Politechnika Wrocławska

Automatyczne rozpoznawanie polskich leksykalnych relacji derywacyjno-semantycznych

Maciej Piasecki, Marek Maziarz,
Radosław Ramocki, Paweł Minda

Grupa Naukowa G4.19

Instytut Informatyki PWr.

Słowosieć (plWordNet) 2.0

- Słowosieć
  - Inaczej: plWordNet
  - 2005: rozpoczęcie prac
  - 2009: wersja 1.0

    26990 jednostek leksykalnych

    17695 synsetów

PLWORDNET
SŁOWOSIEĆ

- www.plwordnet.pwr.wroc.pl

Słowosieć (plWordNet) 2.0

PLWORDNET SŁOWOSIEĆ

- plWordNet 2.0 (2010-2012)
  - plan:
  - 135000 jednostek leksykalnych, 90000-100000 synsetów
  - zakres:
  - czasowniki, przymiotniki, relacje derywacyjne
  - stan obecny:
  - 135300 jednostek leksykalnych, 96 700 synsetów
  - N N516 068637 finansowany przez **MNiSW**
  - POIG.01.01.02-14-013/09 **UE oraz MNiSW**
  - SyNaT projekt strategiczny – **NCBiR**

NARODOWA STRATEGIA SPÓJNOŚCI

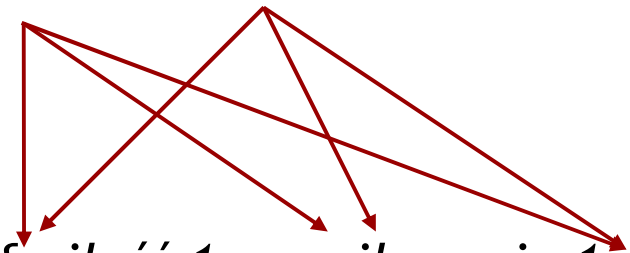
UNIA EUROPEJSKA
EUROPEJSKI FUNDUSZ ROZWOJU REGIONALNEGO

Słowosieć: podstawowe założenia

- Jednostka leksykalna jako element centralny struktury
  - lingwistyczne relacje leksykalne są definiowane dla jednostek leksykalnych, a nie dla pojęć
  - „zleksykalizowane pojęcie” jest nieuchwytne
  - arbitralność relacji między pojęciami i problem spójności
  - we wszystkich wordnetach występują relacje pomiędzy jednostkami leksykalnymi
  - pomocnicze testy podstawieniowe (EuroWordNet) dotyczą jednostek leksykalnych
- Reprezentacja struktury znaczeń leksykalnych oparta na zasadzie minimalnych założeń

Słowosieć: podstawowe założenia

- Rola synsetu
  - grupuje jednostki leksykalne o wspólnych relacjach hiperonimii i holo/meronimii
  - jednostki zawarte w synsecie są uznawane za synonimy
  - jest rodzajem skrótu notacyjnego, np.
  - *{afekt 1, uczucie 2}* —hiperonim→



  - *{miłość 1, umiłowanie 1, kochanie 1}*
- Relacje konstytutywne
  - podstawa konstrukcji synsetu
- Dodatkowe rozróżnienia: rejestr stylistyczny, aspekt

Słowosieć: podstawowe założenia

- Przedmiotem opisu jest język polski
  - wierność opisu
  - porównanie z WordNetem nie było wykorzystywanym kryterium
- Ogólny kierunek procesu budowy:
  - od opisu lematów do struktury całościowej
- Podział na bazy dla poszczególnych części mowy:
  - rzeczowniki, czasowniki, przymiotniki
- Półautomatyczna konstrukcja oparta na wielkim korpusie języka polskiego

Słowosieć: relacje łączące synsety

- Hipo/hiperonimia
  - zdefiniowana dla rzeczowników i czasowników
    - {*audycja 1*, *słuchowisko 1*} —hipo→ {*program 5*}
    - {*wezwać 1*, *zaapelować 1*} —hipo→ {*poprosić 1*, *zwrócić się z prośbą 1*}
  - ale również dla przymiotników
    - {*błękitny 1*} —hipo→ {*niebieski 1*}
    - {*plastyczny 1*} —hipo→ {*artystyczny* 1}
    - {*słowiański*} —hipo→ {*europejski 1*}
    - {*nieorganiczny 1*} —hipo→ {*chemiczny 1*}

Słowosieć: relacje łączące synsety

- Meronimia
  - podtypy: część, element kolekcji, materiał, porcja,
    - {*iskra 1*, *skierka 1*, *skra 1*, *iskierka 1*} —m.cz.→ {*ogień 1*}
    - {*kula 4*, *nabój 2*} —m.e.k.→ {*amunicja 1*}
    - {*drewno 1*, *drzewo 1*} —m.mat.→ {*stolarka 2*}
    - {*akropol 1*} —m.msc.→ {wzgórze *1*, *pagórek 1*, *wzgórek 1*}
    - {*blacha 1*} —m.p.→{*metal 2*}
- Holonimia
  - podtypy: identyczne jak dla meronimii
    - {*kula 4*, *nabój 2*} —h.cz.→ {*łuska 1*, *gilza 1*}
    - {*bydło 2*} —h.e.k.→ {*krowa 1*}
    - {*deszczowiec 1*} —h.m.→ {*tkanina 1*}
  - Nie są relacjami automatycznie odwrotnymi
    - np. *szprycha* —m.cz.→ *koło*

Słowosieć: relacje łączące synsety

- Bliskoznaczność
  - synonimia międzyrejestrowa – łącząca jednostki z istotnie różnych rejestrów stylistycznych
  - współdzielą hiperonim, ale nie współdzielą hiponimów

    np. {*adorator 1*} ←blsk.→ {*zalotnik 1*}

    ←blsk.→ {*absztyfikant 1*}

    {*apollo 2*} ←blsk.→ {*przystojniak 1*}
- Fuzzynimia
  - widoczne powiązanie semantyczne, ale niepasujące do żadnej z pozostałych relacji
- Typ/instancja
  - nazwy własne z synsetami zawierającymi rzeczowniki pospolite

Relacje pomiędzy czasownikowymi synsetami

- *Hipero/hiponimia*
- *Meronimia* i *holonimia*
- Kauzacja
  - *poić* > *pić*
- Procesywność (V-A, V-N)
  - *zaczerwienić się* > *czerwony*
- Inchoatywność
  - *ruszyć* > *poruszać się*
- Stanowość
  - *czerwienieć* > *czerwony*
- Wielokrotność
  - iteratywność
  - dystrybutywność
- *pootwierać* > *otwierać*
- Presupozycja
- Uprzedniość
- Fuzzynimia

Słowosieć: relacje leksykalne

- Synonimia — wyrażana przez synsety
- Antonimia
  - antonimia komplementarna
    - np. *mężczyzna 1* ↔ *kobieta 1*
  - antonimia właściwa
    - rzeczowniki, np. *blask 1* ↔ *cień 1*
    - czasowniki, np. *przedzielić 1* ↔ *połączyć 1*
    - przymiotniki, np. *akustyczny 2* ↔ *dźwiękoszczelny 1*
- Konwersja
  - przeciwieństwo argumentów lub ról – ujmowanie tej samej sytuacji z dwóch różnych punktów widzenia
  - rzeczowniki, np. *biorca 1* ↔ *dawca 1*
  - czasowniki, np. *kupić 2* ↔ *sprzedać 1*

Problem

- Relacje derywacyjne
  - opisywane w wielu wordnetach
  - leksykalne relacje semantyczne oparte na relacjach derywacyjnych
  - podejmowane próby automatyzacji ich opisu
  - w większości oparte na ręcznie skonstruowanych regułach
  - brak narzędzi do automatycznego rozpoznawania relacji derywacyjnych dla wielu języków, w tym języka polskiego

Problem

- Idea
  - oparcie procesu półautomatycznego rozszerzania wordnetu w zakresie relacji derywacyjnych o schemat sprzężenia zwrotnego (tzw. bootstrapping)
  - wejście: ograniczony zbiór par (instancji relacji) opisanych ręcznie
  - wytrenowanie generatora par
  - zastosowanie go do wsparcia procesu rozszerzania wordnetu

Relacje derywacyjne i motywowane derywacyjnie

- Relacja pomiędzy derywatem i jego bazą derywacyjną
- W niektórych przypadkach rozszerzone do leksykalnych relacji semantycznych *sensu stricto*
- Jasne znaczenie
- Istotne w definiowaniu jednostek leksykalnych i synsetów
- Tylko najczęstsze zjawiska w języku polskim
- Najbardziej produktywne schematy derywacyjne
- Ponad 100 000 par w obecnej wersji Słowosieci

Relacje derywacyjne

- **Synonimia międzyparadygmatyczna**

  ~ `NEAR' relations in EWN

  `transpozycyjna' (`składniowa') derywacja

  zmiana części mowy bez żadnej istotnej zmiany znaczenia

  pis-anie (odsłownik) < *pisać* (N-V)

  czerwon-ość < *czerwony* (N-Adj)

Relacje derywacyjne

- **Nosiciel stanu|cechy** (N-Adj)

  ślepi-ec < ślep-y

  starz-ec < *star-y*

  relacja odwrotną jest **stan |cecha**

- **Żeńskość** (N-N)

  żeńskie derywaty od męskich podstaw derywacyjnych

  pisar-ka < *pisarz*

  kot-ka < *kot*

Relacje derywacyjne

- **Nacechowanie** (N-N)

deminutywność (mały, niewielki, miły)

dom-ek < dom

książecz-ka < książka

ekspresywność i augmentatywność (duży, wielki, okropny)

ptasz-ysko < ptak

noch-al < nos

istota młoda (ludzie i zwierzęta)

koci-ę < kot

Relacje derywacyjne

- **Rola (**semantyczne**)** (N < V)

  Podobna do EWN Role oraz Involved Relations

  Agens: *pływ-ak* < *pływ-ać*

  Pacjens | obiekt: *uczeń* < *uczyć*

  Narzędzie | instrument: *pis-ak* < *pis-ać*

  Miejsce: *suszarnia* < *suszyć*

  Wytwór | rezultat: *układ* < *układać*

  Czas: *świt* < *świtać*

  Podtyp nieokreślony: *artykulacja* < *artykułować*

Relacje derywacyjne

- **Zawieranie roli** (V-N)

  Agens: *filozofować* < *filozof*

  Pacjens | obiekt : *kartk-ować* < *kartka*

  Narzędzie | instrument : *pieprz-yć* < *pieprz*

  Miejsce: *aresztować* < *areszt*

  Wytwór | rezultat: *detronizować* < *detronizacja*

  Czas: *ucztować* < *uczta*

  Podtyp nieokreślony: *cwałować* < *cwał*

Relacje derywacyjne

- **Rola przy niewyrażonym predykacie**
  - rzeczowniki, które wypełniają role choć nie łączy ich relacja derywacyjna z czasownikiem, a z rzeczownikiem
  - **Agens**: *gołębiarz* < *gołąb*
  - **Wytwór | rezultat**: *kapuśniak* < *kapusta*
  - **Miejsce**: *kwiaciarnia* < *kwiat*

Relacje motywowane derywacyjnie

- Relacje derywacyjne rozszerzone do leksykalnych relacji semantycznych
- Pomiędzy rzeczownikami:
  - **Mieszkaniec**
- Pomiędzy czasownikami:
  - inchoatywność, stanowość, procesywność, kauzacja, iteratywność, dystrybutywność

Relacje motywowane derywacyjnie

- Mieszkaniec (N-N)

  {Trojańczyk, Trojanin} – {Troja, Ilion}

  Troj-ańczyk, *Troj-anin* < *Troja*

  Troj-ańczyk, *Troj-anin* ~~<~~ *Ilion* (ale przechodzą test)

- Inchoatywność (V-V) `początek sytuacji'

  {wstać, podnieść się, powstać} – {stać}

  w-stać < *stać*

  powstać, *podnieść się* ~~<~~ *stać* (ale przechodzą test)

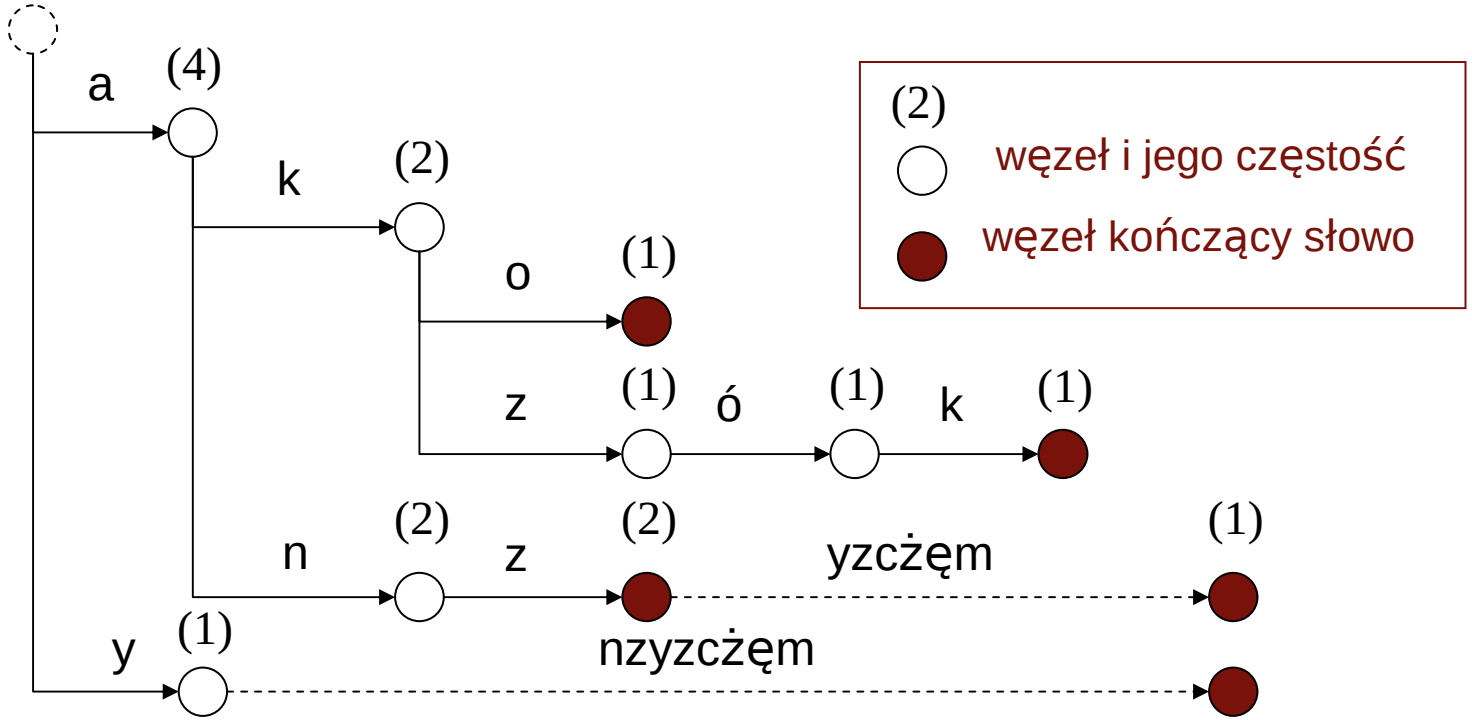
Derywator

- Działanie
  - *wejście*: forma wyrazowa (potencjalny derywat)
  - *wyjście*: podstawa derywacyjna + typ relacji derywacyjnej
- Konstrukcja – idea
  - wykorzystanie *Odgadywacza* rozpoznającego opis morfologiczny po *końcówce*
  - podwójny transduktor – podwójny *Odgadywacz*
  - automatyczne rozpoznanie wymian wewnątrztematowych

Derywator: uczenie *Odgadywacza*

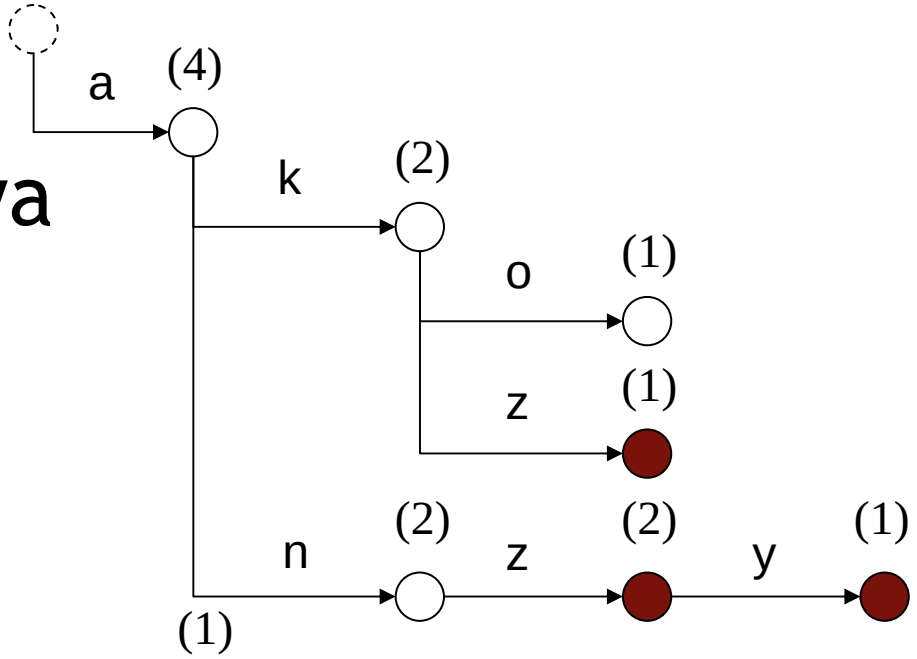
- *Faza 1*: konstrukcja drzewa transduktora
  - zbiór wejściowy: ⟨forma wyrazowa, znacznik, lemat⟩

  odwrócone formy wyrazowe ⇒



Derywator: uczenie *Odgadywacza*

- Reguły rekonstrukcji lematów
  - długość końcówki formy wyrazowej + końcówka lematu
- *Faza 2*: przycinanie drzewa



Derywator: uczenie

- **Wejście:**
  - L = ⟨*derywat*, znacznik relacji, podstawa⟩,
    T – tablica wymian wewnątrztematowych
    (rzutowanie: sekwencja liter -> sekwencja liter)
- **Dla każdego** $e = \langle d, r, b \rangle \in L$:
  - t_p = sekwencja co najwyżej k wymian z T taka że P jest wspólnym początkiem d i b.
  - t_s = jak wyżej tylko w odniesieniu do wspólnego końca S
  - **Jeżeli** *długość*$(P) \geq$ *długość*(S)
  - **wtedy** dodaj $\langle d, r{+}t_p, t_p(b) \rangle$ do przykładów treningowych `normalnego' odgadywacza,
  - **w przeciwnym przypadku** dodaj $\langle$ *rev*(d), $r{+}t_s$, *rev*$(t_s(b))\rangle$ do przykładów odwróconego odgadywacza

Derywator: wymiany wewnątrztematowe

- Tablica wymian *T* - przykłady
  - <ch sz chi>: (ch, sz), (ch, chi), (sz, ch), (sz, chi), (chi, ch), (chi, sz)
  - <o ó>: (o, ó), (ó, o)
- Algorytm
  - w = b
  - Dla kolejnych reguł *r* z tablicy *T*
    - **jeżeli** wspólny_prefiks(d, *r*(w)) > wspólny_prefiks(d,b) **wtedy** *dodaj*(*r*,*t*)
    - w = r(w), b=w
  - proces powtarzamy do *k* wymian dołączonych do *t*

Derywator: wymiany wewnątrztematowe

- Przykład wyznaczonych wymian:
  - *świecznik – świeca*:
    - t = <[’cz’/’c’]>
    - b = *świecza*, prefiks: *świec*
  - *dolatywać – dolecieć*
    - *t* = <[‘la’/‘le’], [‘t’/‘c’]>
    - b =*dolatieć* and P = *dolat*
  - *pszczelarz – pszczoła*
    - *t* = <[‘cze’/‘czo’], [‘le’/‘ła’], [`la’/`le’]>
    - b = *pszczela* P = *pszczela*

Derywator: rozpoznawanie relacji

- Wejście: lemat *l*
- *l* zostaje dostarczony do obydwu modułów-odgadywaczy:
  - {⟨*b*, *t*, *r*⟩: *b* – zrekonstruowana podstawa, *t* – sekwencja podstawień zapamiętana wewnątrz odgadywacza, *r* – relacja}
- **Dla każdej** trójki:
  - *b* jest transformowane za pomocą odwróconej listy podstawień *t*
  - **jeżeli** *b* było wygenerowane przez odgadywacz oparty na prefiksie **wtedy** *b* musi zostać odwrócone
- Filtrowanie morfologiczne
  - reguły zdefiniowane dla każdego podtypu relacji
  - np. żeńskość: rzeczowniki w przypadku mianownika oraz derywat w rodzaju żeńskim

Derywator: rozpoznawanie relacji

- Filtrowanie za pomocą Morfeusza i korpusu
- Morfologiczne reguły filtrowania:
  - [rola:narzędzie]
    subst:*:nom:f:* subst:*:nom:m1:*
    subst:*:nom:m3:* subst:*:nom:n:* ->
    subst:*:nom:* ger:*:nom:* subst:*:nom:f:*
    subst:*:nom:m1:*
  - [żeńskość] subst:sg:nom:f:* ->
    subst:sg:nom:m1:* subst:sg:nom:m2:*
    subst:sg:nom:m3:* subst:sg:nom:n:*

Ocena

- Metody:
  - 10-krotna walidacja krzyżowa z uwzględnieniem podziału na klasy
    - Słowosieć (18.08.2011, 15718 przykładów)
  - ręczna
    - wejście: Morfeusz SGJP (341230 form wyrazowych)
    - do 50 par na klasę
- Uzysk z uczenia
  - porównanie – wersje Słowosieci:
    - 18.08.2011: 15718 przykładów
    - 06.09.2011: 17971 przykładów, 14% więcej

Ocena: gruboziarnista

Typ	Walidacja krzyżowa [%]	
	Dokładność	Kompletność
nacechowanie	97.1	55.8
rola (semantyczna)	75.7	35.6
zawieranie roli	100.0	36.4
synonimia międzyparadygmatyczna	90.5	80.4

Ocena: podtypy

Typ	Instancje	Walidacja krzyżowa [%]		Ocena ręczna [%]	
		Dokład.	Kompl.	Dokł. – podtyp	Dokł. – typ
aspektowość	5835	99.0	75.3	–	–
derywacyjność	1752	77.4	30.5	60.0	60.0
nosiciel stanu|cechy	176	40.0	14.2	34.0	34.0
żeńskość	1458	96.5	57.4	74.0	74.0
mieszkaniec	153	71.7	14.3	8.11	8.11
stan | cecha	188	40.8	9.6	46.0	46.0

Ocena: podtypy

Typ	Instan.	Walidacja krzyżowa [%]		Ręczna [%]	
		Dokład.	Kompl.	Dokład. - podtyp	Dokład. - typ
nacech. : deminutywność	1286	96.9	60.6	72.0	74.0
nacech.: augmentatywność	213	67.9	23.9	32.0	36.0
nacech.: Istota młoda	46	26.7	28.5	10.5	78.9
synonimia między.: N-ADJ	1219	98.5	82.2	81.6	81.6
synonimia między. : N-V	1173	82.1	79.1	—	—
synonimia między.: ADJ-V	99	95.9	66.9	—	—

Ocena: rola (semantyczna)

Podtyp	Instancje	Walidacja aut. [%]		Ocena ręczna [%]	
		Dokład.	Kompl.	Dokł. – podtyp	Dokł. - typ
agens przy niewyrażonym predykacie		89.4	30.5	40.0	42.0
agens	625	58.9	22.2	45.2	61.9
czas	45	11.7	59.5	20.0	20.0
miejsce	129	24.2	26.4	43.2	61.4
miejse przy niewyr.	143	59.2	25.8	40.0	40.0
instrument	327	29.7	36.6	42.0	58.0
pacjens	92	26.7	31.7	44.0	64.0
Nieokreślony	82	0.0	41.2	10.0	10.0
wytwór przy niewyr.	39	35.0	34.2	23.3	23.3
wytwór	614	32.7	54.7	12.0	12.0

Ocena: zawieranie roli

Podtyp	Instancje	Walidacja autom. [%]		Ocena ręczna [%]	
		Dokładność	Kompletn.	Dokład. – podtyp	Dokładn. – typ
zawieranie agensa	108	12.8	20.2	22.0	60.0
zawieranie czasu	14	10.0	45.0	0.0	66.7
zawieranie miejsca	33	20.0	42.5	0.0	84.2
Zawieranie narzędzia	287	38.8	43.2	20.0	84.0
zawieranie pacjensa	79	10.0	35.0	8.0	76.0
zawieranie wytworu	323	38.8	34.4	40.0	66.0
nieokreślony	59	15.0	42.0	45.8	60.4

Uzysk z uczenia

- Rodzaje nowych par
  - nowe instancje
    - derywat i podstawa opisane w wordnecie
    - brakuje instancji relacji
  - nowe derywaty
    - derywat nieopisany w wordnecie od opisanej już podstawy
    - najistotniejsze dla rozszerzania
  - nowe podstawy
    - w większości bardzo rzadkie lematy
  - obydwa nowe

Ocena: uzysk z uczenia

Typ pary	Podstawowy zbiór	Rozszerzony zbiór	Uzysk
Instancja	10582	11105	523
Derywat	37886	41333	3447
Derywat - monosemiczna podstawa	24681	26727	2046
Podstawa	5905	6467	562
Nowa para	103990	107695	3705

Rozszerzanie wordnetu

- Rozszerzanie oparte na *Derywatorze*
  - wygenerowane nowe instancje relacji
  - nowe derywaty i podstawy derywacyjne nieopisane do tej pory w wordnecie
  - ale *Derywator* pracuje na lematach!
  - konieczne jest rzutowanie do jednostek leksykalnych!
- Np. *osad* – *osadzić* (*Derywator*) → rola:pacjens
  - kilka jednostek leksykalnych: *osad* 1 ‘wastwa’, *osadzić* 3 ‘wytworzyć’, *osadzić* 2 ‘przymocować’, *osadzić* 1 ‘zatrzymać’ i *osadzić* 4 ‘umiejscowić’
  - tylko jedna poprawna: *osad* 1 –rola:pacjens → *osadzić* 3

Rozszerzanie wordnetu

- Automatyczne przypisanie nowych lematów do struktury wordnetu
- Filtrowanie semantyczne na podstawie informacji wydobytej z korpusu
- Podejście półautomatyczne
  - redukcja poziomu błędu do wielkości akceptowanej przez lingwistów

Filtrowanie semantyczne: oparte na wordnecie

- Założenia:
  - położenie jednostki leksykalnej w strukturze hiperonimicznej wordnetu charakteryzuje ją semantycznie
  - relacje derywacyjne są zróżnicowane pod względem klas semantycznych elementów par
- Cechy:
  - znaczniki morfo-syntaktyczne derywatu i podstawy
  - klasy semantyczne – synsety – dla derywatu i podstawy
    - niejednoznaczność: Derywator produkuje pary lematów
    - lematowi może odpowiadać kilka jednostek leksykalnych
    - listy klas dla derywatu i podstawy

Filtrowanie semantyczne: oparte na wordnecie

- Automatyczne wyznaczenie klas semantycznych
  - statystyki przynależności do synsetów dla par treningowych
  - wyznaczane poddrzewa hiperonimiczne skupiające większe ilości przykładów

Wyniki filtrowania w oparciu o wordnet

- Ograniczenie: jedynie nowe instancje
- Uczenie:
  - klasyfikator: SVM z biblioteki LibLINEAR (Weka)
  - strategie:
    - osobny klasyfikator (binarny) do klasy,
    - klasyfikator wieloklasowy,
    - kaskada klasyfikatorów binarnych
- Dane treningowo-testowe (Słowosieć 1.6)
  - *pozytywne*: lematów powiązanych relacją derywacyjną
  - *negatywne*: pary wygenerowane przez *Derywator* niepowiązane żadną relacją derywacyjną

Filtrowanie semantyczne: oparte na korpusie

- Założenia:
  - współwystąpienia elementów par derywacyjnych są niezwykle rzadkie
  - konteksty wystąpień derywatów i podstaw ujawniają ograniczenia semantyczne na ich użycie
- Cechy
  - dystrybucyjne leksykalno-syntaktyczne
  - znaczniki morfo-syntaktyczne derywatu i podstawy
  - długość końcówki
  - przynależność końcówki do grupy końcówek

Filtrowanie semantyczne: cechy korpusowe

- Cechy leksykalno-syntaktyczne dla rzeczowników – częstość współwystąpienia z:
  - określonym przymiotnikiem w relacji modyfikacji
  - określonym rzeczownikiem w złożeniu współrzędnym
  - określonym czasownikiem jako jego potencjalny podmiot
  - z określonym rzeczownikiem w dopełniaczu jako modyfikatorem

Filtrowanie semantyczne: cechy korpusowe

- Wydobycie częstości
  - korpus 1,8 miliarda tokenów:
    - Korpus IPI PAN + Korpus Rzeczpospolitej + Wikipedia (XI 2011) + teksty zebrane z Internetu
  - system *SuperMatrix*
  - cechy zapisane w języku ograniczeń morfo-syntaktycznych *WCCL*
  - macierz koincydencji:
    - wiersze: opisywane derywaty i podstawy
    - kolumny: lemat+relacja

Filtrowanie semantyczne: cechy korpusowe

- Transformacja cech
  - dużo rzadkich lematów -> wysoki poziom szumu
  - filtrowanie:
    - wierszy (np. >=50)
    - i kolumn (1% o najwyższej entropii, częstość>=100)
  - przekształcenie redukujące wymiar macierzy (algorytm LDA) – koncentracja informacji

Filtrowanie semantyczne oparte na korpusie: wyniki

- Klasyfikator: SVM z biblioteki LibLINEAR (Weka)
- Strategia:
  - kaskada klasyfikatorów binarnych uporządkowanych wg ich dokładności

Filtrowanie semantyczne oparte na korpusie: wyniki

Relacja	TP	TN	FP	FN	Dokł.	Kompl.
Mieszkaniec	84	2957	3	16	96,55	84,00
Nacechowanie	1097	1799	92	72	92,26	93,84
Rola	791	2068	112	89	87,60	89,89
Żeńskość	672	2353	23	12	96,69	98,25
Deminutywność	967	286	99	42	90,71	95,84
Eks.|augmentatywność	81	1227	35	51	69,83	61,36
Istota młoda	7	1361	5	21	58,33	25,00
Agens przy niew.	744	222	99	40	88,26	94,90
Miejsce przy niew.	59	1025	8	13	88,06	81,94

Wnioski i dalsze plany

- Relatywnie proste podejście
  - gotowy do użycia odgadywacz morfologiczny i analizator
  - narzędzie uczące się produktywnych reguł derywacji o dobrej dokładności na poziomie lematów
- W dużym stopniu podejście niezależne od języka
  - wyjątek: lista możliwych wymian wewnątrztematowych

Ciąg dalszy … już wkrótce!

Dziękuję bardzo za uwagę i cierpliwość!

W imieniu własnym i Współautorów,

Maciej Piasecki