1、中文维基百科数据解压后建议用EmEditor打开；

2、维基百科的每个页面被包含在<page>元素中，以</page>结束；

3、<page>里的子元素 <title>为页面的标题，<text>为维基百科正文内容；

4、需要注意：

* {{ }} 里面的内容对应为页面的表格，如

；

* '''小行星872'''代表页面中的粗体字，；
* == 参考资料 == 代表页面中的标题，；
* [[美国]] 代表含有超链接。