

No anonymizations, so tika will read the text and will do it correctly.;...