

Przetwarzanie języka naturalnego

Natural language processing (NLP)

Tematyka prac. Propozycje badawcze.

Agnieszka Mykowiecka, Alina Wróblewska



INSTITUTE OF COMPUTER SCIENCE
POLISH ACADEMY OF SCIENCES
ul. Jana Kazimierza 5, 01-248 Warszawa

Warszawa, listopad 2018

- giving computers the ability to process human language
- NLP aims at **making computers talk and communicate**, or more precisely, at endowing them with the linguistic ability of humans
- NLP encompasses anything a computer needs to understand natural language (text or speech) and also generate the natural language
- reproduce the natural transmission of information by modelling the speaker's production and the hearer's interpretation on a suitable type of computer



- NLP obejmuje badania z wielu dziedzin: lingwistyki, informatyki, matematyki, statystyki, psychologii, kogniwyistyki ...

Goals of Artificial Intelligence



Reasoning



Automated Learning & Scheduling



Machine Learning



Natural Language Processing



Computer Vision



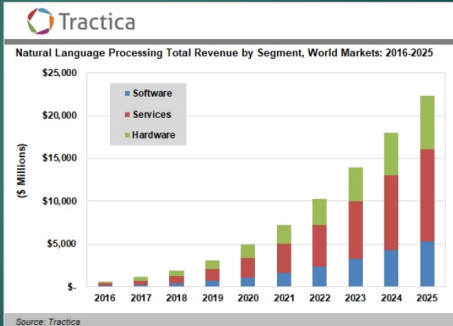
Robotics



General Intelligence

Źródło: <https://www.upwork.com/hiring/for-clients/artificial-intelligence-and-natural-language-processing-in-big-data/>

Natural Language
Processing Market to
Reach \$22.3 Billion
by 2025



Źródło: The comercial NLP landscape. Robert Dale. RANLP 2017.

Sponsorzy konferencji ACL 2017

Platynowi

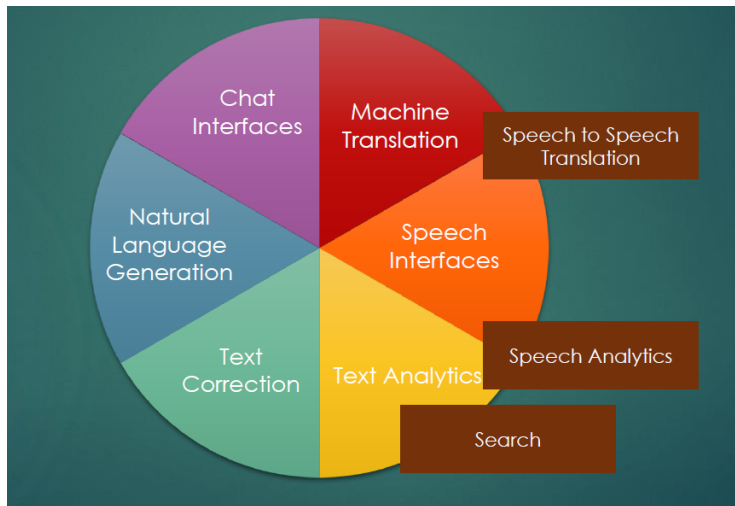


SAMSUNG

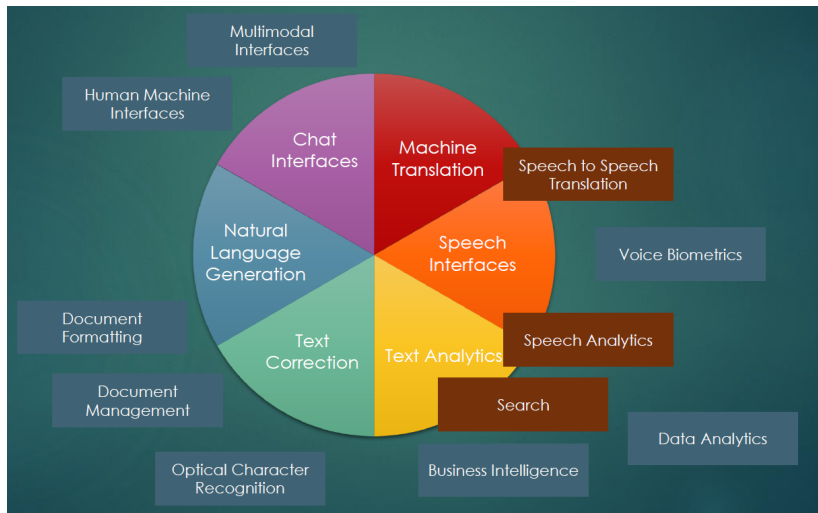
Złoci



<http://acl2017.org/sponsors/overview/>



Źródło: The comercial NLP landscape. Robert Dale. RANLP 2017.



Źródło: The comercial NLP landscape. Robert Dale. RANLP 2017.

- segmentacja – jak poprawnie podzielić tekst?
- analiza morfologiczna – jakie słowa możemy rozpoznać w tekście?
- analiza składniowa – jaka jest struktura tekstu?
- analiza semantyczna – jaki jest sens/znaczenie danego tekstu?
- analiza pragmatyczna – jakie jest znaczenie tekstu w szerszym kontekście?

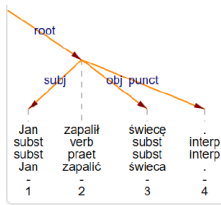
- metody formalne
 - gramatyki formalne opisujące składnię i/lub semantykę
 - logiki, zasady wnioskowania
- metody statystyczne/maszynowe uczenie się
 - praktycznie każda metoda, w szczególności metody etykietowania sekwencyjnego
- sieci neuronowe
- metody hybrydowe, np.
 - gramatyki formalne+statystyka
 - sieci neuronowe+statystyka

segmentacja, analiza morfologiczna

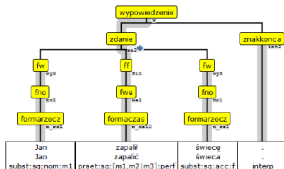
0	1	Jan	Jan:s2	subst:sg:nom:m1	nazwisko
			Jan:s3	subst:sg:pl:nom.gen.dat.acc.inst.loc.voc:f	nazwisko
			Jana:s4	subst:pl:gen:f	nazwisko
		Ewangelia_świętego_Jana	brev:npun		
		Jan:s1	subst:sg:nom:m1	imię	
1	2	zapalił	zapalić	praet:sg:m1.m2.m3:perf	
2	3	świecę	świeca	subst:sg:acc:f	nazwa pospolita
			świecić	fin:sg:pri:imperf	
3	4	.	.	interp	

analiza składniowa

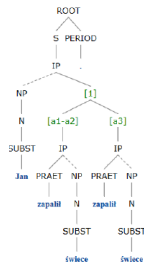
drzewo zależnościowe



drzewo składnikowe



struktura LFG C-structure

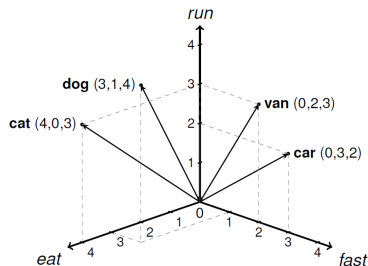


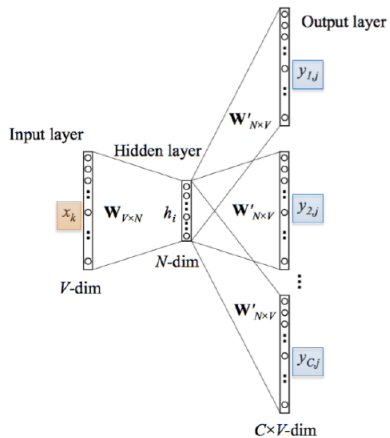
List [Europejskiej Akademii Filmowej]_{ORG} do ministra [Glińskiego]_{PER}

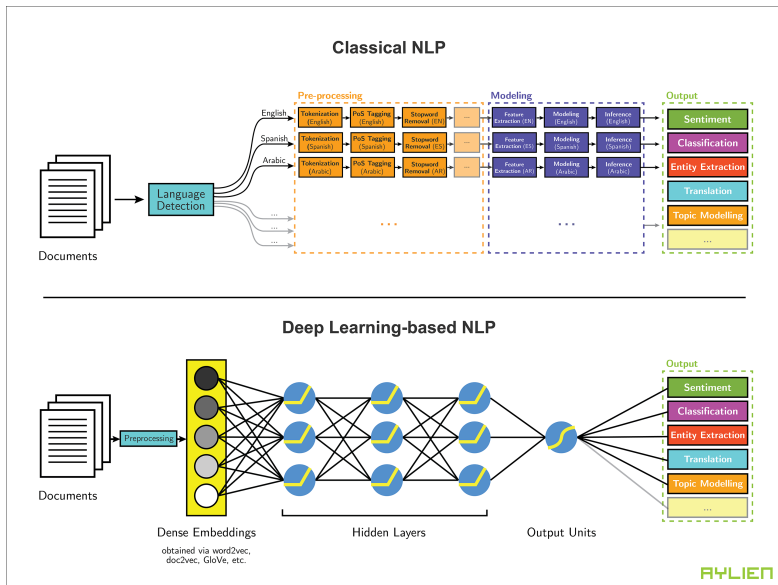
word	POS	capitalisation	chunk	entity type
List	subst	yes	I-NP	O
Europejskiej	adj	yes	I-NP	I-ORG
Akademii	subst	yes	I-NP	I-ORG
Filmowej	adj	yes	I-NP	I-ORG
do	prep	no	I-PP	O
ministra	NNP	no	I-NP	O
Glińskiego	adj	yes	I-NP	I-PER

	<i>dog</i>	<i>drive</i>	<i>eat</i>	<i>fast</i>	<i>play</i>	<i>...</i>	<i>the</i>	<i>wheel</i>
car	0	3	0	2	0	$\vdots$	2	1
cat	1	0	3	0	1	$\vdots$	2	0
dog	0	0	3	0	2	$\vdots$	2	0
van	0	3	0	1	0	$\vdots$	3	1

co-occurrence matrix







- zasoby
 - gromadzenie, opracowywanie, ujednolicanie, oznaczanie, weryfikacja
- narzędzia podstawowe:
 - segmentacja, oznaczanie części mowy, identyfikacja nazw własnych ...
- aplikacje realizujące konkretne zadania, np.
 - system odpowiedzi na pytania,
 - streszczanie wielodokumentowe,
 - upraszczanie tekstu
 - analiza dyskursu (wykrywanie relacji między zdaniem pojedynczymi)

NKJP

1.	przez ten czas nie dała	<u>odczuć</u> [odczucie:subst:pl:gen:n]	gościowi niechęci. "Dla
2.	jest w stanie lepiej ją	<u>odczuć</u> [odczucie:subst:pl:gen:n]	, zrozumieć, przekazać i
3.	i zanim duch cokolwiek znowu	<u>odczuć</u> [odczucie:subst:pl:gen:n]	zdola, zmysły okropnie mnie
4.	laptopami, płaskimi komórkami płaskich	<u>odczuć</u> [odczucie:subst:pl:gen:n]	ułożonych na płask w spłaszczonych
5.	, że jest w stanie	<u>odczuć</u> [odczucie:subst:pl:gen:n]	kamienną niecierpliwość, choć nawet
6.	bardzo blisko mnie, żeby	<u>odczuć</u> [odczucie:subst:pl:gen:n]	ten nieustannie przepływający przeze mnie
7.	. Właśnie pozwoliłam sobie	<u>odczuć</u> [odczucie:subst:pl:gen:n]	głód rodziny, jej ograniczeń
8.	by choć rąbkami świadomości nie	<u>odczuć</u> [odczucie:subst:pl:gen:n]	przez mgnienie swojej przewagi nad
9.	bezduszny, obojętny człowiek mógł	<u>odczuć</u> [odczucie:subst:pl:gen:n]	jego wściekłość. Jego nienawiść
10.	nie chciałem negocjować pani	<u>odczuć</u> [odczucie:subst:pl:gen:n]	i życiowych doświadczeń. Zastanawiał

Sejm

1.	przekonany, iż będę wyrazicielem	<u>odczuć</u> [odczucie:subst:pl:gen:n]	wielu pań i panów posłów
2.	. W Europie daje się	<u>odczuć</u> [odczucie:subst:pl:gen:n]	wrogość i uprzedzenia do chrześcijan
3.	A pacjenci szybko powinni to	<u>odczuć</u> [odczucie:subst:pl:gen:n]	, tak jak w już
4.	NATO. Nie sposób nie	<u>odczuć</u> [odczucie:subst:pl:gen:n]	pokory, kiedy porównujemy to
5.	w Europie. Mogę to	<u>odczuć</u> [odczucie:subst:pl:gen:n]	w wielu miejscach, w
6.	rzeczywiście w Polsce można było	<u>odczuć</u> [odczucie:inf:perf]	wyraźną poprawę w stosunkach pracy
7.	, to polskie przedsiębiorstwa mogą	<u>odczuć</u> [odczucie:inf:perf]	znaczne pogorszenie się sytuacji,
8.	kryzysu dają się coraz bardziej	<u>odczuć</u> [odczucie:inf:perf]	również w naszej gospodarce.
9.	To jest kwestia bardzo subiektywnych	<u>odczuć</u> [odczucie:inf:perf]	. Skoro jednak i telewizja
10.	tej sali, można było	<u>odczuć</u> [odczucie:subst:pl:gen:n]	, jak potężną siłą jest

- Lemmatization of proper names and multi-word phrases
- Temporal expressions
- Word sense disambiguation/identification
- Sentiment analysis
- Coreference resolution
- Entity linking
- Cyberbullying Detection
- ... <http://clip.ipipan.waw.pl/benchmarks>

wąskich dróg leśnych → wąski droga leśny
wąska droga leśna

- dróg oddechowych → drogi oddechowe
- ulicy Białej → ulica Biała
- Bibliotece Wyższej Szkoły Zarządzania →
Biblioteka Wyższej Szkoły Zarządzania
- Bibliotece Głównej Politechniki Warszawskiej →
Biblioteka Główna Politechniki Warszawskiej

Žarko Čabarkapa

Žarko Čabarkapa, (ur. ⁿ¹21 maja 1981 roku w Belgradzie), serbski koszykarz, występujący w barwach KK Budućnost Podgorica, reprezentant Serbii. Ma 211 cm wzrostu, waży 106 kg.

Kariera

Przed ²przebyciem do NBA występował ^{n2 n4 n3 n5 n6} w serbskich drużynach Bleopetrol (^{3 4}1998 - ⁰¹) i Budućnost (⁵2001 - ⁰³). W ⁷2003 roku został wybrany ^{n7 n2} przez Phoenix Suns z numerem 17. w drafcie. W swoim pierwszym sezonie w NBA w barwach Suns ⁸wystąpił w 49 meczach ⁿ⁸zdobywając średnio 4.1 punktu na mecz. W ⁹drugim roku gry w NBA (¹⁰2004 - ⁰⁵) ¹¹rozegrał ^{n8 n9 n10} w barwach Suns zaledwie 3 mecze po czym ¹²przeniósł ⁿ¹¹ się do Golden State Warriors. Tam ¹³wystąpił ⁿ¹² aż w 77 spotkaniach ⁿ¹³zdobywając 6.1 punktu na mecz. W ¹⁴sezonie 2005 - ⁰⁶ w barwach Warriors ¹⁵rozegrał ^{n13 n14} 61 spotkań ⁿ¹⁵zrzucając 3.3 punktu na mecz. ¹⁶16 stycznia 2009 ⁿ¹⁶podpisał kontrakt z KK Budućnost Podgorica.

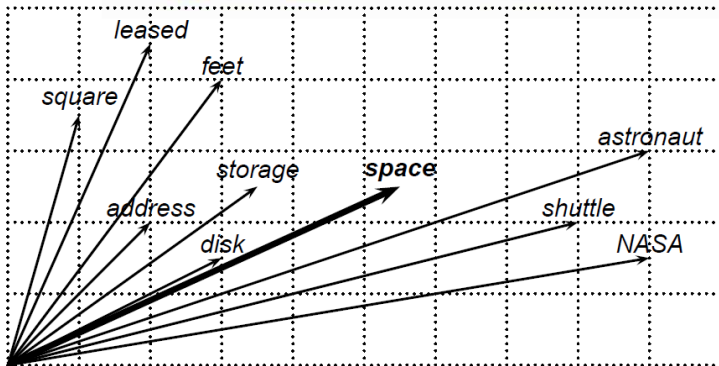
- identyfikacja fraz odnoszących się do czasu i ich typu
 - data, godzina, ...
- identyfikacja relacji
 - przed, po, w trakcie, ...
- ustalenie bezwzględnego określenia punktu/okresu czasu
 - rok temu → listopad 2017

Rozpoznawanie nazw własnych

<http://multiservice.nlp.ipipan.waw.pl/>



- 1 Nie udało mi się , musiałem zostawić puste miejsce – powiedział Onetowi tuż po wyjściu z sali Daniel Moczydłowski , uczeń
- XXIV Liceum Ogólnokształcącego im . C . K . Norwida w Warszawie .



Przykład – Sounding Board

zwycięzcy konkursu Amazona na aplikację dla Alexy, 2017

Turn	User	Sounding Board
1	Let's chat	Hi! This is an Alexa Prize socialbot. How's your day?
2	I'm <u>five</u> .	Glad to hear it! We could talk about robots, Batman, or Superman. Say "next", to chat about other things.
3	Superman.	It looks like you wanna chat about Superman. I was high up in the cloud when I realized: If people don't recognize Clark Kent as Superman because of his glasses, does his eye doctor know his true identity?
4	I guess so.	Did you know that Henry Cavill almost missed the call for the role of Superman cause he was playing "World of Warcraft"
5	Really, I didn't know that.	Weird, right? Speaking of Superman, Did you see: Superman? It was released in 1997. It's a comedy.
6	Yes, it was hilarious.	I'm glad you feel this is hilarious. Which part do you like best about this movie?
7	The part when he met <u>Lewis Leah</u> .	Interesting. Meccartin and Raffi co-directed this film. The movie has a 6.3 out of 10 on IMDB. which seems pretty good! do you like the movie's director?

- Instytut Podstaw Informatyki PAN, ul. Jana Kazimierza 5
- Strona IPI PAN <https://ipipan.waw.pl>
- Strona ZIL <http://zil.ipipan.waw.pl>
- Agnieszka Mykowiecka – agn@ipipan.waw.pl
- Alina Wróblewska – alina@ipipan.waw.pl

- Instytut Podstaw Informatyki PAN, ul. Jana Kazimierza 5
- Strona IPI PAN <https://ipipan.waw.pl>
- Strona ZIL <http://zil.ipipan.waw.pl>
- Agnieszka Mykowiecka – agn@ipipan.waw.pl
- Alina Wróblewska – alina@ipipan.waw.pl

Dziękuję za uwagę!