

# 在线对齐产品手册

2021年8月

# 目录

1. 简介.....	3
(1) 简介.....	3
(2) 主要特点.....	4
(3) 支持文件类型.....	4
(4) 支持语种.....	5
(5) 文件大小限制.....	5
(6) 收费标准.....	5
2. 操作流程.....	6
(1) 选择对齐模式.....	6
(2) 导入文档.....	7
(3) 段落调整.....	8
(4) 句对调整.....	9
(5) 导出文件.....	11
3. 特色功能.....	16
(1) PDF转换.....	16
(2) 快捷键和右键菜单.....	16

# 1. 简介

## (1) 简介

Tmxmall 在线对齐是一款基于浏览器架构的专业语料对齐平台。该平台技术先进、对齐效率高、操作简单。Tmxmall 在线对齐提供了非常人性化的交互界面，方便用户快捷地调整对齐结果，极大程度上提高文档对齐效率和用户对齐体验。其自主研发的智能对齐算法可以自动对齐原文及译文语料中“一对多、多对一、多对多”的句子，使得原本需要人工介入的连线调整工作完全被自动化程序替代，从而大幅降低人工干预的工作量，使对齐工作真正变得高效。具有支持语料制作、TMX文件在线编辑、术语提取、语料去重、查找与替换等功能。



## (2) 主要特点

- 单/双文档对齐

双文档对齐：原文与译文在两个文件中对齐；

单文档对齐：原文与译文以上下/左右形式对照在单个文件中对齐。

- 多语言对

涵盖46种语言，2070种语言对。支持中文、英语、韩语、日语、俄语、德语、西班牙语、法语、阿拉伯语、土耳其语、缅甸语、藏语、彝语、拉丁语等。

- 智能对齐

能自动识别“一对多”、“多对一”、“多对多”句子对应关系，极大地提高对齐的效率以及准确度。

- 支持多种格式

docx, xlsx, pptx, pdf, txt等37种主流格式的导入；tmx, xlsx, docx, txt共4种格式的导出。

- 术语提取

融合深度学习技术，在记忆库制作完成基础上可实现双语术语的智能提取。

- 高级功能

完全重复句对一键去重；一句多译句对一键筛选；原文与译文相同内容一键筛选；检索词快速定位。

## (3) 支持文件类型

- 办公文件

doc/docx、xls/xlsx、ppt/pptx、rtf、pot、dot、vsdx、ods、ots、odg、otp、odt、ott

- 网页文件

html、htm、xhtml

- 技术写作文件

dita、ditamap

- 本地化文件

xliff、sdlxf、xlf、xml

- 排版文件

md、idml、mif、srt

- 其他文件

pdf、chm、txt、csv、yml、json、tmx

#### **(4) 支持语种**

涵盖中文、英语、日语、韩语、俄语、德语、西班牙语、法语、葡萄牙语、阿拉伯语、缅甸语、印尼语、柬埔寨语、菲律宾语、泰语、越南语、老挝语、马来语、希腊语、意大利语、土耳其语、乌克兰语、瑞典语、希伯来语、斯洛伐克语、拉丁语、捷克语、孟加拉语、亚美尼亚语、波兰语、丹麦语、荷兰语、芬兰语、匈牙利语、罗马尼亚语、印地语、藏语、维语、彝语、保加利亚语、斯洛文尼亚语、克罗地亚语、爱沙尼亚语、立陶宛语、马其顿语、阿尔巴尼亚语共46种语言。

#### **(5) 文件大小限制**

文档最大为20M（超出20M请在刷新或关闭浏览器前及时导出记忆库，以防数据丢失）。

#### **(6) 收费标准**

- ① 导出每份文件2元。
- ② 套餐价格：
  - 9.9元（可导出10份文件）
  - 49.9元（可导出60份文件）
  - 99.9元（可导出150份文件）
  - 199.9元（可导出500份文件）

## 2. 操作流程

本节将介绍如何利用Tmxmall在线对齐 (<https://www.tmxmall.com/aligner>) 完成记忆库的制作。使用前请先注册并登录Tmxmall账号，即可开始使用。开始对齐前，可将光标停在页面右上角的“帮助”，了解基本操作。

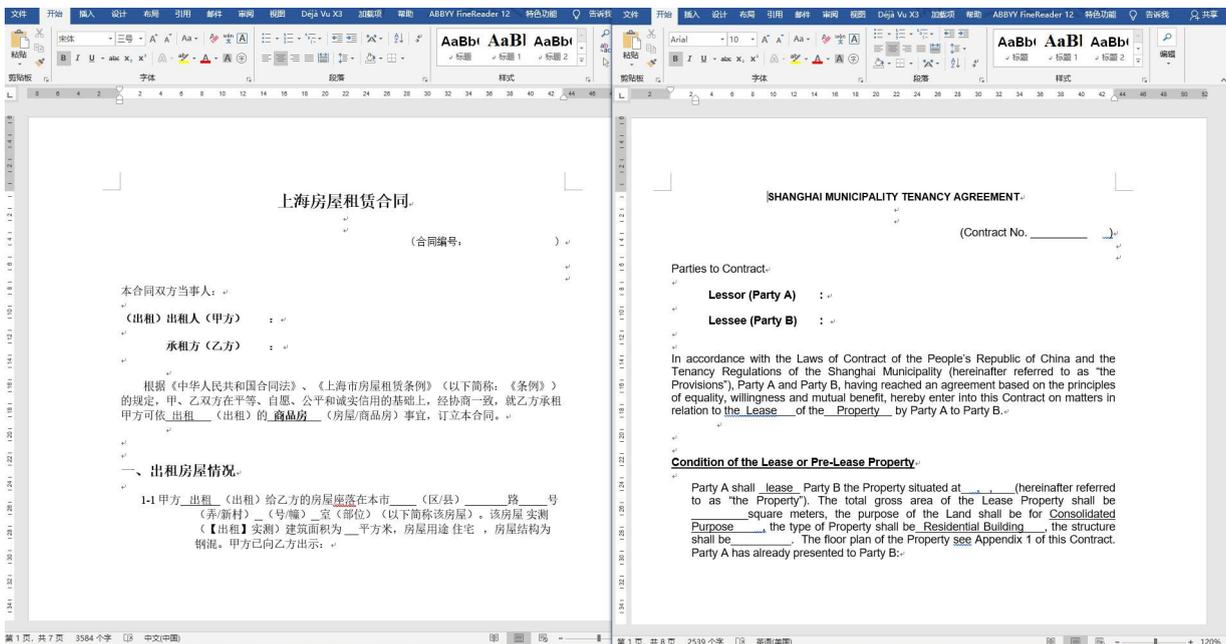


### (1) 选择对齐模式

根据用户文档类型不同，需选择对齐模式：双文档对齐或单文档对齐。

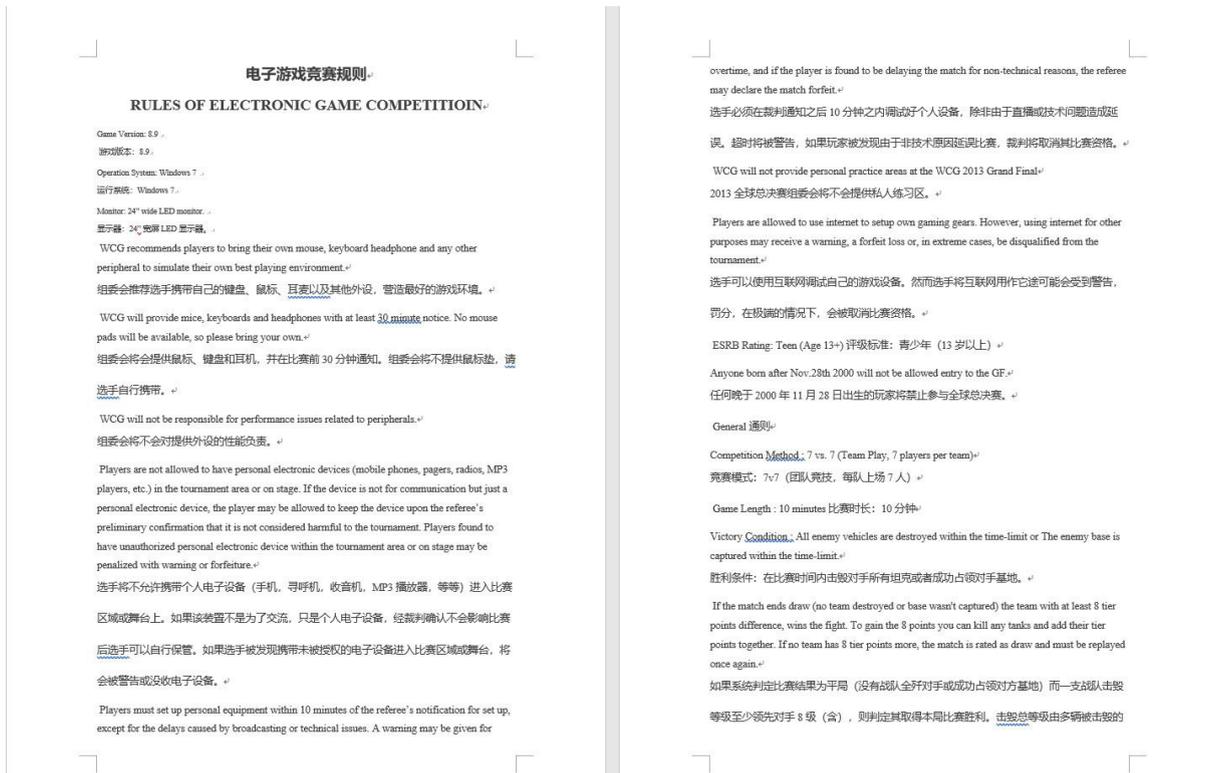
#### ① 双文档对齐

双文档对齐：原文与译文在两个文件中对齐。例如一个中文文档与一个英文文档。



## ① 单文档对齐

单文档对齐：原文与译文以上下/左右形式对照在单个文件中对齐。例：原文译文是段落上下对照模式的一个文档。



## (2) 导入文档

### ① 双文档导入

单击双文档对齐：分别单击蓝色文件夹图标，导入待对齐的文件。



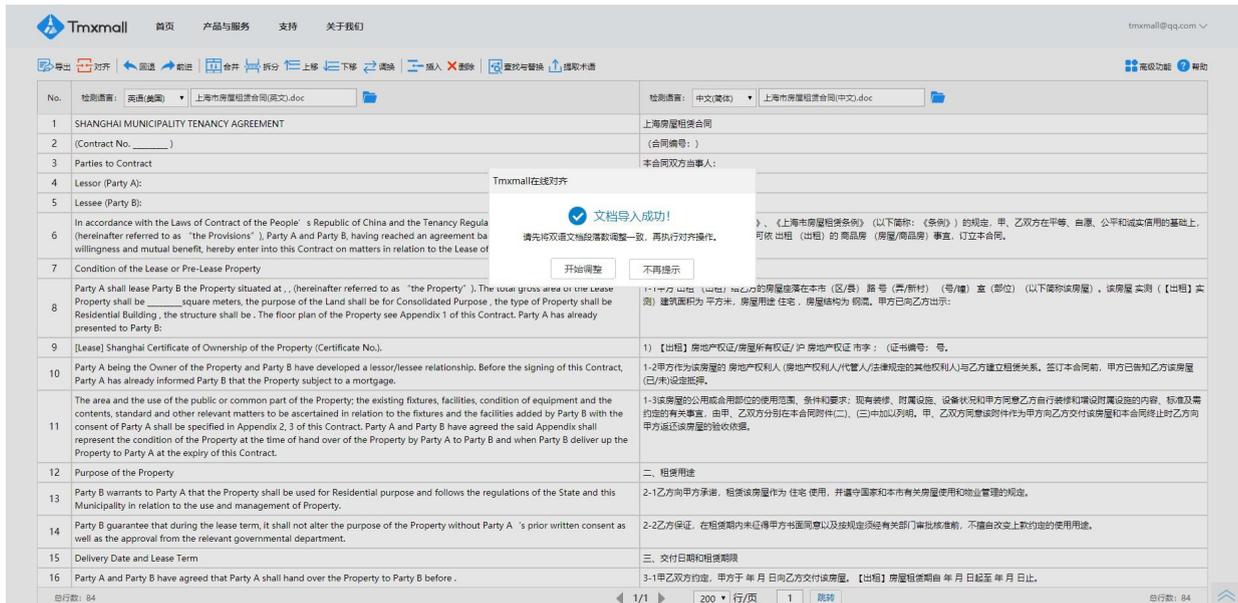
### ② 单文档导入

单击单文档对齐：单击蓝色文件夹图标，导入待对齐的文件。



### (3) 段落调整

文档导入成功后，用户可通过合并、拆分、上移、下移、调换、插入和删除等操作手动调整文档，使左右两列段落语义对应，并将段落调整至总行数一致。



#### 操作方法：

选中：

- ① 选中单元格：单击需选中的单元格；
- ② 选中整行：单击该行所在的序号；
- ③ 多选单元格（行）：按住Shift（或Ctrl）并单击单元格（序号）。

编辑：双击单元格进入编辑状态，可编辑修改单元格文本内容。

回退：撤销上一步操作。

前进：恢复上一步操作。

合并：选中某单元格（行），按住Shift（或Ctrl）并选中其余需要合并的单元格（行），点击“合并”。

拆分：双击单元格进入编辑状态，鼠标单击需要拆分的位置，点击“拆分”。

上移：选中（可多选）单元格（行），点击“上移”。

下移：选中（可多选）单元格（行），点击“上移”。

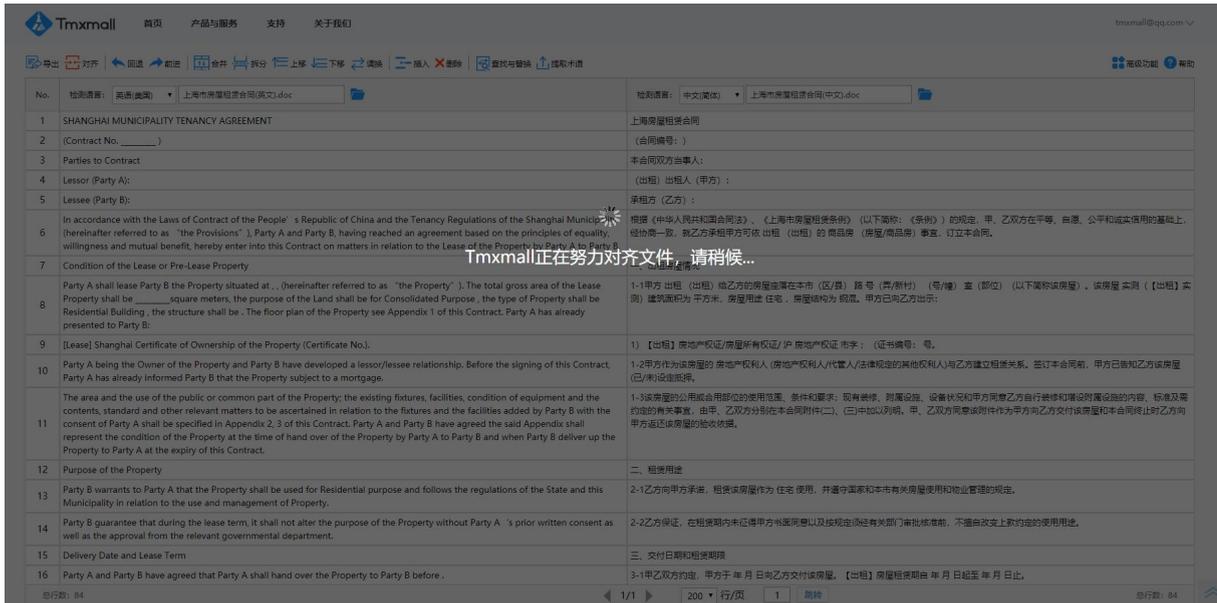
调换：选中（可多选）单元格（行），点击“调换”。

插入：选中某单元格（行），点击“插入”。

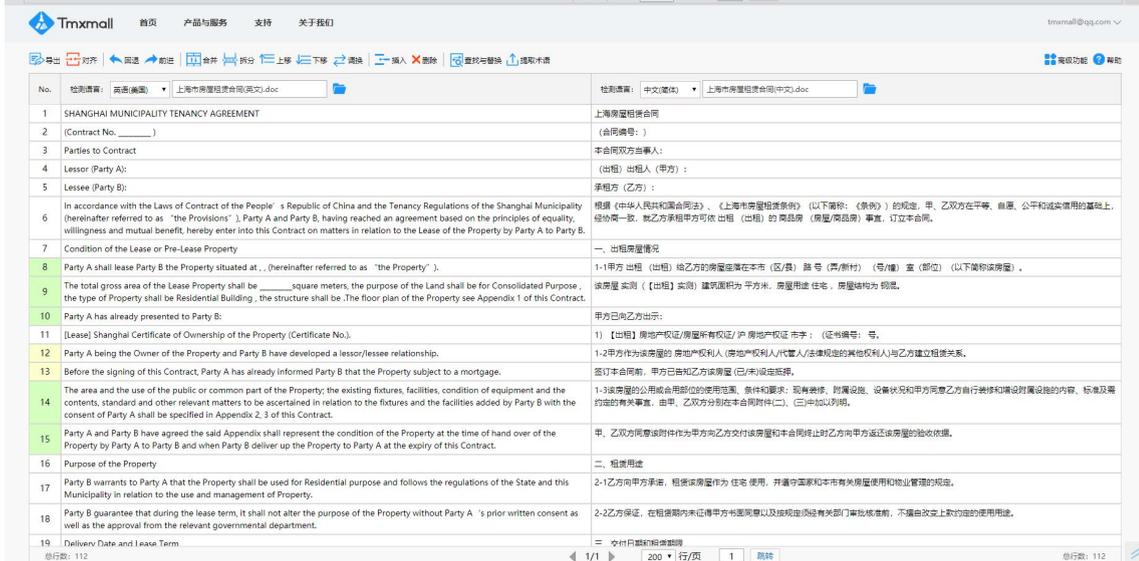
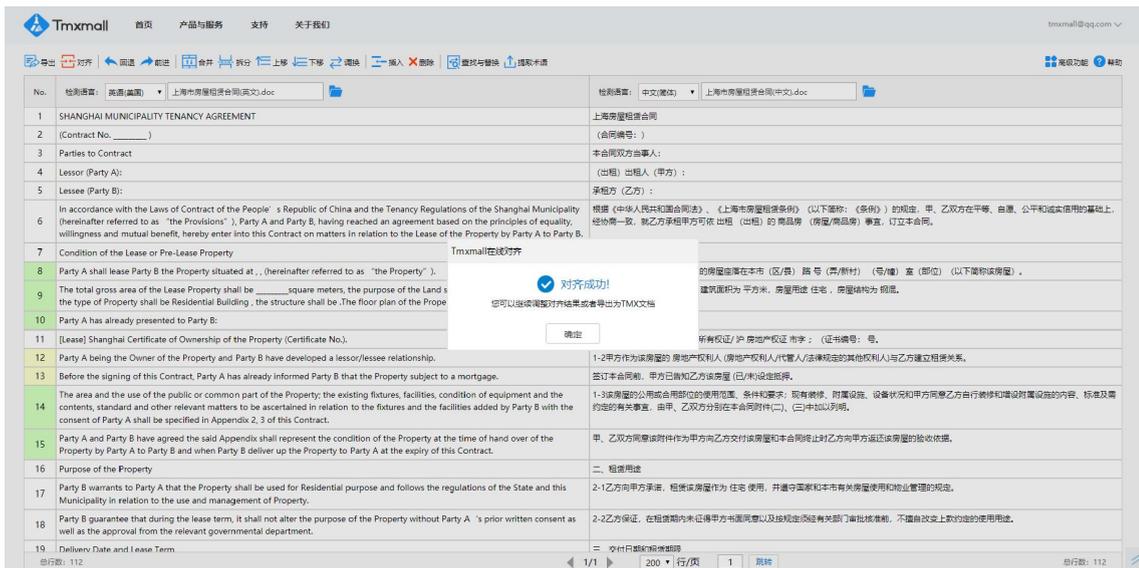
删除：选中（可多选）要删除的单元格（行），点击“删除”。

## (4) 句对调整

段落调整完成后，单击“对齐”按钮，系统将自动将段落拆分成句对。



拆分完成后，用户只需通篇检查拆分是否正确，并通过合并、拆分、上移、下移、调换、插入和删除等操作将拆分错误句对进行微调，将句对调整至总行数一致即可。



## ① 操作方法:

选中:

- ① 选中单元格: 单击需选中的单元格;
- ② 选中整行: 单击该行所在的序号;
- ③ 多选单元格(行): 按住Shift(或Ctrl)并单击单元格(序号)。

编辑: 双击单元格进入编辑状态, 可编辑修改单元格文本内容。

回退: 撤销上一步操作。

前进: 恢复上一步操作。

合并: 选中某单元格(行), 按住Shift(或Ctrl)并选中其余需要合并的单元格(行), 点击“合并”。

拆分: 双击单元格进入编辑状态, 鼠标单击需要拆分的位置, 点击“拆分”。

上移：选中（可多选）单元格（行），点击“上移”。

下移：选中（可多选）单元格（行），点击“上移”。

调换：选中（可多选）单元格（行），点击“调换”。

插入：选中某单元格（行），点击“插入”。

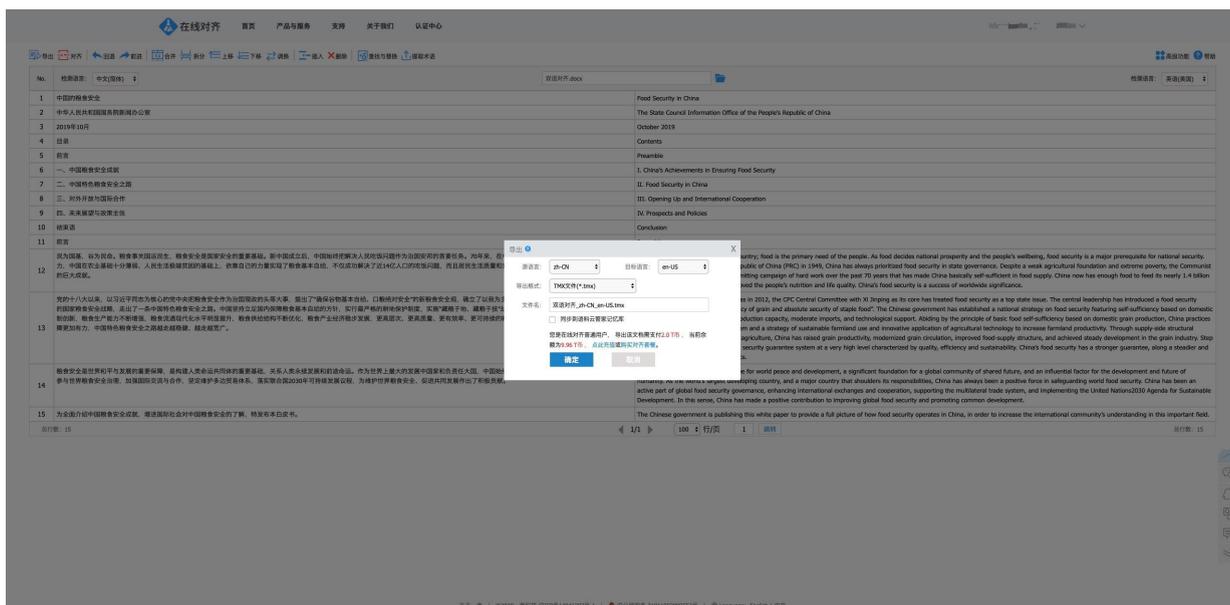
删除：选中（可多选）要删除的单元格（行），点击“删除”。

## ② 颜色标识

绿色和黄色表示含义：系统执行句对齐时，所有执行过段落拆分的句子都会用绿色或黄色标记出来，方便使用者在检查环节重点校对句对拆分的准确性。用绿色和黄色分别代表奇数段和偶数段，以示对相邻两个段落加以区分，该功能可以有效防止用户将不属同一段落的句子错误合并。

## (5) 导出文件

将两列的句对总行数调整一致后，单击“导出”，即可导出对齐文件。用户可自行选择语言方向和导出的格式。



单击“导出”后，浏览器会自动弹出下载文件窗口，即可将记忆下载至本地

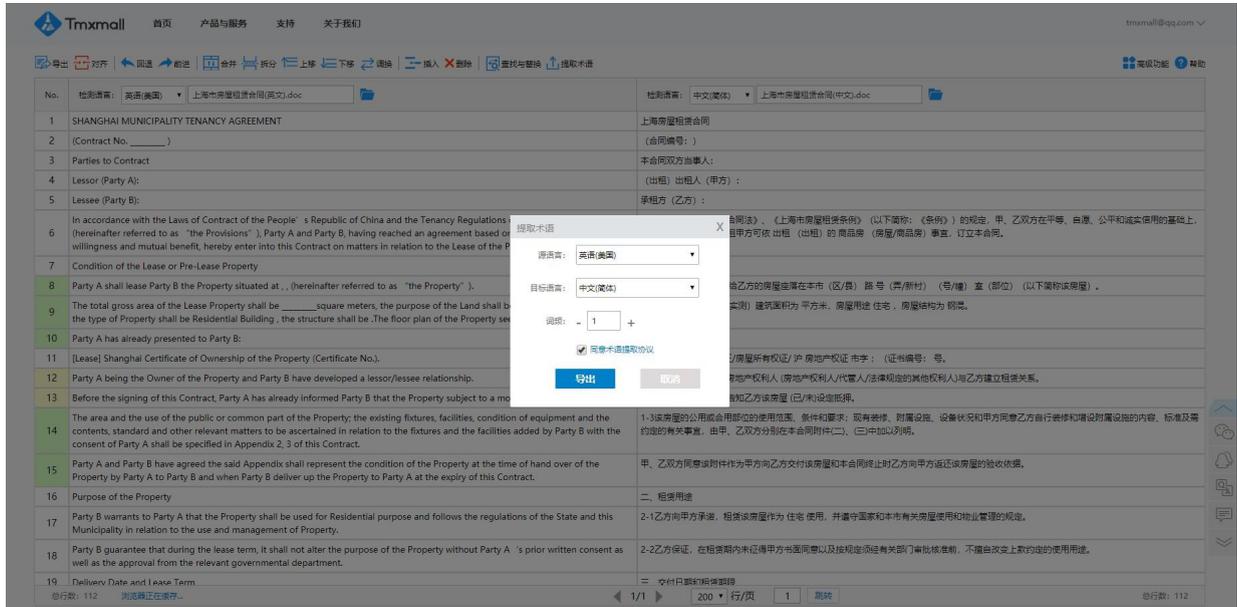
**注：导出文件时仅支持浏览器直接下载，暂不支持迅雷等第三方下载工具。如需取消迅雷等默认下载工具，可参考对齐页面右上角帮助内容；在线对齐为付费服务，导出前需确保T币余额充足。**

## 导出格式

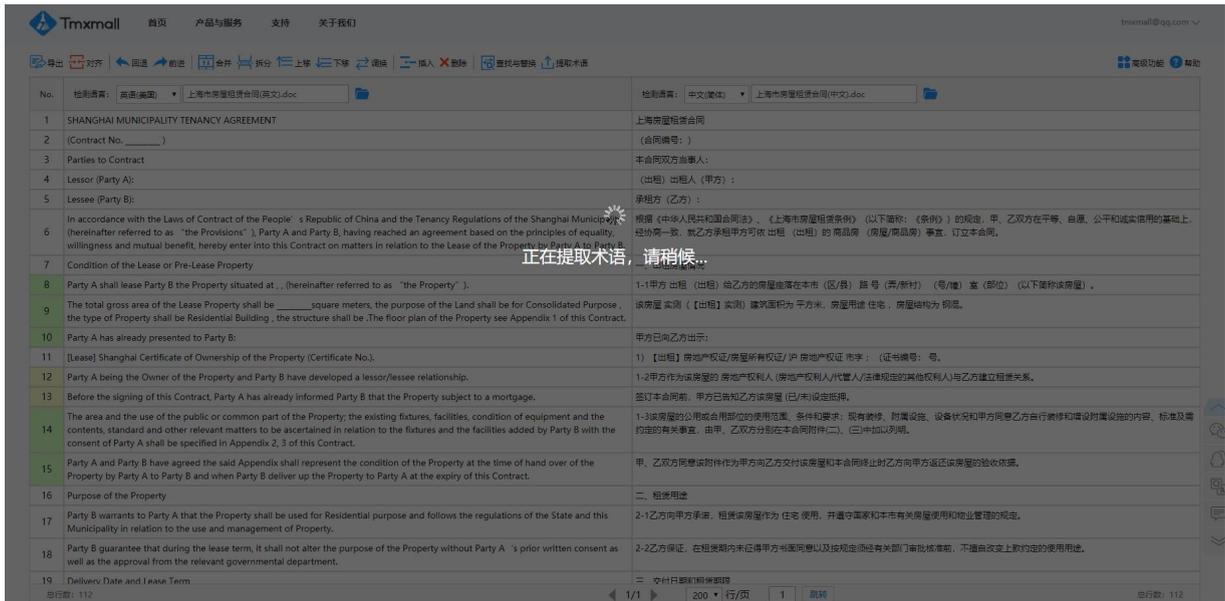
导出格式支持TMX文件 (\*.tmx)、Excel工作簿 (\*.xlsx)、上下对照的Word文档 (\*.docx)、上下对照的文本文档(换行符分隔) (\*.txt)和左右对照的文本文档(制表符分隔) (\*.txt)。

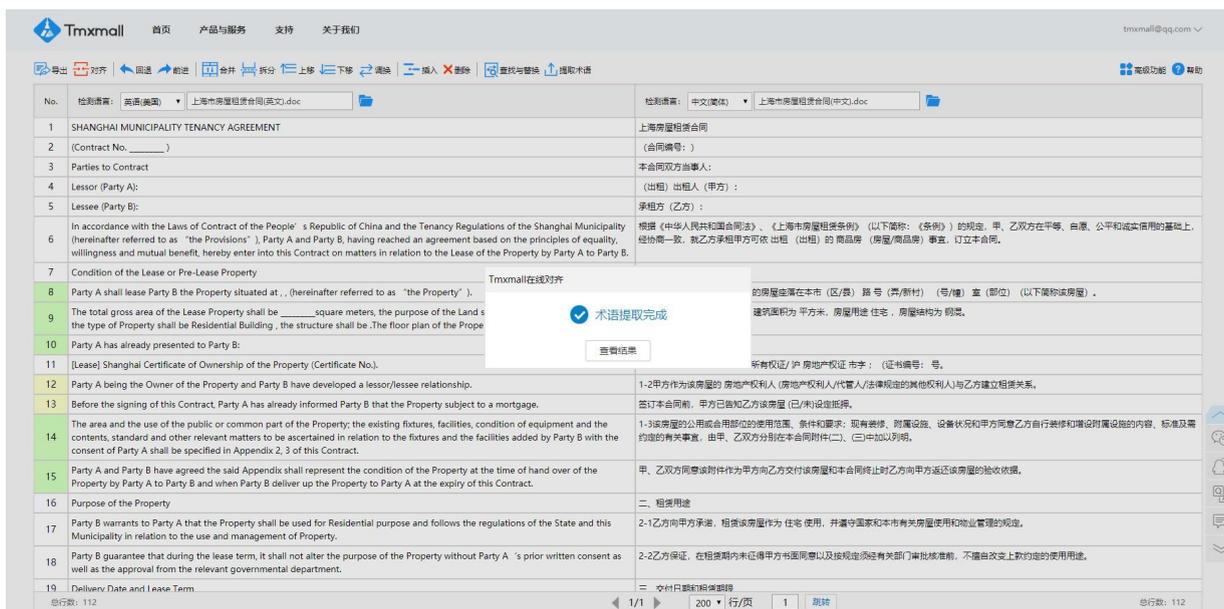
## 术语提取

记忆库制作完成后,可根据需要进行术语提取(仅支持中英术语提取)。单击“提取术语”,设置语言方向和词频。



单击“导出”后,系统将自动提取术语,提取完成后单击“查看结果”。





用户可设置词频筛选术语，并且可以对术语内容进行编辑或删除。编辑完成后，用户可选择勾选要导出的术语，将其导出Excel。

Excel导出

Excel导出(含词频)

## 语料清洗

在语料库制作完成的基础上，为提高语料纯度，还可以按需进行语料清洗工作。

## 查找与替换

用户可查找文本或将其替换为其他内容。

查找下一个：定位到下一个检索词。

替换并查找：替换成当前词并查找下一个检索词。替换：替换成当前词。

替换所有：将所有检索词统一替换成替换词。区分大小写：可按照区分大小写检索。

The screenshot shows the Tmxmall software interface. At the top, there are navigation links: 首页 (Home), 产品与服务 (Products and Services), 支持 (Support), and 关于我们 (About Us). Below the navigation bar is a toolbar with various editing tools like 导出 (Export), 对齐 (Align), 回退 (Undo), 前进 (Forward), 合并 (Merge), 拆分 (Split), 上移 (Move Up), 下移 (Move Down), 调换 (Swap), 插入 (Insert), 删除 (Delete), 查找与替换 (Find and Replace), and 提取术语 (Extract Terms). The main area displays a document with a search and replace dialog box open. The dialog box has two tabs: '查找与替换' (Find and Replace) and '提取术语' (Extract Terms). The '查找与替换' tab is active, showing search criteria: '检测语言: 英语(美国)' (Detection Language: English (US)) and '上海市房屋租赁合同(英文)' (Shanghai Municipality Tenancy Agreement (English)). The search term is 'Party A' and the replacement term is '甲方'. There are buttons for '查找下一个' (Find Next), '替换并查找' (Replace and Find), '替换' (Replace), and '替换所有' (Replace All). The document content is partially visible, showing clauses 1 through 13.

## 筛选

通过输入检索词进行批量筛选。单击“高级功能-筛选”，在文本框中输入要检索的关键词，页面会自动筛选出含有该关键词的句对。关键词可区分大小写。

The screenshot shows the Tmxmall software interface. At the top, there are navigation links: 首页 (Home), 产品与服务 (Products and Services), 支持 (Support), and 关于我们 (About Us). Below the navigation bar is a toolbar with various editing tools like 导出 (Export), 对齐 (Align), 回退 (Undo), 前进 (Forward), 合并 (Merge), 拆分 (Split), 上移 (Move Up), 下移 (Move Down), 调换 (Swap), 插入 (Insert), 删除 (Delete), 查找与替换 (Find and Replace), and 提取术语 (Extract Terms). The main area displays a document with a search and replace dialog box open. The dialog box has two tabs: '查找与替换' (Find and Replace) and '提取术语' (Extract Terms). The '查找与替换' tab is active, showing search criteria: '检测语言: 中文(简体)' (Detection Language: Chinese (Simplified)) and '上海市房屋租赁合同(中文).doc' (Shanghai Municipality Tenancy Agreement (Chinese).doc). The search term is 'party' and the replacement term is '甲方'. There are buttons for '查找下一个' (Find Next), '替换并查找' (Replace and Find), '替换' (Replace), and '替换所有' (Replace All). A '高级功能-筛选' (Advanced Function - Filter) dialog box is also visible, showing the search term 'party' and a checkbox for '区分大小写' (Case Sensitive). The document content is partially visible, showing clauses 1 through 15.

## 语料去重

语料清洗包括清洗：（1）未翻译句段 （2）原文与译文相同句段 （3）原文相同，译文不同句段 （4）原文不同，译文相同句段 （5）去除重复句段

### 未翻译句段

“未翻译句段”指TMX文件中若有译文空白句对，可通过此功能进行筛选。

### 原文与译文相同句段

“原文与译文相同句段”指筛选TMX文件中原文与译文（左右两列）完全相同的句对，用户可按需选择性进行删除。删除完成后，单击高级功能面板中的关闭按钮，可显示剩余数据。

### 原文相同，译文不同句段

“原文相同，译文不同句段”指筛选TMX文件中原文相同，译文不同句段，并可根据需要选择“保留第一条”。

删除完成后，单击高级功能面板中的关闭按钮，可显示剩余数据。

### 原文不同，译文相同句段

“原文不同，译文相同句段”指筛选TMX文件中原文不同，译文相同句段，并可根据需要选择“保留第一条”。

删除完成后，单击高级功能面板中的关闭按钮，可显示剩余数据。

### 去除重复句段

去除重复句段是去除TMX文件中内容完全重复的句对，只保留一条句对。

*注：该操作无法撤销，请谨慎操作。*

### 3. 特色功能

利用特色功能，可使得对齐变得更加高效。

“PDF转换”可将扫描版PDF转换为可识别文字版文件，“快捷键和右键菜单”可在对齐过程中极大提高对齐效率。

#### (1) PDF转换

如有扫描版的PDF，需将文件转换成可识别文字后再执行对齐操作。单击在线对齐页面上“扫描版PDF？点此转换”可进入PDF转Word转换器页面。可将PDF文件直接拖入进行格式转换和文件下载。



#### (2) 快捷键和右键菜单

为提高对齐效率，在线对齐引入了快捷键和右键菜单功能。用户在对齐过程中利用快捷键结合右键菜单，可大幅提高对齐效率。

##### \*快捷键说明：\*

回退（Ctrl+Z）：撤销上一步操作。

前进（Ctrl+Y）：恢复上一步操作。

合并 (Ctrl+C/M)：选中某单元格 (行)，按住Shift (或Ctrl) 并选中其余需要合并的单元格 (行)，点击合并，完成合并句对操作。

拆分 (Ctrl+Enter)：双击单元格进入编辑状态，鼠标单击需要拆分的位置，点击拆分，完成拆分句对操作。

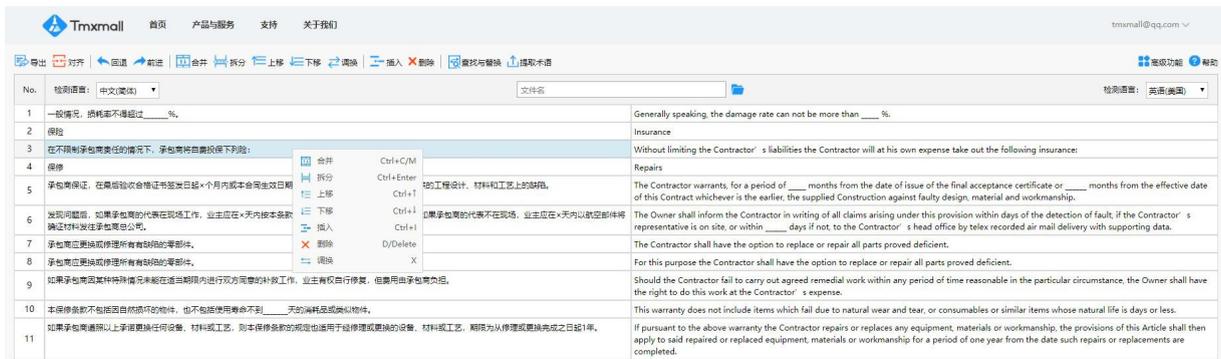
上移 (Ctrl+↑)：选中 (可多选) 单元格 (行)，点击上移，完成操作。

下移 (Ctrl+↓)：选中 (可多选) 单元格 (行)，点击下移，完成操作。

调换 (X)：选中 (可多选) 单元格 (行)，点击对调，完成同一行左右两列内容对调。

插入 (Ctrl+I)：选中某单元格 (行)，点击插入，完成操作。

删除 (D/Delete)：选中要删除的单元格 (行)，点击删除，完成操作。



## 4. 常见问题

### (1) 使用在线对齐时，为何提示文件上传错误？

- ① 请检查您的文件格式是否支持。
- ② 请检查您的文件大小是否超出限制。

请检查您的文件类型是否支持，在线对齐支持导入的文件类型有：

#### 双文档对齐：

（Office系列） doc/docx、xls/xlsx、ppt/pptx、wps、rtf、pot、odt等；

（网页文件） xml、html、xhtml等；

（其他常用文件） chm、txt、pdf等。

#### 单文档对齐：

（Office系列） doc/docx、xls/xlsx、ppt/pptx、wps、rtf、pot、odt等；

（网页文件） xml、html、xhtml等；

（其他常用文件） chm、tmx、txt、pdf等。

其中txt文档最大不超过1M；其余类型的文档最大不超过20M。

在线对齐中绿色和黄色代表什么？

绿色和黄色表示执行句对齐时，所有执行过段落拆分的句子都会用绿色或黄色标记出来，方便使用者在检查环节重点校对句对拆分的准确性。用绿色和黄色分别代表奇数段和偶数段，以示对相邻两个段落加以区分，该功能可以有效防止用户将不属同一段落的句子错误合并。

### (2) 使用在线对齐或在线翻译时电脑突然死机或断电怎么办？

在线对齐或在线翻译皆包含本地缓存功能（文件大小在5M之内），在您下次启动Tmxmall在线对齐或在线翻译时，系统会提示您是否继续完成上次的操作。

### (3) 下载的记忆库tmx版本默认是多少？

平台默认下载记忆库tmx版本为 tmx 1.4。tmx 1.4可在当今主流CAT工具中使用。如在tmxmall平台下载的记忆库不能成功导入Trados2006、Trados2007等软件，可加官方QQ群（435947659）寻求技术帮助。

#### **(4) 购买Tmxmall相关服务后如何索取发票？**

Tmxmall默认开具**电子发票**，您可在下方表单中填写发票信息，工作人员收到申请后会在**七个工作日**内开具并发送至您的邮箱。

#### **(5) 人工语料对齐的收费标准**

10元/1000中文字符起（不计空格），需评估客户样例文本的具体格式和处理难度后再给出定价。

#### **(6) 人工语料对齐的整体流程？**

语料评估→报价→支付预付款→交付语料库→支付尾款

#### **(7) 人工语料对齐能否提供发票？**

可提供发票（普票/专票），默认开具**增值税普通发票**。

#### **(8) 人工语料对齐是否有保密协议？**

双方可以签订人工语料对齐业务数据保密协议，盖章后生效。

#### **(9) 人工语料对齐的准确率是多少？**

默认对齐准确率高于99%，具体根据客户的原文本质量和对齐要求进行商定。