

برچسب گذاری اجزای واژگانی کلام عمل انتساب برچسبهای واژگانی به کلمات و نشانههای تشکیلی دهنده متن است، به صورتی که این برچسبها نشاندهنده نقش کلمات و نشانهها در جمله باشد. برچسبگذارهای اجزای واژگانی کلام و پیکرههای برچسب خورده با این برچسبها در بسیاری از حوزههای دیگر پردازش زبان طبیعی مورد استفاده قرار میگیرند که از این آنها میتوان به خطاب، تبدیل متن به گفتار، سیستمهای تشخیص خودکار گفتار و ترجمه ماشینی اشاره کرد.

پیکره متنی زبان فارسی پیکرههای است که مجموعه برچسب آن شامل ۹۰ برچسب است که ۱۴ عدد از آنها برچسبهای مقولات اصلی است. ساختار برچسب کلمات در پیکره مذکور بر اساس استاندارد انگلیز و به صورت سلسله مراتبی است. استفاده از ساختار برچسبگذاری سلسله مراتبی امکان نمایش مقولات اصلی، زیرمقولات، وندهای تصریفی، واژهبست و دیگر ویژگیهای کلمات را فراهم میسازد. استفاده از ساختار سلسله مراتبی باعث شده که تعداد برچسبهای پیکره به بیش از ۵۸۰ مورد برسد.

به تازگی وبگاه «پردازش زبان فارسی» شروع به کار نموده است. در این وبگاه با استفاده از پیکره متنی زبان فارسی یک برچسبگذار اجزای واژگانی ایجاد شده که نسخه اول آن به نمایش گذاشته شده است. مجموعه برچسب برچسبگذار شامل مقولات اصلی کلمات است. دقت برچسبگذار ۹۵٪ است.

این است که نسخه جدید برچسبگذار به زودی عرضه گردد. در نسخه جدید تعداد برچسبهای مجموعه برچسب افزایش یافته (تا حدود ۴۰۰ برچسب) تا ویژگیهای بیشتری از کلمات را بتوان نشان داد.

آدرس این وبگاه <http://www.persianp.ir> میباشد.