Atitit 文本文档的压缩解决方案（无损压缩与有损压缩 au

文本文档的压缩有俩个方面的压缩，一个是内容压缩，一个是样式的压缩。

文本的内容与样式相互分离，可以采用不同的压缩技术。。

文本文档的内容无损压缩，主要是可以多占用格式大的格式转换为小的格式。比如对word，pdf转换为html，转化为markdown ，转化为txt等。。 对文本内容无损，但是对样式稍有损伤。。

文本文档的内容有损压缩。。主要采用摘要的方式。以及提取子标题，提取大纲结构模式来压缩。抽取骨架