Word–topic probabilities – wget website scrape