

المعالجة الآلية للغة العربية (التحليل الصرفي)

عبد الإله السفياني

بِسْمِ اللّٰهِ الرَّحْمٰنِ الرَّحِیْمِ

الحوسبة العربية (تعريب الحاسبات)

المعالجة الآلية للغة العربية

(التحليل الصرفي)

عبد الإله السفياني

المسار

- مقدمة
- ما هو المحلل المصرفي؟
- تعاريف
- متطلبات عمل المحلل
- خطوات عمل المحلل
- الخوارزمية المستخدمة
- الخلاصة

مقدمة

• مراحل المعالجة الآلية للغة:

١. التحليل الصرفي:

تحليل كل كلمة إلى أجزائها.

٢. تحليل التركيب:

تحويل الجملة (سلاسل الكلمات) إلى تراكيب معينة تبين ترابط الكلمات ببعضها.

٣. تحليل الدلالة (المعنى):

تحويل المدخلات إلى تمثيل داخلي يحتفظ بالمعنى.

ما هو المحلل الصرفي؟

- هو برنامج يقوم بدراسة بناء الكلمة، وتحويلها إلى جذرها الأصلي.

- من تطبيقاته:

١. المدقق الإملائي: تحليلي.

٢. البحث: تحليل + توليد.

٣. مترجم المفردات.

تعريف

- المشتقات و الجوامد:

- المشتق: و هو المولد أو المتفرع من جذر، مثل: مكتبة مأخوذه من الجذر (ك ت ب).
- الجامد: و هي التي ليس لها جذر، مثل: بئس، نعم، إنسان ...

- **الجذر:** و طولة من ٣-٤ أحرف و الجذر هو الحروف الأصلية في الكلمة، و الجذر ليس بكلمة و إنما مجموعة من الحروف.

- **الوزن:** الوزن = الجذر + حروف أو حركات ضبط بالشكل مزيدة

- **السوابق:** الحروف الزائدة في بداية الكلمة.

- **الواحق:** الحروف الزائدة في نهاية الكلمة.

متطلبات عمل المحلل

- ١ . قائمة بجميع السوابق الممكنة. (Prefix)
ال التعريف، تاء الفاعل، ... إلخ
- ٢ . قائمة بجميع اللواحق الممكنة. (Suffix)
واو الجماعة، نون النسوة، ... إلخ
- ٣ . قائمة بجميع الأوزان الأساسية. (Patterns)
فعل، فعول، مفاعيل، ... إلخ
- ٤ . قاعدة بيانات بجميع الجذور الثلاثية منها و الرباعية.
- ٥ . قاعدة بيانات بجميع الكلمات الجامدة.

خطوات عمل المحلل..

١. ابحث عن الكلمة في قائمة الكلمات الجامدة

إذا وجدت: انتهى؛

إذا لم توجد: اذهب إلى الخطوة الثانية؛

٢. باستخدام قائمة السوابق واللوح احذف جميع السوابق واللوح المتصلة بالكلمة؛

٣. اختر الوزن المناسب للكلمة

٤. استخرج الجذر من الكلمة

خطوات عمل المحلل (مثال) ..

- رابعاً: نستخرج الجذر من الكلمة:
- باختيار الأحرف المقابلة ل- ف ع ل -

المدخلات	ا	ل	م	ر	ا	ب	ظ	و	ن
الوزن			م	ف	ا	ع	ن		
الجذر				ر		ب	ظ		

خطوات عمل المحلل (مثال)

- أخيراً: نبحث عن الجذر المستخرج (ربط) في قاعدة بيانات الجذور .

إذا موجود → إذا الكلمة صحيحة

إذا غير موجود → إذا هنالك خطأ إملائي

الخوارزمية

1. check if the word is invariable:

1.1 using the prefix table remove the largest prefix associated with the word.

1.2 if the remaining part of the word is in the list of invariable words then:

1.2.1 if the prefix is compatible with the word and the variable-invariable flag is false then return the type; value; grammatical rule

1.2.2 else go to step 2 below.

الخوارزمية

2. Analyze the word assuming that is variable.

2.1 Using the suffix table remove the largest suffix associated with the word.

2.2 Using the morphology pattern list, locate patterns that match the word.

A pattern is said to match a word if it satisfies the following conditions:

- * the pattern is equal to the word in length.

- * the root identified by the pattern is in the root table.

- * the removed suffix and prefix is compatible with the selected pattern.

- * the root is compatible with the pattern.

2.3 For each matching pattern return the type of the corresponding word, the root, the suffix and prefix.

2.4 Add the removed suffix back to the word letter by letter, starting with the first letter in the suffix. For each letter added repeat steps 2.2 and 2.3.

2.5 Add the removed prefix back to the word letter by letter, starting with the first letter in the prefix. For each letter added repeat steps 1 and 2.

الخلاصة

- مراحل المعالجة الآلية للغة
- ما هو المحلل الصرفي؟
- متطلبات عمل المحلل
- خطوات عمل المحلل
- الخوارزمية المستخدمة

شكراً

أي أسئلة...