

Инструкция по разметке корпуса для задачи выделения сущностей.

Сущность – непрерывная подстрока текста, подпадающая под один из типов сущностей.

Отношение – сочетание двух сущностей, связанных по смыслу одним из предопределенных типов отношений.

Цель работы: ручная разметка корпуса для обучения на нем алгоритмов машинного обучения.

Задачи:

- 1) Выделение сущностей в тексте в формате BIOES-style
- 2) Выделение отношений в тексте в формате coNLL

В финале предполагается получить корпус, на котором можно обучать модели выделения именованных сущностей, и, что не менее важно, выделения отношений. Поэтому размечать какую-то сущность следует в двух случаях:

1. там, где эти конкретные отношения потенциально выделяются.
2. когда сущность, которая потенциально может участвовать в таком отношении, встречается без этого отношения

Дополнительно, если вы сомневаетесь, какие слова стоит вносить в сущность, то подумайте, важно ли конкретное слово в текущей сущности, например:

- "Проводится программа соцподдержки малоимущих семей".

В данном случае не стоит выделять "малоимущих семей" в сущность SOC, т.к. сущность "<ACT> программа соцподдержки малоимущих семей <ACT>" потенциально выделяется в отношение TSK - BIN : ACT.

Возникающие неоднозначности предполагается снимать посредством коммуникации и [указателя](#). Искать встретившееся словосочетание по всем уже размеченным файлам можно, например, при помощи VSCode, комбинация ctrl+shift+f.

Внимание: добавляйте частицы, важные для смысла высказывания в сущности например:

"MET курс рубля к доллару MET BIN не увеличивается BIN"

Типы сущностей:

- 1) **MET** (*metric*) – численный индикатор\показатель, объект, на котором определена операция сравнения. Примеры:
 - a. Производительности труда
 - b. Плановых и фактических значений показателей
 - c. Сейсмичность территории

- d. Вероятность возникновения нарушения
- e. Экономического роста
- f. Степень износа здания

2) **ECO** (*economics*) – экономическая сущность (из тех, что не подходят под определение **MET**) или объект инфраструктуры. Примеры:

- a. Биологических ресурсов
- b. Инновационного потенциала
- c. Внутреннего рынка
- d. Запасов и резервов продукции камчатских производителей
- e. Ресурсной базы рыболовства
- f. Мультипликативных эффектов развития
- g. Региональной экономики
- h. Источником энергии
- i. Потенциала энергетического сектора
- j. Энергетический баланс
- k. Частные компании
- l. Свободы экономической деятельности
- m. Частного бизнеса
- n. Бюджетного финансирования
- o. ООО "Моя фирма"
- p. Современной производственно-технологической базы судов
- q. Рыбопромыслового флота
- r. Современного оборудования
- s. Автомобильных дорог
- t. Топливо-энергетического комплекса
- u. Гидроэлектростанций приливной энергетики
- v. Порто-хаба
- w. Транзитных транспортных узлов
- x. Библиотечные и музейные фонды
- y. Система информационного обслуживания населения

3) **BIN** (*binary*) – одноразовое действие\бинарная характеристика (есть или нет) Пример:

- a. Создание
- b. Строительство

- c. Оказание
- d. Обеспечение
- e. Развитие
- f. Приобретение
- g. Формирование
- h. Диверсификация
- i. Вовлечение
- j. Освоение
- k. Строительство
- l. Модернизация
- m. Возобновление
- n. Размещены
- o. Ветшание
- p. Стимулирование
- q. Осуществление
- r. Функционирует
- s. Отсутствие
- t. Сохраняется
- u. Проводится

4) CMP (*compare*) - сравнительная характеристика. Примеры:

- a. Повышение
- b. Насыщение
- c. Увеличение
- d. Повышению
- e. Снизит
- f. Получит развитие
- g. Активизация
- j. Превышает аналогичный показатель
- k. Превышает их пропускную способность
- l. Уточнение [показателей, объемов финансовых средств и т.п.]
- m. Положительная динамика
- n. Совершенствование
- o. Расширение

5) QUA (*qualitative*) – качественная характеристика. Примеры:

- a. Стабильность
- b. Ограниченность
- c. Высокая
- d. Неэффективная
- e. Ограниченный
- f. Большое
- g. Склонен к самовозгоранию
- h. Не подлежит длительному хранению
- i. Самодостаточная
- j. Слабая
- k. Безопасная

7) ACT (*activity*) — принимаемые меры, проводимые мероприятия (часто сочетается с BIN: *проводится* [BIN] ремонт внутренних помещений [ACT], *запущен* [BIN] образовательный проект [ACT]). Примеры:

- a. Профилактика наркомании
- b. Профилактика сиротства
- c. Оповещение граждан о предстоящих мероприятиях по трудоустройству
- d. Антинаркотические акции
- e. Психолого-педагогическая помощь
- f. Операции "МАК", «КАНАЛ»
- g. Сопровождение приёмных семей и т.п.
- h. Реставрационные работы
- i. Праздничные мероприятия
- j. Образовательный проект «Серебряный университет»
- k. Пилотный проект по организации культурных мероприятий

8) INST (*institutions*) — различные учреждения, заведения, структуры и организации

- a. Учреждения культурно-досугового типа
- b. Центр Досуга и Творчества
- c. Этно-культурный центр
- d. Сеть муниципальных учреждений культуры

- e. Организации поддержки семьи и детства
- f. Центр занятости молодежи
- g. Органы опеки и попечительства
- h. Сеть организаций для детей-сирот

9) **SOC** (*social*) – социальный объект. Примеры:

- a. Населения страны
- b. специалистами и кадрами рабочих профессий
- c. Кадровая система
- d. Система льгот и преференций
- e. Профессиональных кадров
- f. Доступности жилья
- g. Качества жилищного обеспечения
- h. Обеспечению жильем отдельных категорий граждан
- i. Качества жизни
- j. Социальной защиты населения
- k. Досуг населения
- l. Народное творчество
- m. Научный и просветительский потенциал
- n. Историческое наследие

Типы отношений:

- 1) Текущее положение дел
 - a. Отрицательное **NNG** (*Now NeGative*)
 - b. Нейтральное **NNT** (*Now NeutTral*)
 - c. Позитивное **NPS** (*Now PoSitive*)
- 2) Реализованные изменения / итоги (*о прошлом*)
 - a. Отрицательное **PNG** (*Past NeGative*)
 - b. Нейтральное **PNT** (*Past NeutTral*)
 - c. Позитивное **PPS** (*Past PoSitive*)
- 3) Прогнозы
 - a. Отрицательное **FNG** (*Future NeGative*)
 - b. Нейтральное **FNT** (*Future NeutTral*)
 - c. Позитивное **FPS** (*Future PoSitive*)
- 4) Цели (что-то абстрактное)
 - a. **GOL** (*goal*)

5) Задачи (для достижения цели) (конкретные действия)
a. **TSK** (*task*)

Примеры отношений:

1) **NNG**

- a. Стоимость жилья – высокая (**MET - QUA**)
- b. Система дорог – слабая (**INF – QUA**)

2) **NNU**

- a. Качество жизни – удовлетворительное (**SOC - QUA**)
- b. Экономический показатель – на прежнем уровне (**ECO – CMP**)

3) **NPS**

- a. Цен на бензин – снижение (**MET – CMP**)
- b. Развитие автомобильных дорог – достаточное (**MET – QUA**)

4) **PNG**

- a. Пенсионный возраст – повышен (**MET – CMP**)

5) **PPS**

- a. Работы по восстановлению - завершены (**ACT – BIN**)

6) **FNG**

- a. Курса рубля – продолжение падения (**ECO – CMP**)

7) **FPS**

- a. Уровня зарплат - ожидается рост (**MET – CMP**)

8) **GOL**

- a. Повышение - уровня качества транспортного обслуживания населения; (**MET – CMP**)
- b. Повышение - доступности транспортных услуг; (**MET – CMP**)

9) **TSK**

- a. Ввод - новых линий (**BIN - INF**)
- b. Открытие - новых станций метрополитена (**BIN - INF**)
- c. Обновление - подвижного состава метрополитена (**BIN – INF**)
- d. Создание - преимуществ проезда транспортом общего пользования (**BIN - SOC**)

Для разметки используется формат coNLL. Каждое слово находится в своей строке, сущности размечаются в соответствующем столбце, где проставлено **OUT** - пустая сущность. Отношения

размечаются после разметки сущностей для предложения: указываются номера начал сущностей и тип отношения между ними. Пример:

количество рабочих мест в этом году увеличилось

1	количество	MET
2	рабочих	MET
3	мест	MET
4	в	OUT
5	этом	OUT
6	году	OUT
7	увеличилось	CMR

&& PPS:1:7

Для разметки можно также использовать текстовый редактор, например Notepad++. При работе в Notepad++ рекомендуется настроить макросы для каждой сущности: ставим курсор после тега **OUT**; кликаем **Macro -> Start Recording**; стираем **OUT**; пишем свой тег, например, **INF**; переносим курсор после следующего тега **OUT** на строку ниже; нажимаем **Macro -> Stop recording**; Сохраняем макрос с помощью **Macro -> Save Current Recorded Macro**; в диалоговом окне выбираем комбинацию клавиш (мне было удобно Alt + первая буква тега);

Приложение. Пример разметки

плотность населения города Москвы - MET
транспорте общего пользования - INF
корректировка сумм бюджетных значений - ESO
около 20 процентов жителей - SOC
пассажирских перевозок АСТ
Функционирует - бинарная (есть\нет) - BIN
превышает аналогичный показатель - CMR
слабая - качественная характеристика - QUA

Приложение к постановлению Правительства Москвы г. № Приложение к постановлению Правительства Москвы от 2 сентября 2011 г. № 408-ПП по состоянию на 1 января 2017 г. П А С П О Р Т Государственной программы города Москвы «Развитие транспортной системы»

Характеристика текущего состояния транспортной системы города Москвы Город Москва является крупнейшим транспортным узлом России.

Население города Москвы составляет 12,33 млн. человек.

Площадь города Москвы составляет 2561 кв. км.

На транспорте общего пользования в городе Москве совершается более 14 млн. поездок в сутки и более 5 млрд. поездок за год.

В настоящее время в городе функционирует транспортная система, в которую входят следующие виды транспорта: Московский метрополитен (далее – метрополитен), Московская монорельсовая транспортная система, железнодорожный транспорт, наземный городской пассажирский транспорт, которые перевозят в год около 3,75 млрд. экономически активных пассажиров. Состояние и проблемы транспортной

Commented [1]: NNT

отрасли города Москвы характеризуются следующими параметрами (по состоянию на конец 2011 года): плотность населения города Москвы (107,3 человек/га) превышает аналогичный показатель крупнейших мегаполисов мира; плотность улично-дорожной сети (3,38 км/кв. км) в 2–4 раза ниже аналогичных показателей крупнейших мегаполисов мира. Этот показатель на территории Троицкого и Новомосковского административных округов города Москвы составляет 0,88 км/кв. км; фактическая загрузка всех видов транспорта общего пользования в утренние часы пик при движении в центр составляет 1,060 млн. человек и превышает провозную способность (870 тыс. человек) в среднем на 22 процента; загрузка автомобильных дорог превышает их пропускную способность на 42 процента; среднее время поездки на транспорте общего пользования пассажира из жилых районов до «мест приложения труда» составляет 67 минут, а около 20 процентов жителей города Москвы тратят более трех часов в день на дорогу от дома до работы и обратно; объем перевозок на метрополитене в 1,5–2 раза превышает соответствующие показатели крупнейших мегаполисов мира при меньшей протяженности линий метрополитена, средняя наполняемость вагонов электропоездов метрополитена составляет 5,2 человек/кв. м; недостаточное развитие наземных видов транспорта общего пользования, низкая плотность маршрутной сети, отсутствие преимуществ для движения транспорта общего пользования в транспортном потоке; слабая система хордовых связей между радиальными автомобильными дорогами; отсутствие дублирующих направлений магистральных радиальных автомобильных дорог, их недостаточная пропускная способность и несоответствие параметров автомобильных дорог на границе города Москвы и Московской области планировочным параметрам; недостаточный уровень технического состояния искусственных сооружений и проезжей части на значительном протяжении автомобильных дорог федерального и регионального значения; отсутствие транспортных развязок на разных уровнях на пересечениях улично-дорожной сети, в том числе с железными дорогами и реками; отсутствие эффективной системы управления дорожным движением; недостаточное взаимодействие перевозчиков, обслуживающих автобусные маршруты, отсутствие скоординированной системы управления перевозками,

Сохраняется устойчивая тенденция к увеличению затрат времени при поездках по городу. При отсутствии мер по улучшению транспортной ситуации к 2025 году ожидается практически удвоение перегрузки по всем видам транспорта, несмотря на прогнозируемый умеренный прирост населения по 0,8 процента в год. Время в пути на личном транспорте вырастет на 25 процентов – до 75 минут. Доля жителей, которые будут тратить более трех часов в день на дорогу от дома на работу и обратно, возрастет до 30 процентов. В ходе реализации мероприятий Государственной программы города Москвы «Развитие транспортной системы» (далее – Государственная программа) в 2012–2016 годах были существенно увеличены темпы метростроения, строительства, реконструкции и содержания дорог, проведено обновление подвижного состава, развернуты работы по развитию железнодорожного транспорта, сформированы пешеходное и велосипедное пространства. Более подробная информация о достигнутых в ходе реализации Государственной программы результатах изложена в составе соответствующих подпрограмм. В результате принятых мер достигнуты следующие результаты по основным показателям, интегрально характеризующие улучшение транспортной ситуации: среднее время поездки на транспорте общего пользования в утренние часы пик из жилых районов вблизи Московской кольцевой автомобильной дороги до центра города уменьшилось с 67 до 56,8 минут; годовой объем пассажирских перевозок на транспорте общего пользования (без учета льготных категорий пассажиров) вырос с 2,82 до 3,75 млрд. пассажиров. Прогноз развития транспортной системы и планируемые макроэкономические показатели по итогам реализации Государственной программы Для

Commented [2]: NNG

Commented [3]: NNG

Commented [4]: NNG

Commented [5]: NNG

Commented [6]: NNG

Commented [7]: NNG

Commented [8]: NNG

Commented [9]: NNG

Commented [10]: NNG

Commented [11]: NNG

Commented [12]: NNG

Commented [13]: NNG

Commented [14]: NNG

Commented [15]: NNG

Commented [16]: NNG

Commented [17]: NNG

Commented [18]: FNG

Commented [19]: FNG

Commented [20]: FNT

Commented [21]: FNG

Commented [22]: FNG

Commented [23]: PPS

Commented [24]: PPS

Commented [25]: PPS

Commented [26]: PPS

Commented [27]: PPS

Commented [28]: PPS

Commented [29]: GOL

улучшения транспортной ситуации предлагается осуществить ряд мер, основные из которых направлены на снижение среднего времени поездки в часы пик (до 55 минут при движении по городу Москве и до 70 минут из Московской области в город Москву) к 2020 году за счет увеличения провозной способности транспорта общего пользования на 40 процентов и сдерживания использования личного автомобильного транспорта (снижение доли поездок на личном транспорте в часы пик с 12 до 8 процентов на въезде в пределы Третьего транспортного кольца, с 26 до 18 процентов при въезде на Московскую кольцевую автомобильную дорогу). В рамках Государственной программы планируется приступить к реализации этих мер, рассчитанных на период до 2020 года. Результатом выполнения Государственной программы должны стать: снижение среднего времени поездки на транспорте общего пользования в утренние часы

Commented [30]: GOL

Commented [31]: GOL

Commented [32]: GOL

Commented [33]: GOL

Commented [34]: GOL

Commented [35]: GOL

Описание целей и задач Государственной программы Целью Государственной программы является обеспечение комфортных условий жизнедеятельности населения города Москвы путем развития устойчиво функционирующей, безопасной, привлекательной и удобной для всех групп населения транспортной системы, как части Московского транспортного узла. Основными задачами Государственной программы, направленными на достижение поставленной цели, определены следующие: повышение уровня качества транспортного обслуживания населения; повышение доступности транспортных услуг для всех групп населения. Для достижения цели Государственной программы и реализации направленных на ее достижение задач разработаны следующие основные мероприятия: ввод новых линий и открытие новых станций метрополитена, обновление подвижного состава метрополитена; создание преимуществ проезда транспортом общего пользования. Адресный перечень объектов капитального строительства, реконструкции или технического перевооружения, финансирование которых планируется осуществить полностью или частично за счет средств бюджета города Москвы при реализации Государственной программы, формируется в адресной инвестиционной программе города Москвы в установленном порядке. Основными концептуальными направлениями развития Московского транспортного узла определены: обеспечение преимущества транспорта общего пользования при осуществлении пассажирских перевозок; обеспечение объемов пассажирских и грузовых перевозок в соответствии с прогнозными потребностями; оптимизация и координация грузовых перевозок на основе комплексного взаимоувязанного развития транспортно-логистической системы, функционирующей на территории Московского региона

Commented [36]: GOL

Commented [37]: GOL

Sergei Troshin
FPS

Commented [38]: GOL

Commented [39]: GOL

Commented [40]: TSK

Commented [41]: TSK

Commented [42]: TSK

Commented [43]: GOL

Commented [44]: GOL

Sergei Troshin
GOL

Commented [45]: TSK

Commented [46]: GOL

Commented [47]: GOL

Commented [48]: TSK

Commented [49]: FNT

В течение предстоящего 2017 года возможна корректировка сумм бюджетных назначений при наличии дополнительных поступлений в местный бюджет. Благодаря накопленному опыту, кадровому составу, действующей нормативно-правовой базе органы местного самоуправления реализуют возложенные на них полномочия по решению вопросов местного значения. Пояснительная записка обоснование параметров прогноза социально-экономического развития В разрезе социально-экономического анализа муниципального образования основными характеристиками являются его численность и возрастные категории населения; наполняемость бюджета; расходная часть бюджета; доля расходов местного бюджета, направленная на содержание органов местного самоуправления, в общей сумме расходов местного бюджета; сумма средств местного бюджета, направленная на проведение благоустройства территории, в расчете на одного жителя муниципального образования; сумма средств местного бюджета, направленная на проведение программных мероприятий за исключение благоустройства территории; количество детей, переданных на семейные формы устройства. Проводя анализ численности населения МО МО Ульяновка, можно говорить о тенденции к его плавному устойчивому росту, начиная с 2014 года. К 2019 году прогнозируемая численность населения МО МО Ульяновка составит 77,0 тыс. чел.

Commented [50]: TSK

Commented [51]: TSK

Commented [52]: TSK

Наполняемость бюджета зависит от перечня доходов, нормативов отчислений доходов, собираемости налогов. В 2017-2019 гг из перечня доходов исключен налог на имущество физических лиц.

Commented [53]: PNT