink kiértékelése valós feladatokon
 - Knowledge base population (Kovács and Recski, 2018)
 - Natural language inference
 - Machine comprehension (Kovács and Gémes, 2018)

4lang: a formalizmus (Kornai, 2010; Kornai et al., 2015)

41ang: a formalizmus (Kornai, 2010; Kornai et al., 2015)

Irányított fogalmi gráfok, 3 típusú él:

41ang: a formalizmus (Kornai, 2010; Kornai et al., 2015)

Irányított fogalmi gráfok, 3 típusú él:

- 0 típusú él

4lang: a formalizmus (Kornai, 2010; Kornai et al., 2015)

Írányított fogalmi gráfok, 3 típusú él:

- 0 típusú él
 - tulajdonság: $\text{dog} \xrightarrow{0} \text{large}$

41ang: a formalizmus (Kornai, 2010; Kornai et al., 2015)

Irányított fogalmi gráfok, 3 típusú él:

- 0 típusú él
 - tulajdonság: $\text{dog} \xrightarrow{0} \text{large}$
 - IS_A reláció (hipernima): $\text{dog} \xrightarrow{0} \text{mammal}$

41ang: a formalizmus (Kornai, 2010; Kornai et al., 2015)

Irányított fogalmi gráfok, 3 típusú él:

- 0 típusú él
 - tulajdonság: $\text{dog} \xrightarrow{0} \text{large}$
 - IS_A reláció (hipernima): $\text{dog} \xrightarrow{0} \text{mammal}$
 - állítmány: $\text{dog} \xrightarrow{0} \text{bark}$

41ang: a formalizmus (Kornai, 2010; Kornai et al., 2015)

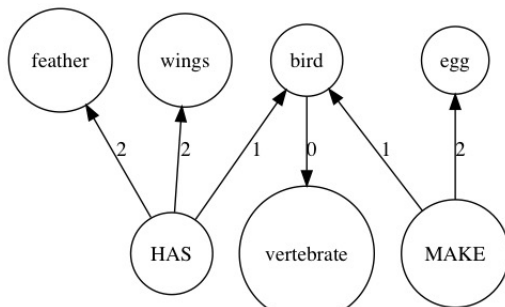
Irányított fogalmi gráfok, 3 típusú él:

- 0 típusú él
 - tulajdonság: $\text{dog} \xrightarrow{0} \text{large}$
 - IS_A reláció (hipernima): $\text{dog} \xrightarrow{0} \text{mammal}$
 - állítmány: $\text{dog} \xrightarrow{0} \text{bark}$
- 1- és 2-es él bináris állítmányt köt össze az argumentumokkal.

41ang: a formalizmus (Kornai, 2010; Kornai et al., 2015)

Irányított fogalmi gráfok, 3 típusú él:

- 0 típusú él
 - tulajdonság: $\text{dog} \xrightarrow{0} \text{large}$
 - IS_A reláció (hipernima): $\text{dog} \xrightarrow{0} \text{mammal}$
 - állítmány: $\text{dog} \xrightarrow{0} \text{bark}$
- 1- és 2-es él bináris állítmányt köt össze az argumentumokkal.



A 4lang szolgáltatás

A 4lang szolgáltatás

- Éles service a text_to_4lang modulra építve

A 4lang szolgáltatás

- Éles service a text_to_4lang modulra építve
- Magasan automatizált gráfgenerálás

A 4lang szolgáltatás

- Éles service a text_to_4lang modulra építve
- Magasan automatizált gráfgenerálás
- Online demo
 - <http://4lang.hlt.bme.hu>

A 4lang szolgáltatás

- Éles service a text_to_4lang modulra építve
- Magasan automatizált gráfgenerálás
- Online demo
 - <http://4lang.hlt.bme.hu>
- Github
 - <https://github.com/adaamko/4lang>

KBP

KBP

- Cél: rejtett tények felszínre hozása entitásokról egy tudásbázisban

KBP

- Cél: rejtett tények felszínre hozása entitásokról egy tudásbázisban
- Legtöbb megoldás különbözik a miénktől

KBP

- Cél: rejtett tények felszínre hozása entitásokról egy tudásbázisban
- Legtöbb megoldás különbözik a miénktől
 - Strukturálatlan adat

KBP

- Cél: rejtett tények felszínre hozása entitásokról egy tudásbázisban
- Legtöbb megoldás különbözik a miénktől
 - Strukturálatlan adat
 - Mi következtetést használtunk

KBP

- Cél: rejtett tények felszínre hozása entitásokról egy tudásbázisban
- Legtöbb megoldás különbözik a miénktől
 - Strukturálatlan adat
 - Mi következtetést használtunk
- Tudásbázis: WikiData

KBP

- Cél: rejtett tények felszínre hozása entitásokról egy tudásbázisban
- Legtöbb megoldás különbözik a miénktől
 - Strukturálatlan adat
 - Mi következtetést használtunk
- Tudásbázis: WikiData
- Szemantikai elemző: 4lang

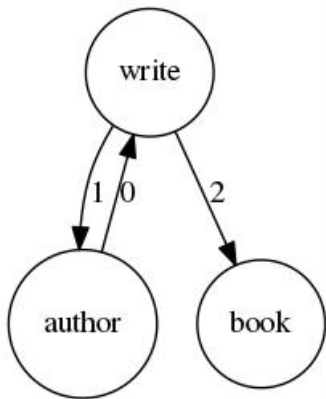
Szabályok

Szabályok

author(George Orwell, 1984)

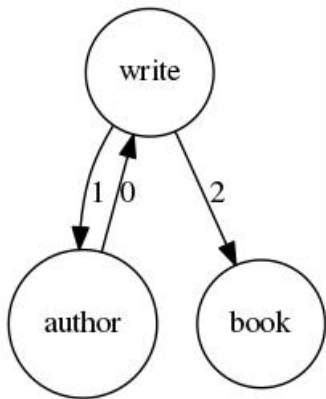
Szabályok

author(George Orwell, 1984)



Szabályok

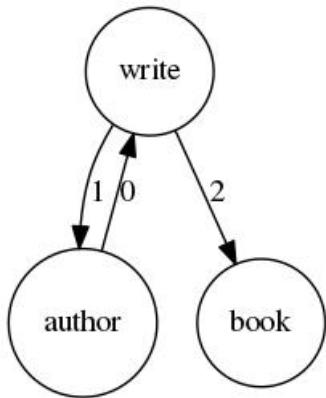
author(George Orwell, 1984)



- *mit ki/mi mivel*

Szabályok

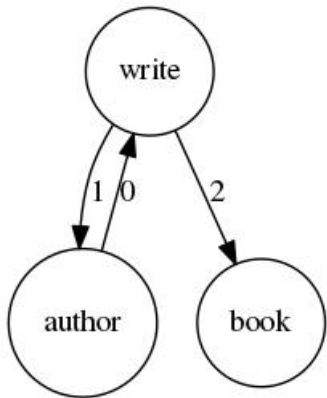
author(George Orwell, 1984)



- *mit ki/mi mivel*
- Wikidata: 86.3m triplet és 893 állítás

Szabályok

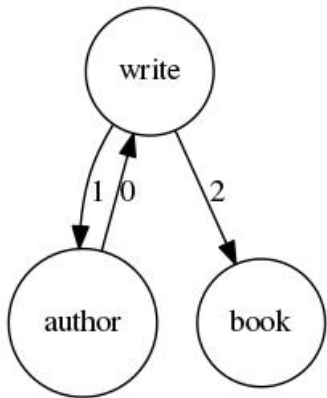
author(George Orwell, 1984)



- *mit ki/mi mivel*
- Wikidata: 86.3m triplet és 893 állítás
- Orwell $\xrightarrow{0}$ write

Szabályok

author(George Orwell, 1984)



- *mit ki/mi mivel*
- Wikidata: 86.3m triplet és 893 állítás
- Orwell $\xrightarrow{0}$ write
- 1984 $\xrightarrow{0}$ book

Eredmények

	1-szabály	2-szabály	teljes
állítmany	84	25	109
helyes	55	17	72
új tények	8.2 millió	0.83 millió	9 millió
helyes	7.6 millió	0.74 millió	8.3 millió
accuracy	0.92	0.89	0.92

Table: Evaluation results

Natural language inference

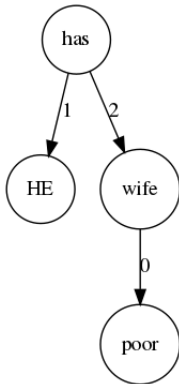
Natural language inference

My poor wife!

Natural language inference

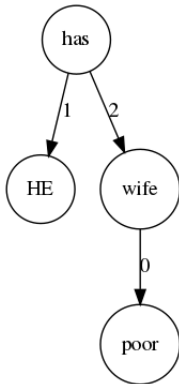
I feel bad for my wife!

My poor wife!

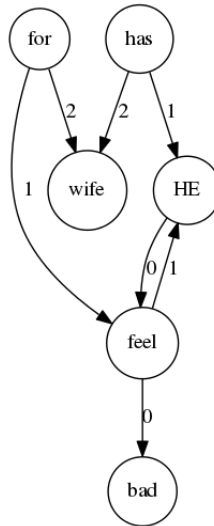


Natural language inference

My poor wife!



I feel bad for my wife!



Natural language inference

Natural language inference

- **Cél**
 - egy állítás következik-e egy feltevésből?
 - 2017 RepEval task

Natural language inference

- **Cél**
 - egy állítás következik-e egy feltevésből?
 - 2017 RepEval task
- **Megvalósítás**
 - Gráfok hasonlósága
 - Mikor hasonló két gráf?

Natural language inference

- **Cél**

- egy állítás következik-e egy feltevésből?
- 2017 RepEval task

- **Megvalósítás**

- Gráfok hasonlósága
- Mikor hasonló két gráf?

- My poor wife! (G1)

Natural language inference

- **Cél**

- egy állítás következik-e egy feltevésből?
- 2017 RepEval task

- **Megvalósítás**

- Gráfok hasonlósága
- Mikor hasonló két gráf?

- My poor wife! (G1)

- I feel bad for my wife!
(G2)

Natural language inference

- **Cél**

- egy állítás következik-e egy feltevésből?
- 2017 RepEval task

- **Megvalósítás**

- Gráfok hasonlósága
- Mikor hasonló két gráf?

- My poor wife! (G1)

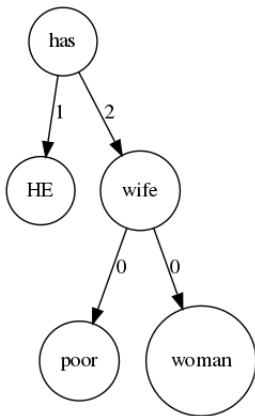
- I feel bad for my wife!
(G2)

-

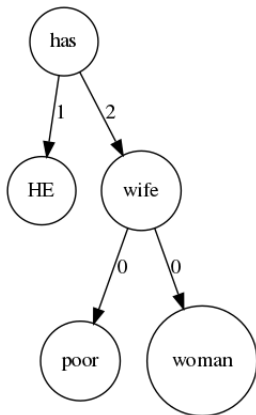
$$\frac{|E(G_1) \cap E(G_2)|}{|E(G_2)|}$$

Az "expand" és a "substitution" funkció

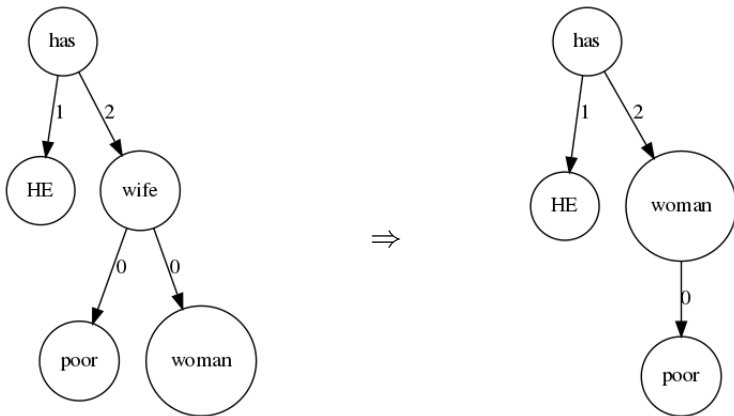
Az "expand" és a "substitution" funkció



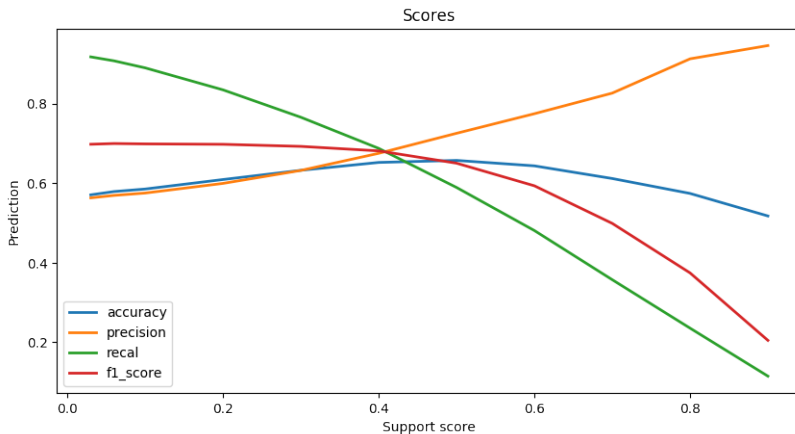
Az "expand" és a "substitution" funkció

 \Rightarrow

Az "expand" és a "substitution" funkció



Eredmények



A gépi szövegértés feladat (Chen et al., 2018; Wang et al., 2018)

Szöveg: "Today we decided to paint the extra room in our house. Were going to have visitor coming next month so hopefully the painting ain't that smelly anymore. I made sure that the wall is clean and clear of all the nuisance. We already bought the pain and we decided the new wall pain is sky blue. My husband is putting newspaper on the floor to avoid any spill on our floor/carpet....."

Kérdés: "Did anyone help her?"

1. **válaszlehetőség:** "It was the narrator and her husband"
2. **válaszlehetőség:** "No, she worked alone."

A gépi szövegértés feladat (Chen et al., 2018; Wang et al., 2018)

A gépi szövegértés feladat (Chen et al., 2018; Wang et al., 2018)

- **2018-as Semeval Task**
 - *Machine comprehension using commonsense knowledge*

A gépi szövegértés feladat (Chen et al., 2018; Wang et al., 2018)

- **2018-as Semeval Task**
 - *Machine comprehension using commonsense knowledge*
- Rövid szövegekből egyszerű feleletválasztós kérdések

A gépi szövegértés feladat (Chen et al., 2018; Wang et al., 2018)

- **2018-as Semeval Task**

- *Machine comprehension using commonsense knowledge*
- Rövid szövegekből egyszerű feleletválasztós kérdések

- Két legjobb rendszer

- HFL-RC
 - Yuanfudao

A gépi szövegértés feladat (Chen et al., 2018; Wang et al., 2018)

- **2018-as Semeval Task**
 - *Machine comprehension using commonsense knowledge*
- Rövid szövegekből egyszerű feleletválasztós kérdések
- Két legjobb rendszer
 - HFL-RC
 - Yuanfudao
- **MCScript**
 - tanító és teszt adat

Baseline

- Minden kérdés válasz párhoz mergelt gráf
- A "jobban" hasonló lesz a helyes válasz
- **68,3** accuracy score

Yuanfudao rendszer

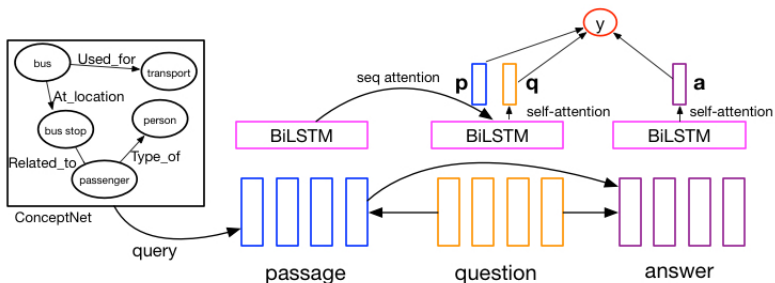


Figure: A rendszer struktúrája

Eredmények

modell	dev adat	teszt adat
előtanított TriAN, ConceptNet nélkül	83.7%	81.9%
előtanított TriAN, ConceptNettel	82.5%	80.3%
előtanított TriAN, 4langgal	84.2%	81.5%
előtanított TriAN, mindkettővel	83.4%	82.9%
TriAN, ConceptNet nélkül	82.8%	80.2%
TriAN, ConceptNettel	82.7%	80.5%
TriAN, 4langgal	83.2%	80.9%
TriAN, mindkettővel	83.1%	80.8%

Table: A 4lang és a ConceptNet hatása az eredményeken

Összefoglalás

Összefoglalás

- Automatizált módszer és metrikák hasonlóság mérésére.

Összefoglalás

- Automatizált módszer és metrikák hasonlóság mérésére.
- Erős baselineok különböző feladatokra

Összefoglalás

- Automatizált módszer és metrikák hasonlóság mérésére.
- Erős baselineok különböző feladatokra
- Baseline módszerünk alkalmazása state-of-the-art rendszerben, kis javulást hozva

Összefoglalás

- Automatizált módszer és metrikák hasonlóság mérésére.
- Erős baselineok különböző feladatokra
- Baseline módszerünk alkalmazása state-of-the-art rendszerben, kis javulást hozva
- A kód elérhető <https://github.com/adaamko/4lang>

Összefoglalás

- Automatizált módszer és metrikák hasonlóság mérésére.
- Erős baselineok különböző feladatokra
- Baseline módszerünk alkalmazása state-of-the-art rendszerben, kis javulást hozva
- A kód elérhető <https://github.com/adaamko/4lang>
- Éles Rest API és online demo
<http://4lang.hlt.bme.hu>

Köszönöm a figyelmet!

- Chen, Zhipeng et al. (2018). "HFL-RC System at SemEval-2018 Task 11: Hybrid Multi-Aspects Model for Commonsense Reading Comprehension". In: *arXiv preprint arXiv:1803.05655*.
- Kornai, András (2010). "The algebra of lexical semantics". In: *Proceedings of the 11th Mathematics of Language Workshop*. Ed. by Christian Ebert, Gerhard Jäger, and Jens Michaelis. LNAI 6149. Springer, pp. 174–199.
- Kornai, András et al. (2015). "Competence in lexical semantics". In: *Proceedings of the Fourth Joint Conference on Lexical and Computational Semantics (*SEM 2015)*. Denver, Colorado: Association for Computational Linguistics, pp. 165–175.
- Kovács, Ádám and Kinga Gémes (2018). *Semantic parsing with graph transformations*. Tech. rep. Budapest: Budapest University of Technology and Economics. URL: <http://tdk.bme.hu/VIK/DownloadPaper/Szemantika-elemzes-graftranszformaciokkal>.
- Kovács, Ádám and Gábor Recski (2018). "Knowledge base population using natural language inference". In: *AACS*.
- Wang, Liang et al. (2018). "Yuanfudao at SemEval-2018 Task 11: Three-way Attention and Relational Knowledge for Commonsense Machine Comprehension". In: *arXiv preprint arXiv:1803.00191*.