
Morfolojik Analiz

— Fatih Emin Karahan - 502331027 —

Morfolojik Analiz Nedir?

- Morfoloji, kelimelerin yapısını ve kelimelerin kökleri, ekleri, zaman, kişi gibi özelliklerini inceleyen dilbilim dalıdır.
- Doğal dil işlemede, morfolojik analiz kelimeleri köklerine ayırmak ve gramer bilgilerini anlamak için kullanılır.

Örnek

Kelime: kitaplarımızdan

Kök: kitap

Ekler:

- "-lar" (çoğul eki)
- "-ımız" (iyelik eki, birinci çoğul şahıs)
- "-dan" (ayrılma durumu eki)

Morfolojik Analizin Doğal Dil İşlemedeki Önemi

Dilin Kurallı Yapısını Çözerek Kelimelerin Anlamını Çıkartmak

Her dilin, kelimelerin nasıl oluşturulacağını belirleyen bir dizi gramer kuralı vardır. Morfolojik analiz, bir kelimenin kökünü ve eklerini ayırarak bu kuralları çözümleyip kelimenin anlamını ortaya çıkarmaya yardımcı olur.

Örneğin, Türkçede "geliyorum" kelimesi:

- **Kök:** gel (fil)
- **Ek:** -iyor (şimdiki zaman eki)
- **Ek:** -um (birinci tekil şahıs eki)

Sözcüklerin Farklı Biçimlerini Analiz Ederek Dilsel Zenginliği Anlamak

Bir kelimenin farklı eklerle türetilebilmesi, o dilin zenginliğini gösterir. Morfolojik analiz, bir kelimenin hangi biçimlerde kullanılabileceğini ortaya koyar.

Örneğin, "kitap" kelimesi şu biçimlere gelebilir:

- kitaplar (çoğul hali)
- kitapçı (meslek adı)
- kitapsız (olumsuzlama)

Diğer Doğal Dil İşleme Adımları İçin Veri Hazırlığı

Morfolojik analiz, doğal dil işleme sürecindeki birçok adım için ön hazırlık sağlar. Bu adımların bazıları şunlardır:

Anlamsal Analiz: Morfolojik analiz, kelimelerin kök anlamını çıkartarak cümle içinde kelimeler arası anlam ilişkilerinin doğru bir şekilde kurulmasını sağlar. Örneğin, bir kelimenin hangi kipte veya hangi şahsa ait olduğunu bilmek, anlam çözümlemesi açısından önemlidir.

Makine Çevirisi: Çeviri sistemleri, bir kelimenin kökünü ve eklerini doğru analiz ederek hedef dile uygun biçimde çevirmelidir. Örneğin, Türkçeden İngilizceye çeviri yapılırken "kitaplarımızdan" kelimesini doğru şekilde "from our books" olarak çevirebilmek için morfolojik analiz gereklidir.

Metin Anlamlandırma (NLP Pipelines): Doğal dil işleme süreçlerinde (ör. duygu analizi, özetleme, soru cevaplama), kelimelerin doğru bir biçimde analiz edilmesi gereklidir. Morfolojik analiz, metinlerdeki kelimelerin hangi zaman diliminde kullanıldığını, kimlerle ilgili olduğunu ya da hangi olumsuzlama/olumlu anlam taşıdığını anlamaya yardımcı olur.

Adlandırılmış Varlık Tanıma (Named Entity Recognition, NER): Morfolojik analiz sayesinde isim, yer, tarih gibi özel varlıkların tanınması kolaylaşır. Bu varlıkların anlamı ve cümledeki bağlamını doğru şekilde belirlemek için morfolojik bilgi gereklidir.

Morfolojik Analizin Zorlukları

- **Çok Şekillilik (Polimorfizm):** Kelimenin farklı bağlamlarda farklı anlamlara gelebilmesi.
- **Karmaşık Ek Yapıları:** Özellikle Türkçe gibi eklemeli dillerde eklerin çokluğu ve dilbilgisel rolüne göre çeşitliliği.
- **Sözlük Tabanlı ve Kural Tabanlı Yaklaşımlar:** Her iki yaklaşımın avantajları ve dezavantajları.

Morfolojik Analiz Yöntemleri

Kural Tabanlı Yaklaşımlar:

- Gramer kurallarına dayanarak kelimeleri analiz etme.
- Özellikle dil bilgisi açısından zengin dillerde (Türkçe, Fince) kullanılır.

İstatistiksel ve Makine Öğrenmesi Yöntemleri:

- Veriye dayalı öğrenme yöntemleri, etiketlenmiş veri üzerinden öğrenir.
- Derin öğrenme yöntemlerinin (LSTM, Transformer) bu alandaki kullanımı.

Morfolojik Analiz için Örnek Uygulamalar

Makine Çevirisi: Metinleri doğru şekilde çevirmek için morfolojik analize ihtiyaç duyulması.

Dil Anlamlandırma: Kullanıcıların girdiği kelimelere örnek cümle oluşturma.

Sentiment Analizi: Kelimelerin duygu taşıyan anlamlarını belirlemek.

Makale Özeti

<https://www.sciencedirect.com/science/article/pii/S1319157817301611>

Amaç: Arapça morfolojik analizinin hızını ve doğruluğunu artırmak için FPGAs kullanılarak paralel donanım işlemcileri geliştirilmiştir.

Zorluklar: Arapça'nın karmaşık eklemeleri ve iç eklerle (infix) çalışmak, analizi daha zor hale getirmektedir.

Yöntem: Hem yazılım hem de donanım versiyonları oluşturulmuş, paralel yapılar kullanılarak daha hızlı işlem süreleri elde edilmiştir.

Sonuçlar: Donanım, yazılıma göre 28.873 kat hızlanmış, doğruluk oranı %90'a ulaşmıştır.

Uygulama

Kod

```
def remove_affixes_turkish(word):
    # Türkçede yaygın ön ekler
    prefixes = ['en', 'en çok', 'en az']

    # Türkçede yaygın son ekler
    suffixes = ['ler', 'lar', 'ci', 'cı', 'lik', 'lık', 'da', 'de', 'ta', 'te', 'den', 'dan', 'e', 'a', 'im', 'ım', 'sin', 'sın', 'siniz', 'sunuz', 'mız', 'miz', 'niz', 'nız', 'nız', 'imiz', 'ımız', 'iniz', 'ınız', 'yız', 'yız', 'lerim', 'larım', 'ım', 'im', 'um', 'üm', 'm', 'n', 'u', 'ü', 'yim', 'yım']

    for prefix in prefixes:
        if word.startswith(prefix):
            word = word[len(prefix):].strip()
            break # ilk bulunan prefix çıkarılır

    for suffix in suffixes:
        if word.endswith(suffix):
            word = word[:-len(suffix)]
            break # ilk bulunan suffix çıkarılır

    return word

def morphological_analysis(word):
    root = remove_affixes_turkish(word)

    if len(root) < 3:
        return word
    return root

words = ['en küçük', 'kitaplar', 'ev', 'arabanız', 'okulda', 'çocukta', 'kedi', 'kalemim']

for word in words:
    root = morphological_analysis(word)
    print(f"Asıl kelime: {word} -> Kök: {root}")
```

Çıktı

Asıl kelime: en küçük -> Kök: küçük

Asıl kelime: kitaplar -> Kök: kitap

Asıl kelime: ev -> Kök: ev

Asıl kelime: arabanız -> Kök: araba

Asıl kelime: okulda -> Kök: okul

Asıl kelime: çocuktan -> Kök: çocuk

Asıl kelime: kedi -> Kök: kedi

Asıl kelime: kalemim -> Kök: kalem

Türkiyeden Örnek Uygulama - Zemberek

- <https://github.com/ahmetaa/zemberek-nlp>

Kaynakça

GeeksforGeeks. (2021). *Morphological Analysis in NLP*.

<https://www.geeksforgeeks.org/morphological-analysis-in-nlp/>

Ömer Akgül. (2011). Türkçe Kelimelerin Morfolojik Analizi

<https://akgulomer.wordpress.com/2011/01/23/turkce-kelimelerin-morfolojik-analizi/>

Dergipark. (2024). Türkçe doğal dil işleme: ses bilgisi ve morfolojik analiz.

<https://dergipark.org.tr/tr/pub/jiens/issue/82875/1472513>