

Névelemfelismerés vs. entitáskinyerés - módszerek és alkalmazások

Pancza Judit

conTEXT 2023

2023.11.14.



Clementine

NLP MEGOLDÁSOK

Magyar nyelvű
szöveganalitika

2007

2009

2013

2015

2016

2020

2022

Banki, biztosítói
ügyfélszolgálati megoldások

Hálózatvizualizációt
(i2) és az
adatelemzést ötvöző
megoldások
implementációja

Folyamatos
support,
támogatás

Komplex
rendszerek
kialakítása

Entitáskinyerés
rendvédelmi
területen

Magyar nyelvű
hanganalitika

NÉVELEM FELISMERÉS

„Named Entity Recognition is a commonly used Natural Language Processing task. NER is the process of identifying and categorizing words, expressions, or names in unstructured data into predefined categories such as persons, organizations, and locations.”



Rendvédelem

Entitások és rejtett
kapcsolatok
Bűnszövegek
felderítése



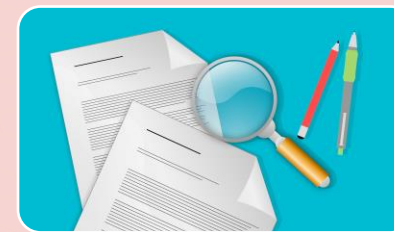
Csalásfelderítés

Entitások és rejtett
kapcsolatok



Ügyfélszolgálat

Beérkező levelek
feldolgozásának
támogatása
Ügyféladatok
kigyűjtése



NLP feladatok

Topic modellezés
Szentiment elemzés

NÉVELEM FELISMERÉS

- Tulajdonnevek vs. névelemek
- Kigyűjtés és osztályba sorolás
- Szótár és szabály alapú megközelítés vs statisztikai tanulás

...ügynöki források azt gyanítják, hogy a Bögre Béla
(Kaposvár, 1971.03.20., anyja neve: Nagy Veronika)
otthonát drogelosztó pontként használja.

...ügynöki források azt gyanítják, hogy a [PER] Bögre Béla [LOC] (Kaposvár,
1971.03.20., anyja neve: [PER] Nagy Veronika) otthonát drogelosztó pontként
használja.

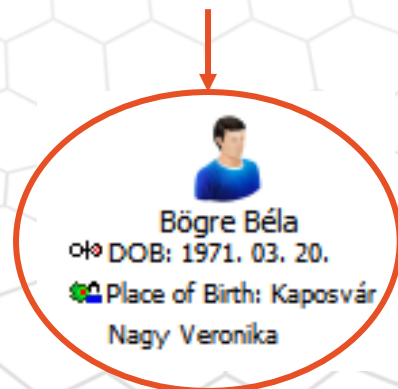
<https://juniper.nytud.hu/demo/ner>

NÉVELEM FELISMERÉS VS ENTITÁS AZONOSÍTÁS

...ügynöki források azt gyanítják, hogy a Bögre Béla (Kaposvár, 1971.03.20., anyja neve: Nagy Veronika) otthonát drogelosztó pontként használja.

...ügynöki források azt gyanítják, hogy a **[PER] Bögre Béla** **[LOC] (Kaposvár,** 1971.03.20., anyja neve: **[PER] Nagy Veronika)** otthonát drogelosztó pontként használja.

<https://juniper.nytud.hu/demo/ner>



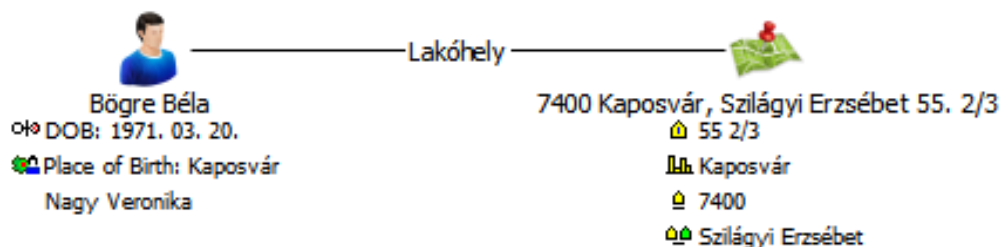
ENTITÁSAZONOSÍTÓ RENDSZER SZINTJEI

Szakértői szint

- Célok
- Mi az entitás?
- Mikor azonos két entitás?



Tervezés

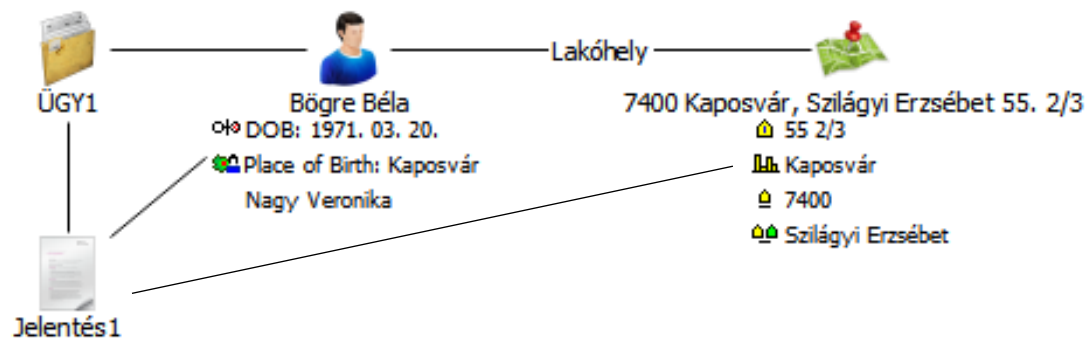


Technikai szint

- NLP
- Egyértelműsítés
- Azonosság vizsgálat



Megvalósítás



Módszertani szint

- Új eszközök, új elemzési módszertan
- Nem közvetlenül NLP, de érteni a rendszer működését



Alkalmazás

Módszerek, alkalmazások, példák



NÉVELEM FELISMERÉS – SZÓTÁR + SZABÁLYOK

hagar	Entire and End	<input type="checkbox"/>	magyar_noi	utonev
hajnal	Entire and End	<input type="checkbox"/>	magyar_noi	utonev
hajnalka	Entire and End	<input type="checkbox"/>	magyar_noi	utonev
hajni	Entire and End	<input type="checkbox"/>	magyar_noi	utonev
hajnácska	Entire and End	<input type="checkbox"/>	magyar_noi	utonev
hanga	Entire and End	<input type="checkbox"/>	magyar_noi	utonev
hanna	Entire and End	<input type="checkbox"/>	magyar_noi	utonev
hannadórá	Entire and End	<input type="checkbox"/>	magyar_noi	utonev
hannaliza	Entire and End	<input type="checkbox"/>	magyar_noi	utonev
hannaróza	Entire and End	<input type="checkbox"/>	magyar_noi	utonev

Name:

Example:

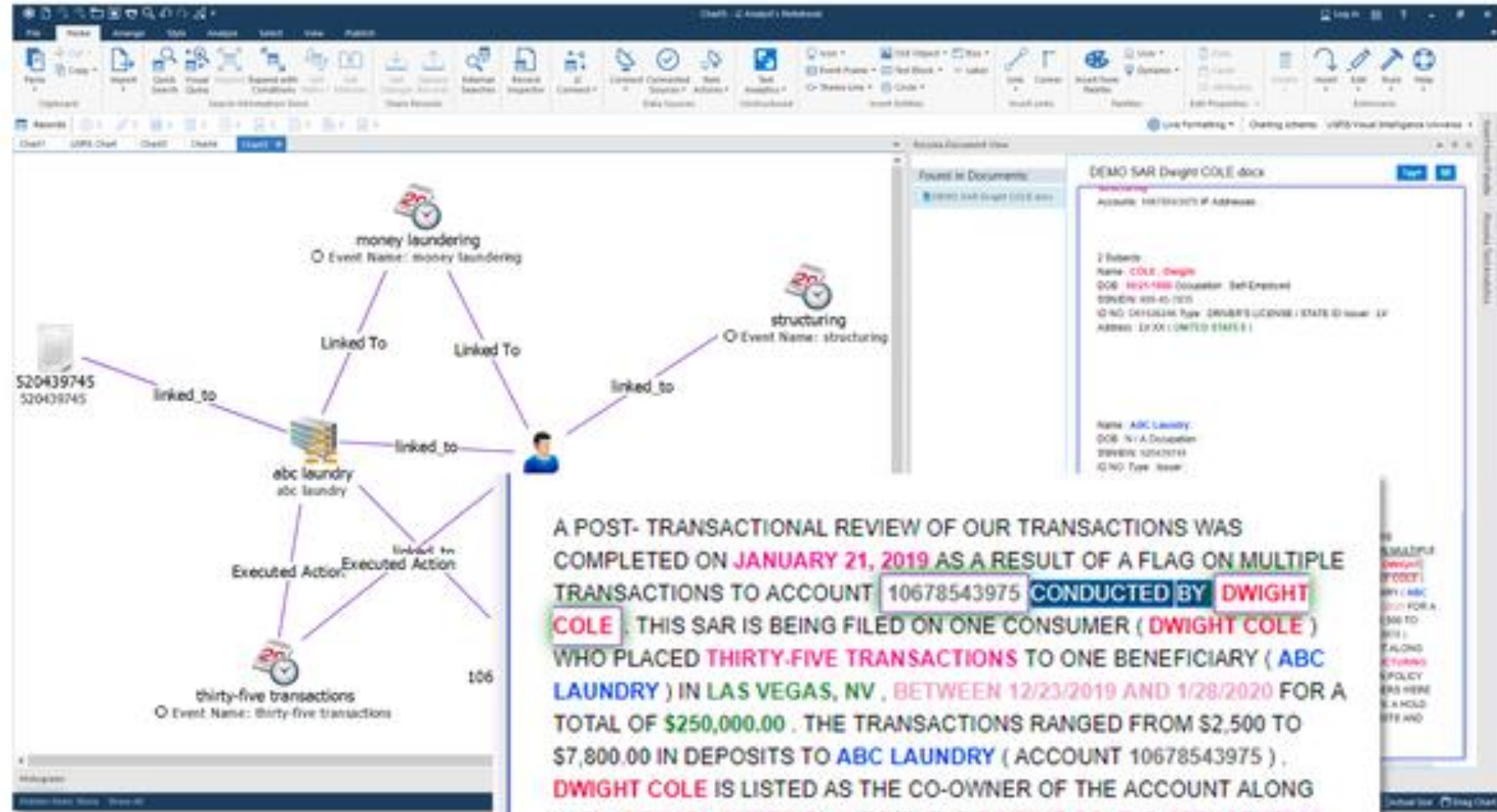
Rule Value table:

	Element	Quantity	Example Token
1	magyar_nev	Exactly 1	bögre béla
2		Between 0 and 2	?
3	telepules	Exactly 1	kaposvár
4		Between 0 and 2	?
5	Date	Exactly 1	03/20/1971
6			

NÉVELEM FELISMERÉS – SZÓTÁR + SZABÁLYOK

- **5 szótár** a különböző adatokhoz (pl. hely, autó, utónevek, címek)
- **21 stream és script**, amely a felismerést végzi
- **200+ szabály** a kapcsolatok feltárására
- **1000+ reguláris kifejezés** a rendszámok, telefonszámok, azonosítók kigyűjtésére
- **27 ezer elemű szótár**, ami a személy azonosítást és név tisztítást végzi

I2 TEXT CHART



NLP RENDSZER, DE..

- Egy komplex rendszerben nem elég az NLP képesség
- Szakértői és NLP szoros együttműködés
- Elemzési módszertan kialakítás



KÖSZÖNÖM A FIGYELMET!

Pancza Judit
jpancza@clementine.hu

