

Warsztat nr 4

Narzędzia do automatycznej analizy odniesień w tekstach

Testowanie reguł Wccl Match na korpusie znakowanym wyrażeniami temporalnymi i analiza wyników oceny

Czas

15 minut

Cel

- Zapoznanie się z funkcjonalnością interfejsu Wccl Match Tester.
- Analiza wyników rozpoznanych anotacji z anotacjami w korpusie wzorcowym
- Filtrowanie anotacji

Materiały

1. Plik z regułami do rozpoznawania wyrażen temporalnych `wccl-match-timex.txt`

Do samodzielnego wykonania

1. Otwarcie strony http://inforex.clarin-pl.eu/index.php?page=wccl_match_tester
2. Pobranie pliku <http://tools.clarin-pl.eu/warsztaty/wccl-match-timex.txt> i skopiowanie jego zawartości
3. Wklejenie skopiowanej zawartości do pola **Rules**.
4. Wybór korpusu *KPWr 1.2.2 TimeML Train - A (1-100)* w polu **Corpora**
5. Kliknięcie przycisku **Evaluate**
6. Odczekanie, aż system zakończy wykonywanie reguł. Przetwarzanie może zająć kilkanaście sekund. Wyniki są aktualizowane co 50 dokumentów. Status przetwarzania widoczny jest po prawej stronie od przycisku **Evaluate**. Zakończenie przetwarzania zostanie zasygnalizowane poprzez ukrycie ikonki przetwarzania i aktywowanie przycisku **Evaluate**.
7. Odczytanie liczby niepoprawnie rozpoznanych oraz nierozpoznanych anotacji typu *t3_time*:
 - a. liczba niepoprawnie rozpoznanych anotacji *t3_time* wynosi 3 (kolumna *False Posiitives*),
 - b. liczba nierozpoznanych anotacji *t3_time* wynosi 1 (kolumna *False Negative*),
8. Wyświetlenie zdań zawierających nierozpoznane anotacje *t3_time* poprzez kliknięcie w komórkę z wartością 1 będącej przecięciem wiersza *t3_time* i kolumny *False Negative*.
9. Wyświetlone zostanie jedno zdanie z trzema zbiorami anotacji:
 - a. Zdanie z anotacjami pomocniczymi (*Auxiliary annotations*)
 - b. Zdanie z rozpoznanymi anotacjami (*Recognized annotations*)
 - c. Zdanie z nierozpoznanymi anotacjami (*Missing annotations*)

W tym przypadku nierozpoznaną anotacją jest „12.00 w Nowy Rok”. W jej miejsce zostały o znaczone dwie krótsze anotacje, tj. „12.00” jako *t3_time* oraz „Nowy Rok” jako *t3_date*.

Zapoznanie się ze składnią reguł Wccl Match i interpretowaniem błędów w regułach

Czas

15 minut

Cel

- Zapoznanie się ze składnią reguł Wccl Match
- Interpretowanie komunikatów błędu

Materiały

1. Plik z opisem anotacji i regułami `wccl-match-errors.txt`

Do samodzielnego wykonania

1. Otwarcie strony Wccl Match dla korpusu zaimportowanego z DSpace
2. Pobranie pliku <http://tools.clarin-pl.eu/warsztaty/wccl-match-errors.txt>
3. Skopiowanie pierwszych trzech linii z pliku i wklejenie w pole **Annotations**
4. Skopiowanie pozostałej treści pliku i wklejenie w pole **Rules**.
5. Wykonanie reguł poprzez kliknięcie w przycisk **Run**.
6. Pojawi się komunikat: **Błąd w regule WCCL: Not a tagset symbol: clas (tagset nkjp (2))**, który oznacza, że „clas” nie jest poprawnym atrybutem tokenu w tagsecie nkjp. W linii nr 5 należy zamienić „clas” na „class” i ponownie wykonać reguły.
7. Pojawi się komunikat **line 6:7: expecting ')', found 'is'**, który oznacza, że w linii nr 6 parser oczekiwał nawiasu zamykającego. W sekcji match znajdują się dwa operatory (equal i is), które powinny być oddzielone przecinkiem. Na końcu linii nr 5 należy dodać przecinek i ponownie wykonać reguły.
8. Pojawi się komunikat **line 13:3: expecting ')', found 'apply'**. Operatory apply reprezentujące kolene reguły powinny być oddzielone znakiem ;. W tym przypadku zabrakło tego znaku. Na końcu linii nr 13 należy dodać znak średnika i ponownie wykonać reguły.
9. W panelu Matches powinny zostać wyświetlone wyniki dopasowania.

Zapoznanie się z funkcjonalnością interfejsu Wccl Match (Corpus)

Czas

15 minut

Cel

- Zapoznanie się z funkcjonalnością interfejsu Wccl Match (Corpus)
- Uruchamianie reguł na korpusie zaimportowanym z DSpace
- Definiowanie anotacji do wyświetlenia

Materiały

1. Plik z regułami do rozpoznawania wyrażen temporalnych `wccl-match-anotacje.txt`

Do samodzielnego wykonania

1. Otwarcie strony Wccl Match dla korpusu zaimportowanego z DSpace
2. Pobranie pliku <http://tools.clarin-pl.eu/warsztaty/wccl-match-anotacje.txt>
3. Skopiowanie pierwszych dwóch linii z pliku i wklejenie w pole **Annotations**
4. Skopiowanie pozostałej treści pliku i wklejenie w pole **Rules**
5. Wykonanie reguł poprzez kliknięcie w przycisk **Run**
6. W panelu **Matches** zostaną wyświetlone zdania, które zawierają frazę „w nam_loc”. Fraza zostanie podkreślona na niebiesko.
7. Aby podświetlić dodatkowe anotacje występujące w zdaniu, np. nazwy organizacji i budowli należy w polu Annotations dodać:
nam_fac red no
nam_org green no
8. Aby wyświetlić zdania, które zawierają zarówno poszukiwaną frazę (`nam_w_loc`) i `nam_fac` należy w linii **nam_fac red no** zamienić **no** na **yes**