

تفاوت اصلی بین "Lemmatization" و "Stemming" در پردازش متن این است که:

:Stemming

این فرآیند به حذف پیشوندها و پسوندهای کلمات به منظور یافتن ریشه (stem) کلمه می‌پردازد. این معمولاً با استفاده از الگوریتم‌های ساده مانند Porter Stemmer یا Snowball Stemmer انجام می‌شود. نتیجه stemming ممکن است به یک کلمه غیرواقعی یا اشتباهی منجر شود.

:Lemmatization این فرآیند تلاش می‌کند کلمات را به شکل اساسی یا لم (lemma) خود تبدیل کند. این به معنای انتقال کلمات به شکل معنایی و موجود در دیکشنری یا واژه‌نامه است. این فرآیند معمولاً با استفاده از دیکشنری‌ها و قوانین گرامری برای هر زبان انجام می‌شود و نتایج دقیق‌تری از نظر معنا و کاربرد دارد.