НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ

Кафедра прикладной математики и информатики

Компьютерная лингвистика. Практика.

#6

Нижний Новгород, 2019

Лабораторная работа #6.

POS-tagging, NER, dependency parsing

Задача.

1. Используя библиотеку spaCy, выполнить POS-tagging, NER и dependency parsing для своего корпуса. Оценить качество
2. Построить HMM для корпуса Brown с помощью алгоритма Витерби

```
def viterbi(probs_emission, probs_transition, words, states):
    V = np.ndarray((len(words), states.shape[0]), np.float128)

    # V[0, x_i] = P(y_0|x_i) * p(x_i)
    for i_state, state in enumerate(states):
        prob_e = probs_emission.get( (words[0], state), LOG_PROB_OF_ZERO )
        w = 0.0
        if state == START_SYMBOL or prob_e == LOG_PROB_OF_ZERO:
            w = 1.0
        V[0, i_state] = prob_e * w

    # V[y_i, x_i] = P(y_i|x_i) * max(t[x, x_i] * V[y_i - 1, x])_{x} = max(P(y_i|x_i) * t[x, x_i] * V[y_i - 1, x])_{x}
    for i_word in range(1, len(words)):
        for i_state in range(states.shape[0]):
            prob_e = probs_emission.get(
                (words[i_word], states[i_state]), LOG_PROB_OF_ZERO
            )
            tmp = probs_transition[:, i_state].flatten() * V[i_word - 1]
            V[i_word, i_state] = prob_e * np.max(tmp)

    return V
```

Лекция #5.

- Тематическое моделирование. Перплексия.
- LSA, LDA (sklearn, gensim)

Перплексия

- Модель языка - *распределение слов в документах*

Перплексия коллекции D для языковой модели $p(w|d)$ (чем меньше, тем лучше):

$$\mathcal{P}(D) = \exp\left(-\frac{1}{n}\sum_{d\in D}\sum_{w\in d} n_{dw}\ln p(w|d)\right)$$

$$n = \sum_{d\in D}\sum_{w\in d} n_{dw}$$

- **Перплексия** ~ правдоподобие, усредненное по всем словам и документам
- **Перплексия** - мера различности слов в тексте
- **Перплексия** - степень *ветвления* текста (сколько слов ожидается после каждого другого слова)

Лабораторная работа #7.

Вычисление перплексии

Задача.

1. Вычисление перплексии для собственного корпуса, используя частоты юниграмм НКРЯ

НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
УНИВЕРСИТЕТ