text extractor test