

In [1]:

```
from text_analyzer.grammar import Grammar
from text_analyzer.dictionary import Dictionary
from text_analyzer.text_analyzer import TextAnalyzer
```

In [2]:

```
Dict = Dictionary()
Dict.set_syn_api(use_syn_api=True, syns_max_lvl=2)
```

In [3]:

```
Dict.add(u'ИМЯ', lambda E: True if
        'Name' in E.wtag or (len(E.word)==1
                             and E.ends=='.') else False)
Dict.add(u'ОТЧЕСТВО', lambda E: True if 'Patr' in E.wtag
        or (len(E.word)==1 and E.ends=='.') else False)
Dict.add(u'ФАМИЛИЯ', lambda E: True if 'Surn' in E.wtag else False)
Dict.add(u'КАЧЕСТВО', lambda E: True if 'ADJF' in E.wtag else False)
Dict.add(u'МЕСТОИМЕНЕНИЕ', lambda E: True if 'NPRO' in E.wtag else False)
Dict.add(u'ФРАЗА_В_КАВЫЧКАХ', lambda E: True
        if E.begs in ['"', '"'] and E.ends in ['"', '"'] else False)
Dict.add(u'НАЧИНАЕТСЯ_С_ЗАГЛАВНОЙ', lambda E: True
        if len(E.word)>1 and (E.begs.lower() != E.begs or
                             E.word.lower() != E.word) else False)
Dict.add(u'ЗАГЛАВНЫМИ_БУКВАМИ', lambda E: True if
        E.word.lower().upper() == E.word else False)
Dict.add(u'ФИЗ_ФОРМА_ФИРМЫ', words=[u'фирма', u'компания',
        u'организация'])
Dict.add(u'ЮР_ФОРМА_ФИРМЫ', words=[u'ooo', u'оao', u'зao',
        u'акционерное общество'])
Dict.add(u'ДОЛЖНОСТЬ', words=[u'директор', u'вице-президент',
        u'начальник', u'рабочий', u'председатель'])
Dict.add(u'ДЕЙСТВИЕ_ИЗМЕНЕНИЯ', words=[u'менять', u'поменять', u'покидать'])
Dict.add(u'ПРИНАДЛЕЖНОСТЬ', words=[u'есть', u'иметь', u'обладать'])
Dict.add(u'СОЦИАЛЬНАЯ_РОЛЬ', words=[u'муж', u'жена', u'сын', u'дочь',
        u'мать', u'отец'])
Dict.add(u'ПРЕДМЕТ', words=[u'машина', u'квартира'])
# предметы, конечно, должны задаваться иначе(например, как все существительные,
# отличные от уже принадлежащих к иным категориям, но для простоты
# оставим простейший вариант)
Dict.add(u'ЧЕЛОВЕК')
Dict.add(u'НАЗВАНИЕ')
Dict.add(u'ФИРМА')
```

In [4]:

```
Gram = Grammar()
```

In [5]:

```
Gram.add_repl([u'ИМЯ', u'НАЧИНАЕТСЯ_С_ЗАГЛАВНОЙ'],  
              u'ИМЯ')  
Gram.add_repl([u'ОТЧЕСТВО', u'НАЧИНАЕТСЯ_С_ЗАГЛАВНОЙ'],  
              u'ОТЧЕСТВО')  
Gram.add_repl([u'ФАМИЛИЯ', u'НАЧИНАЕТСЯ_С_ЗАГЛАВНОЙ'],  
              u'ФАМИЛИЯ')  
Gram.add_repl([u'ЮР_ФОРМА_ФИРМЫ', u'НАЧИНАЕТСЯ_С_ЗАГЛАВНОЙ'],  
              u'ЮР_ФОРМА_ФИРМЫ')  
Gram.add_repl([u'ЮР_ФОРМА_ФИРМЫ', u'ЗАГЛАВНЫМИ_БУКВАМИ'],  
              u'ЮР_ФОРМА_ФИРМЫ')  
Gram.add_repl([u'НАЗВАНИЕ', u'ФРАЗА_В_КАВЫЧКАХ'],  
              u'НАЗВАНИЕ')  
Gram.add_repl([u'НАЗВАНИЕ', u'НАЧИНАЕТСЯ_С_ЗАГЛАВНОЙ'],  
              u'НАЗВАНИЕ')  
Gram.add_repl([u'МЕСТОИМЕНИЕ', u'НАЧИНАЕТСЯ_С_ЗАГЛАВНОЙ'],  
              u'МЕСТОИМЕНИЕ')
```

In [6]:

```
Gram.add_seq([u'ИМЯ', u'ОТЧЕСТВО', u'ФАМИЛИЯ'],  
             u'ЧЕЛОВЕК')  
Gram.add_seq([u'ИМЯ', u'ФАМИЛИЯ'],  
             u'ЧЕЛОВЕК')  
Gram.add_seq([u'ИМЯ', u'ОТЧЕСТВО'],  
             u'ЧЕЛОВЕК')  
Gram.add_seq([u'ИМЯ'],  
             u'ЧЕЛОВЕК')  
Gram.add_seq([u'ФАМИЛИЯ', u'ИМЯ', u'ОТЧЕСТВО'],  
             u'ЧЕЛОВЕК')  
Gram.add_seq([u'ФАМИЛИЯ', u'ИМЯ'],  
             u'ЧЕЛОВЕК')  
Gram.add_seq([u'ФАМИЛИЯ'],  
             u'ЧЕЛОВЕК')  
Gram.add_seq([u'ФРАЗА_В_КАВЫЧКАХ'],  
             u'НАЗВАНИЕ')  
Gram.add_seq([u'НАЧИНАЕТСЯ_С_ЗАГЛАВНОЙ'],  
             u'НАЗВАНИЕ')  
Gram.add_seq([u'КАЧЕСТВО', u'ФИЗ_ФОРМА_ФИРМЫ',  
              u'ЮР_ФОРМА_ФИРМЫ', u'НАЗВАНИЕ'],  
              u'ФИРМА')  
Gram.add_seq([u'ФИЗ_ФОРМА_ФИРМЫ', u'ЮР_ФОРМА_ФИРМЫ', u'НАЗВАНИЕ'],  
              u'ФИРМА')  
Gram.add_seq([u'ЮР_ФОРМА_ФИРМЫ', u'НАЗВАНИЕ'],  
              u'ФИРМА')
```

In [7]:

```
Gram.add_link([u'ЧЕЛОВЕК', u'ДЕЙСТВИЕ_ИЗМЕНЕНИЯ', u'ФИРМА'],
              u'Человек взаимодействует с фирмой')
Gram.add_link([u'ЧЕЛОВЕК', u'ПРИНАДЛЕЖНОСТЬ', u'ФИРМА'],
              u'Человек руководит фирмой')
Gram.add_link([u'ЧЕЛОВЕК', u'ДОЛЖНОСТЬ', u'ФИРМА'],
              u'Человек работает в фирме')
Gram.add_link([u'ЧЕЛОВЕК', u'ФИРМА'],
              u'Человек связан с фирмой')
Gram.add_link([u'ЧЕЛОВЕК', u'МЕСТОИМЕНИЕ',
              u'ДЕЙСТВИЕ_ИЗМЕНЕНИЯ', u'ЧЕЛОВЕК'],
              u'Два человека связаны', max_sent=2)
Gram.add_link([u'ЧЕЛОВЕК', u'ЧЕЛОВЕК', u'МЕСТОИМЕНИЕ'],
              u'Два человека связаны', max_sent=2)
Gram.add_link([u'ЧЕЛОВЕК', u'ДЕЙСТВИЕ_ИЗМЕНЕНИЯ', u'ЧЕЛОВЕК'],
              u'Два человека связаны')
Gram.add_link([u'ЧЕЛОВЕК', u'ПРИНАДЛЕЖНОСТЬ', u'ЧЕЛОВЕК'],
              u'Два человека связаны социальной ролью')
Gram.add_link([u'МЕСТОИМЕНИЕ', u'ПРИНАДЛЕЖНОСТЬ', u'ЧЕЛОВЕК'],
              u'Два человека связаны социальной ролью')
Gram.add_link([u'ЧЕЛОВЕК', u'ПРИНАДЛЕЖНОСТЬ', u'МЕСТОИМЕНИЕ'],
              u'Два человека связаны социальной ролью')
Gram.add_link([u'ЧЕЛОВЕК', u'ПРИНАДЛЕЖНОСТЬ', u'ПРЕДМЕТ'],
              u'У человека есть предмет')
Gram.add_link([u'МЕСТОИМЕНИЕ', u'ПРИНАДЛЕЖНОСТЬ', u'ПРЕДМЕТ'],
              u'У человека есть предмет')
Gram.add_link([u'ЧЕЛОВЕК', u'ПРИНАДЛЕЖНОСТЬ', u'СОЦИАЛЬНАЯ_РОЛЬ'],
              u'Два человека связаны социальной ролью')
Gram.add_link([u'ЧЕЛОВЕК', u'МЕСТОИМЕНИЕ', u'СОЦИАЛЬНАЯ_РОЛЬ'],
              u'Два человека связаны социальной ролью')
Gram.add_link([u'МЕСТОИМЕНИЕ', u'СОЦИАЛЬНАЯ_РОЛЬ', u'ЧЕЛОВЕК'],
              u'Два человека связаны социальной ролью')
Gram.add_link([u'ЧЕЛОВЕК', u'СОЦИАЛЬНАЯ_РОЛЬ', u'ЧЕЛОВЕК'],
              u'Два человека связаны социальной ролью')
Gram.add_link([u'НАЗВАНИЕ', u'КАЧЕСТВО'],
              u'Свойство объекта')
```

Example 1

In [8]:

```
TA = TextAnalyzer(Dict, Gram)
txt = u'''Пётр Сергеевич Иванов покинул должность вице-президента
известной фирмы ООО "Анкор". Его заменил Иван Алексеевич Сидоров.'''
TA.analyze(txt)
```

```
-----> TextAnalyzer
-----| Result of text analysis:
{E1 ЧЕЛОВЕК sent_start }
  {E1.1 ИМЯ sent_start } |Пётр| (NF:пётр)
  {E1.2 ОТЧЕСТВО} |Сергеевич| (NF:сергей)
  {E1.3 ФАМИЛИЯ} |Иванов| (NF:иванов)
{E2 ДЕЙСТВИЕ_ИЗМЕНЕНИЯ} |покинул| (NF:покинуть)
{E3 } |должность| (NF:должность)
{E4 ДОЛЖНОСТЬ} |вице-президента| (NF:вице-президент)
{E5 ФИРМА}
  {E5.1 КАЧЕСТВО} |известной| (NF:известный)
  {E5.2 ФИЗ_ФОРМА_ФИРМЫ} |фирмы| (NF:фирма)
  {E5.3 ЮР_ФОРМА_ФИРМЫ} |ООО| (NF:ooo)
  {E5.4 НАЗВАНИЕ}
    {E5.4.1 НАЧИНАЕТСЯ_С_ЗАГЛАВНОЙ} |Анкор| (NF:анкор)
{E6 МЕСТОИМЕНИЕ sent_start } |Его| (NF:он)
{E7 ДЕЙСТВИЕ_ИЗМЕНЕНИЯ} |заменил| (NF:заменить)
{E8 ЧЕЛОВЕК}
  {E8.1 ИМЯ} |Иван| (NF:иван)
  {E8.2 ОТЧЕСТВО} |Алексеевич| (NF:алексей)
  {E8.3 ФАМИЛИЯ} |Сидоров| (NF:сидоров)
-----| Result of links analysis:
L(Человек взаимодействует с фирмой): E1 - E2 - E5
L(Человек работает в фирме): E1 - E4 - E5
L(Человек связан с фирмой): E1 - E5
L(Два человека связаны): E1 - E6 - E7 - E8
```

In [9]:

```
# Пример использования результатов анализа.
# Выясним все что известно про человека с именем "Иван":
TA.get_info(u'Иван')
```

```
--- Искомая сущность:
{E8 ЧЕЛОВЕК}
  {E8.1 ИМЯ} |Иван| (NF:иван)
  {E8.2 ОТЧЕСТВО} |Алексеевич| (NF:алексей)
  {E8.3 ФАМИЛИЯ} |Сидоров| (NF:сидоров)
--- Связанные сущности:
{E1 ЧЕЛОВЕК sent_start }
  {E1.1 ИМЯ sent_start } |Пётр| (NF:пётр)
  {E1.2 ОТЧЕСТВО} |Сергеевич| (NF:сергей)
  {E1.3 ФАМИЛИЯ} |Иванов| (NF:иванов)
{E7 ДЕЙСТВИЕ_ИЗМЕНЕНИЯ} |заменил| (NF:заменить)
{E6 МЕСТОИМЕНИЕ sent_start } |Его| (NF:он)
--- Связи:
L(Два человека связаны): E1 - E6 - E7 - E8
```

In [10]:

```
# Пример использования результатов анализа.
# Выясним все что известно про человека с именем "Петр":
TA.get_info(u'Петр')
```

```
--- Искомая сущность:
{E1 ЧЕЛОВЕК sent_start }
  {E1.1 ИМЯ sent_start } |Пётр| (NF:пётр)
  {E1.2 ОТЧЕСТВО} |Сергеевич| (NF:сергей)
  {E1.3 ФАМИЛИЯ} |Иванов| (NF:иванов)
--- Связанные сущности:
{E6 МЕСТОИМЕНИЕ sent_start } |Его| (NF:он)
{E2 ДЕЙСТВИЕ_ИЗМЕНЕНИЯ} |покинул| (NF:покинуть)
{E7 ДЕЙСТВИЕ_ИЗМЕНЕНИЯ} |заменял| (NF:заменить)
{E5 ФИРМА}
  {E5.1 КАЧЕСТВО} |известной| (NF:известный)
  {E5.2 ФИЗ_ФОРМА_ФИРМЫ} |фирмы| (NF:фирма)
  {E5.3 ЮР_ФОРМА_ФИРМЫ} |ООО| (NF:ooo)
  {E5.4 НАЗВАНИЕ}
    {E5.4.1 НАЧИНАЕТСЯ_С_ЗАГЛАВНОЙ} |Анкор| (NF:анкор)
{E4 ДОЛЖНОСТЬ} |вице-президента| (NF:вице-президент)
{E8 ЧЕЛОВЕК}
  {E8.1 ИМЯ} |Иван| (NF:иван)
  {E8.2 ОТЧЕСТВО} |Алексеевич| (NF:алексей)
  {E8.3 ФАМИЛИЯ} |Сидоров| (NF:сидоров)
--- Связи:
L(Человек взаимодействует с фирмой): E1 - E2 - E5
L(Человек работает в фирме): E1 - E4 - E5
L(Человек связан с фирмой): E1 - E5
L(Два человека связаны): E1 - E6 - E7 - E8
```

In [11]:

```
# Пример использования результатов анализа.
# Выясним все что известно про человека с именем "Анкор":
TA.get_info(u"Анкор")
```

```
--- Искомая сущность:
{E5 ФИРМА}
  {E5.1 КАЧЕСТВО} |известной| (NF:известный)
  {E5.2 ФИЗ_ФОРМА_ФИРМЫ} |фирмы| (NF:фирма)
  {E5.3 ЮР_ФОРМА_ФИРМЫ} |ООО| (NF:ooo)
  {E5.4 НАЗВАНИЕ}
    {E5.4.1 НАЧИНАЕТСЯ_С_ЗАГЛАВНОЙ} |Анкор| (NF:анкор)
--- Связанные сущности:
{E1 ЧЕЛОВЕК sent_start }
  {E1.1 ИМЯ sent_start } |Пётр| (NF:пётр)
  {E1.2 ОТЧЕСТВО} |Сергеевич| (NF:сергей)
  {E1.3 ФАМИЛИЯ} |Иванов| (NF:иванов)
{E4 ДОЛЖНОСТЬ} |вице-президента| (NF:вице-президент)
{E2 ДЕЙСТВИЕ_ИЗМЕНЕНИЯ} |покинул| (NF:покинуть)
--- Связи:
L(Человек взаимодействует с фирмой): E1 - E2 - E5
L(Человек работает в фирме): E1 - E4 - E5
L(Человек связан с фирмой): E1 - E5
```

Example 2

In [12]:

```
TA = TextAnalyzer(Dict, Gram)
txt = u'''У Ивана Петрова есть сын. Марья Петрова его жена. '''
TA.analyze(txt)
```

```
-----> TextAnalyzer
-----| Result of text analysis:
{E1 ЗАГЛАВНЫМИ_БУКВАМИ sent_start } |У| (NF:y)
{E2 ЧЕЛОВЕК}
  {E2.1 ИМЯ} |Ивана| (NF:иван)
  {E2.2 ФАМИЛИЯ} |Петрова| (NF:петров)
{E3 ПРИНАДЛЕЖНОСТЬ} |есть| (NF:есть)
{E4 СОЦИАЛЬНАЯ_РОЛЬ} |сын| (NF:сын)
{E5 ЧЕЛОВЕК sent_start }
  {E5.1 ИМЯ sent_start } |Марья| (NF:марья)
  {E5.2 ФАМИЛИЯ} |Петрова| (NF:петров)
{E6 МЕСТОИМЕНИЕ} |его| (NF:он)
{E7 СОЦИАЛЬНАЯ_РОЛЬ} |жена| (NF:жена)
-----| Result of links analysis:
L(Два человека связаны): E2 - E5 - E6
L(Два человека связаны социальной ролью): E2 - E3 - E4
L(Два человека связаны социальной ролью): E5 - E6 - E7
```

In [13]:

```
# Пример использования результатов анализа.
# Выясним все что известно про человека с именем "Иван":
TA.get_info(u'Иван')
```

```
--- Искомая сущность:
{E2 ЧЕЛОВЕК}
  {E2.1 ИМЯ} |Ивана| (NF:иван)
  {E2.2 ФАМИЛИЯ} |Петрова| (NF:петров)
--- Связанные сущности:
{E4 СОЦИАЛЬНАЯ_РОЛЬ} |сын| (NF:сын)
{E6 МЕСТОИМЕНИЕ} |его| (NF:он)
{E3 ПРИНАДЛЕЖНОСТЬ} |есть| (NF:есть)
{E5 ЧЕЛОВЕК sent_start }
  {E5.1 ИМЯ sent_start } |Марья| (NF:марья)
  {E5.2 ФАМИЛИЯ} |Петрова| (NF:петров)
--- Связи:
L(Два человека связаны): E2 - E5 - E6
L(Два человека связаны социальной ролью): E2 - E3 - E4
```

In [14]:

```
# Пример использования результатов анализа.
# Выясним все что известно про человека с именем "Марья":
TA.get_info(u'Марья')
```

```
--- Искомая сущность:
{E5 ЧЕЛОВЕК sent_start }
  {E5.1 ИМЯ sent_start } |Марья| (NF:марья)
  {E5.2 ФАМИЛИЯ} |Петрова| (NF:петров)
--- Связанные сущности:
{E7 СОЦИАЛЬНАЯ РОЛЬ} |жена| (NF:жена)
{E6 МЕСТОИМЕНИЕ} |его| (NF:он)
{E2 ЧЕЛОВЕК}
  {E2.1 ИМЯ} |Ивана| (NF:иван)
  {E2.2 ФАМИЛИЯ} |Петрова| (NF:петров)
--- Связи:
L(Два человека связаны): E2 - E5 - E6
L(Два человека связаны социальной ролью): E5 - E6 - E7
```

Example 3

In [15]:

```
TA = TextAnalyzer(Dict, Gram)
txt = u'''У Кузи есть машина. Машина очень быстрая. '''
TA.analyze(txt)
```

```
-----> TextAnalyzer
-----| Result of text analysis:
{E1 ЗАГЛАВНЫМИ_БУКВАМИ sent_start } |У| (NF:y)
{E2 ЧЕЛОВЕК}
  {E2.1 ИМЯ} |Кузи| (NF:кузя)
{E3 ПРИНАДЛЕЖНОСТЬ} |есть| (NF:есть)
{E4 ПРЕДМЕТ} |машина| (NF:машина)
{E5 НАЗВАНИЕ sent_start }
  {E5.1 ПРЕДМЕТ ; НАЧИНАЕТСЯ_С_ЗАГЛАВНОЙ sent_start } |Машина| (NF:м
ашина)
{E6 } |очень| (NF:очень)
{E7 КАЧЕСТВО} |быстрая| (NF:быстрый)
-----| Result of links analysis:
L(У человека есть предмет): E2 - E3 - E4
L(Свойство объекта): E5 - E7
```

In [16]:

```
# Пример использования результатов анализа.
# Выясним все что известно про Кузю:
TA.get_info(u'Кузя')
```

```
--- Искомая сущность:
{E2 ЧЕЛОВЕК}
  {E2.1 ИМЯ} |Кузи| (NF:кузя)
--- Связанные сущности:
{E3 ПРИНАДЛЕЖНОСТЬ} |есть| (NF:есть)
{E4 ПРЕДМЕТ} |машина| (NF:машина)
--- Связи:
L(У человека есть предмет): E2 - E3 - E4
```