

## ИНСТРУКЦИЯ К ШАБЛОНУ

TEI - это международный проект по стандартизации языковой разметки.

Язык TEI основан на системе кодирующих тегов.

Общая структура словаря в нашем TEI-шаблоне выглядит следующим образом:

```
< tei >
  < file_desc > < /file_desc >
  < front > < /front >
  < body >
    < div >
      < superEntry >
        < metalemma > < /metalemma >
        < entry > < /entry >
      < /superEntry >
    < /div >
  < /body >
< /tei >
```

Блок **< file\_desc >** включает в себя некоторую основную информацию про словарь.

В **< entry >** и **< superEntry >** находятся словарные входы.

**< entry >** содержит один структурированный словарный вход.

**< superEntry >** объединяет словарный вход (несколько словарных входов, например, омографы) и металемму.

**< metalemma >** нужна для поиска всех соответствующих друг другу словарных статей (например, "нет" у

Ожегова, "нет" у Ушакова, "нѣтъ" у Даля).

В **< div >** можно объединить все словарные входы на одну букву или части двуязычного словаря (например, один div - рус-англ, второй - англ-рус).

Блок **< file\_desc >** содержит имя того, кто скомпилировал словарь, название словаря, ссылку на словарь в интернете, число словарных входов (здесь должно быть число!) и краткое сообщение о словаре. Вся остальная информация про словарь находится в **< front >**

**< file\_desc >**

**< respStmt >**

**< name >** имя того, кто скомпилировал файл со словарём **< /name >**

**< /respStmt >**

**< extent >** количество словарных входов: ЧИСЛО **< /extent >**

**< sourceDesc >**

**< ref target="ссылка на словарь" >** текст ссылки **< /ref >**

**< p >** сообщение о словаре (e.g. 'This database was converted from the XXX dictionary by XXX) **< /p >**

**< /sourceDesc >**

**< /file\_desc >**

Блок **< front >** состоит из блока **< head >** . В нём указывается автор словаря, редактор, издательство, языки и пр.

**< front >**

**< head >**

**< title dict\_id="не заполняется" >** название словаря **< /title >**

**< volume n="номер тома" of="количество томов" >** **< /volume >**

**< author >** автор словаря/авторы (через запятую) **< /author >**

**< publisher >** название издательства **< /publisher >**

**< editor >** редактор **< /editor >**

< **edition** > редакция < /**edition** >  
< **translator** > переводчик < /**translator** >  
< **pubdate** > год издания < /**pubdate** >  
< **isbn** > международный стандартный книжный номер < /**isbn** >  
< **language type=content/source/target** > язык, на котором написан словарь/перевод с какого языка/на какой  
    < **idno type=iso639-3** > код языка < /**idno** >  
    < /**language** >  
    < **script type="тип орфографии"** > < /**script** >  
< /**head** >  
< /**front** >

Nota Bene: для современной русской орфографии в < **script type=""** > нужно указать "rus".

Блок < **entry** > может относиться к одному из семи перечисленных ниже типов.

< **entry** > обязательно содержит блок < **form** > , который, в свою очередь, относится к одному из пяти перечисленных ниже специальных типов (если тип общий, он не указывается).

В < **form** > указывается заголовочное слово, ударение/транскрипция и различные пометы. В случае если форма слова не является начальной (type "inflection"), падеж/число/род и пр. указываются в блоке < **inflection** > .

< **entry type="1)main 2)hom[homograph] 3)xref[cross reference] 4)affix 5)abbr[abbreviation] 6)supplemental 7)foreign"** >

    < **index** > для словаря Зализняка < /**index** >  
    < **symbol** > для словаря Зализняка < /**symbol** >  
    < **form type="/inflected/abbreviation/full/contradiction/compound"** >  
        < **orth type="lemma" extent="1)full[full form] 2)pref[prefix] 3)suff[suffix] 4)inf[infix] 5)part[partial]"** >  
заголовочное слово < /**orth** >  
    < **orth type="var"** > вариант написания заголовочного слова < /**orth** >  
    < **hyph** > вариант написания слова через дефис < /**hyph** >

< phon type="main/var/alter[чередование]" > транскрипция < /phon >  
    < stress left="" right="" > ударение < /stress >  
< usg type="acc" > прав./неправ. < /usg >  
< usg type="dom" > профессиональное поле < /usg >  
< usg type="geo" > местность < /usg >  
< usg type="plev" > частотность < /usg >  
< usg type="reg" > тип < /usg >  
< usg type="style" > стиль < /usg >  
< usg type="time" > время < /usg >  
< usg type="editor" > пометы редактора < /usg >  
< usg type="other" > другие пометы про употребление < /usg >  
< lbl > пометы не про употребление < /lbl >  
< inflection >  
    < case > падеж < /case >  
    < num > число < /num >  
    < tns > время < /tns >  
    < gender > род < /gender >  
    < orth > всё слово, если не выделена изменяемая часть (е.g. матери); несколько форм слова через запятую (е.g. take, took, taken) < /orth >  
    < orth extent="part" > выделенная изменяемая часть слова (е.g. -ери) < /orth >  
    < phon extent="" > транскрипция < /phon >  
< /inflection >

Nota Bene: тег < usg > заполняется в соответствии с унифицированным блоком помет.

В < form > также вкладывается блок < gramGrp > со всей грамматической информацией о слове.

< gramGrp >  
    < paradigm > парадигма < /paradigm >  
    < pos > часть речи < /pos >  
    < gender > род < /gender >

< **num** > число < /**num** >  
< **per** > лицо < /**per** >  
< **animat** > одушевлённость/неодушевлённость < /**animat** >  
< **declen** > склонение < /**declen** >  
< **mood** > наклонение < /**mood** >  
< **asp** > совершенный/несовершенный вид < /**asp** >  
< **transit** > переходность/непереходность < /**transit** >  
< **refl** > возвратность < /**refl** >  
< **genus** > залог < /**genus** >  
< **pron\_type** > личное/неличное местоимение < /**pron\_type** >  
< **gov** > управление (для предлогов и глаголов:с какими падежами используются) < /**gov** >  
< **iType** > словоизменительный класс < /**iType** >  
< /**gramGrp** >  
< /**form** >

Nota Bene: тег < **pos** > заполняется в соответствии с унифицированным блоком помет.

В < **sense** > , следующем блоке в составе < **entry** > , главным образом, содержится определение слова в блоке < **def** > . Здесь также допускаются различные пометы.

< **sense** n="1/2/3..." >  
  < **usg** type="" > см. выше < /**usg** >  
  < **lbl** > см.выше < /**lbl** >  
  < **def** >  
    < **text** > текст определения < /**text** >  
    < **ref** type="/external/xr code="" > ссылка < /**ref** >  
    < **src** code="" > источник, откуда взято определение < /**src** >  
    < **pic** > картинка < /**pic** >  
  < /**def** >

Наряду с блоком **< def >** в **< sense >** может быть блок **< cit >** . Здесь приводятся примеры, переводы, коллокации, идиомы и пр., типы перечислены ниже.

**< [cit](#) type="1)example 2)translation 3)example\_translation 4)collocation 5) idiom 6)gender\_parallel 7)etym\_true 8)etym\_false 9)etym 10)cognate 11)compare" >**

**< quote >** текст примера/перевода и т.д. **< /quote >**

**< form >** текст примера/перевода и т.д. **< /form >**

**< sense >** текст примера/перевода и т.д. **< /sense >**

**< language type="content/source/target/example/etym\_source/etym\_target/etym" >**

**< idno type="iso639-3" >** код языка **< /idno >**

**< /language >**

**< script type="" > < /script >**

**< ref type="/external" code="" > < /ref >**

**< src code="" > < /src >**

**< /cit >**

**< /sense >**

В состав **< entry >** также может входить блок для вложенных входов **< re >** .

**< [re](#) >** вложенные входы: всё, что может быть в entry **< /re >**

В блок **< etym >** в составе **< entry >** содержится вся информация об этимологии.

**< etym >**

**< text >** текст этимологической статьи **< /text >**

**< cit type="etym\_true/etym\_false/etym" > < /cit >**

**< /etym >**

Последний возможный блок **< entry >** , **< xr >** , включает в себя информацию о синонимах/антонимах и т.д.

< xr type="synonym/antonym/hyperonym/hyponym/analogue/converse" >  
    < lbl > < /lbl >  
    < ref > слово < /ref >  
    < set > синонимические/антонимические ряды и т.д. < /set >  
  < /xr >  
< /entry >

### NOTA BENE:

- 1) В шаблоне используется ряд так называемых свободных тегов, которые можно ставить в абсолютно любое место: < language >, < stress >, < usg >, < lbl >, < cit >, < ref >, < note >, < cut >
- 2) < note > используется во всех случаях, когда информация не помещается ни в один из имеющихся тегов.
- 3) < cut > используется для сокращения длинных статей.
- 4) < gramGrp > может быть вложен не только в < form > , но также в < sense > и < cit >