osstex textractor

doc parser