**Точное ТЗ для программиста**

1. Распарсить эти вордовские файлы (все три тома), превратить их в машиночитаемый иксель.

<https://drive.google.com/drive/folders/19YJhBLWB1JCOabShvaWlLe4rTp0TXl6s>

2. По ключу (видимо, это будет первая строчка стихотворения) привязываем эти варианты к стихотворениям из нашей основной базы. Все это нужно, чтобы потом иметь возможность на сайте представлять разные варианты стихотворений, из разных изданий.

3. Дополнить xml-файлы стихотворений шапками, куда мы внесем еще разные метаданные, которые частично уже собраны.

<https://drive.google.com/drive/u/0/folders/0BxF2JHOecpGhT1JNRkZxRnYtX1k>

У стихов в шапке должны быть:

1. уникальный id (проще всего - по первой строчке + по краткому обозначению издания)
2. дата создания
3. вариант издания, по которому приводится текст. Но здесь нужно продумать - нужно сделать так, чтобы разные варианты одного текста были связаны между собой.
4. много разных строчек для остальных метаданных (размер, кол-во строк и т.д.), можно пока оставить пустые строки для дальнейшего заполнения
5. кто ответственен за этот проект и подготовку файлов (типа, Высшая школа экономики, Мандельштамовское общество, центр Digital humanities + персонально, по именам).

2. Связать конкорданс со стихотворениями (по номеру стихотворения в файлах и в словах конкорданса).

Конкорданс по частям речи: <https://rvb.ru/mandelstam/m_o/concordance/index.htm> (нужно скачать архив с файлами)