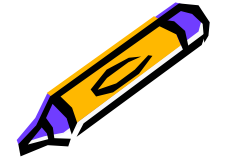


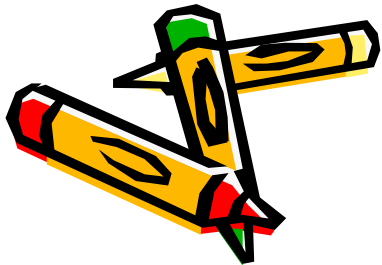
التطبيقات الحاسوبية للغة العربية
في المعهد العالي للعلوم التطبيقية
والتكنولوجيا

د. أميمة الدكاك، د. ندى غنيم



المحتوى

- مقدمة
- تحليل وتركيب الإشارة الكلامية
- تعرف الإشارة الكلامية
- اللغة العربية في المعهد العالي للعلوم التطبيقية والتكنولوجيا
- خاتمة

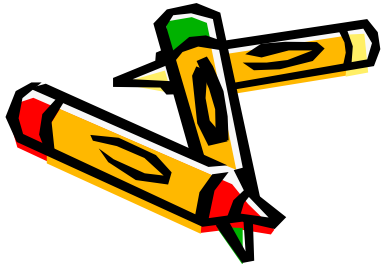


التواصل الكلامي مع الآلة

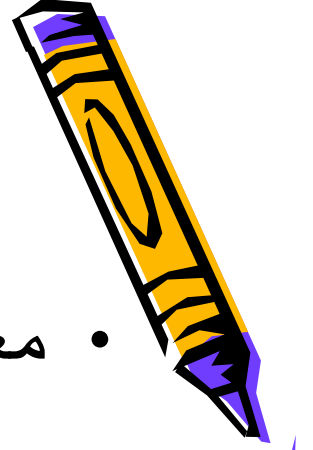


- الحاجة إلى استخدام الحاسوب في الكثير من التطبيقات الحيوية
- إنشاء الجيل الخامس من الحواسيب الذي يتواصل كلامياً مع المستخدم
- اللغة العربية:

- يتحدثها ما لا يقل عن 330 مليون عربي
- يهتم بتعلمها قرابة ١.٢ مليار من المسلمين
- تحظى باهتمام متزايد حتى من الأوساط غير المسلمة وغير الناطقة بالعربية



معالجة اللغات



• معالجة الإشارة الكلامية

- تركيب الكلام

- تعرف الكلام

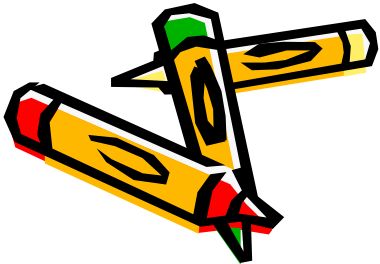
• المعالجة النصية

- المستوى الصرفي

- المستوى النحوي

- المستوى الدلالي

- المستوى المقامي أو الاستعرافي



تركيب الكلام

• تختلف مركبات الكلام هذه بحسب:

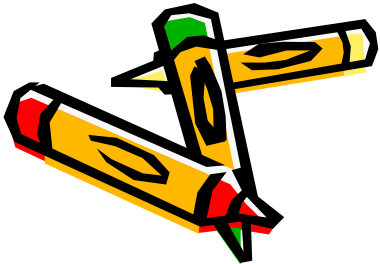
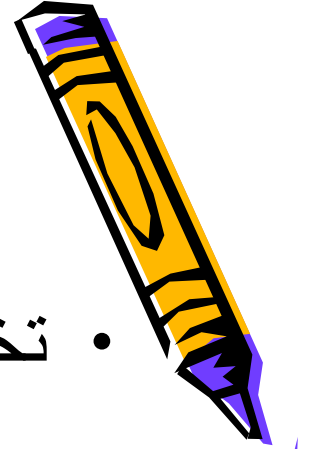
- الوحدات الصوتية الأساسية

- تقنيات التركيب :

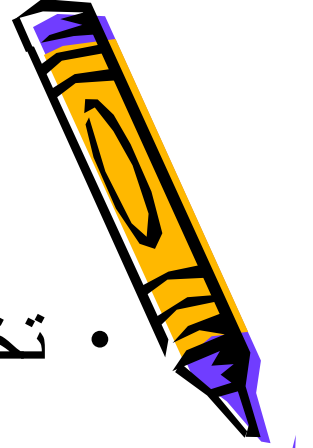
• تحليل الإشارة في المجال الزمني

• تحليل الإشارة في المجال الترددي

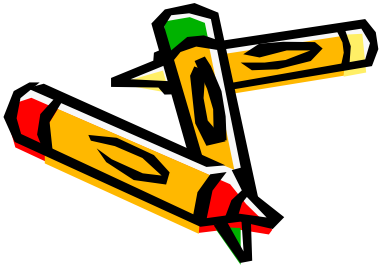
• تحويلات مغايرة.



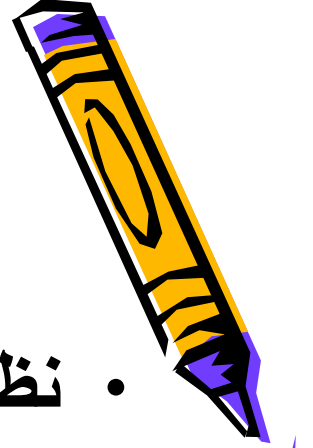
تعرف الكلام



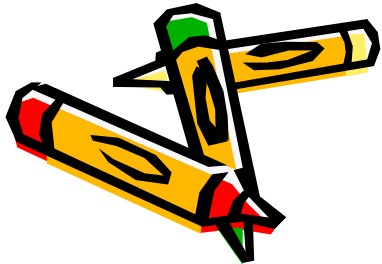
- تختلف نظم التعرف بحسب :
 - الوحدة الصوتية التي يجري تعرفها: كلمات معزولة أم كلمات متصلة أم كلام طبيعي مستمر
 - حجم الوحدات الصوتية: عدد الكلمات المتعرفة.
 - تعرف بصوت متحدث وحيد أم عدة متحدثين أم بدون شروط على المتحدث.
 - موثوقية نظام التعرف بحسب تغيير ظروف استثماره.



اللغة العربية في المعهد العالي للعلوم التطبيقية والتكنولوجيا

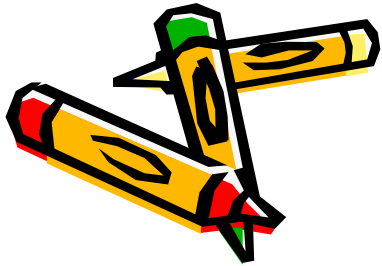
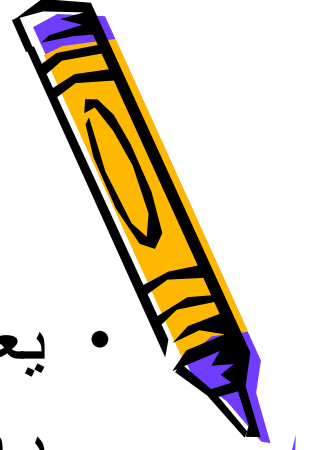


- نظام اشتقاق وتصريف [البواب ٨٤]
- قاعدة معطيات قواعد الإعراب [باكير ٠٢]
- قاموس إلكتروني "قاعدة معطيات المعجم الوسيط"
- محلل صرفي للغة العربية (سنبل 08)
- نواة نظام تشكيل النصوص
- مركب كلام باستخدام الثنائيات الصوتية
- نواة لمركب باستخدام أنصاف المقاطع
- نظام تعرف كلمات معزولة
- تعرف الخط العربي



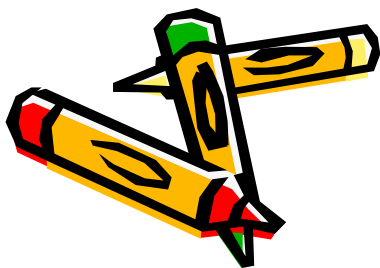
نظام اشتقاق وتصريف

- يعتمد معجماً يتضمن ٥٥٨٨ جذراً ثلاثياً و ١٩٣٢ جذراً رباعياً
- يحوي المعجم على ١١٧٩٠ مصدراً سماعياً و ٨٦٠ صفة مشبهة بالفعل خاصة بالأفعال الثلاثية المجردة.
- تولد الأسماء المشتقة القياسية وكذلك المصادر القياسية الخاصة بالأفعال الثلاثية المزيدة والأفعال الرباعية وفق قواعد محددة دون الرجوع إلى المعجم.



قاعدة معطيات قواعد الإعراب

- قاعدة معطيات تتضمن إعراب ١٢٠٠ شاهد من الأدب الرفيع إعراباً مفصلاً
- تناسب مختلف المراحل الدراسية
- تغطي جميع الدروس النحوية.
- شُرحت الشواهد مفردات فقط
- يجري العمل على تسويقه

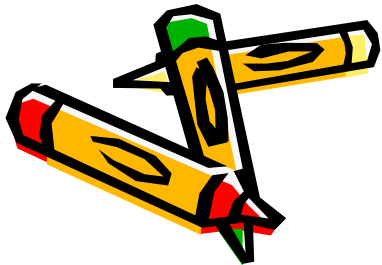


قاموس إلكتروني "قاعدة معطيات المعجم الوسيط"

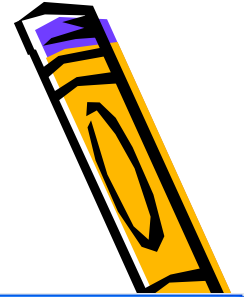


- إدخال مفردات المعجم في قاعدة معطيات
- تصنيفها في ١٦ جدولاً تتضمن المعلومات الصرفية والدلالية المتعلقة بها.
- نواة أساسية لأعمال تحليل صرفي ونحوي ودلالي لاحقة
- مثال يتضمن جدول الأفعال في القاعدة عشرة حقول هي:

- رقم مفتاح الفعل
- شيوعه
- جذره
- الفعل نفسه
- وزنه
- تعديته
- اسم الفاعل منه
- حرف التعدية
- اسم المفعول منه
- معناه.



واجهة المستثمر



المعجم الوسيط

نتائج البحث: أفعال

عدد المداخل	وزن الفعل	الجذر
٤٩		
٣		
١		
٩		

أدخل كلمة
ضرب

أسماء الأفعال

المعجم الوسيط

أدخل كلمة

طريقة البحث

حسب الكلمة والجذر
حسب الجذر
حسب الجذر والمشابهات
حسب الكلمة والجذر

مساعدة

حول البرنامج

إغلاق

المعجم الوسيط

نتائج البحث: أسماء

عدد المداخل	الجذر	الكلمة
٢	ضرب	أضْرَاب
١	ضرب	ضْرَب
١	ضرب	ضَرِب
٧	ضرب	ضَرْب

أدخل كلمة
ضرب

طريقة البحث

حسب الكلمة والجذر

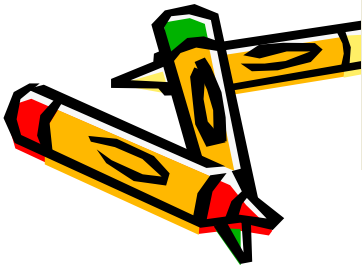
بحث

مساعدة

حول البرنامج

إغلاق

أسماء الأفعال أسماء الأصوات الأدوات **الأسماء** الأفعال



واجهه المستثمر



واجهة مستخدم لبرنامج تحليل لغوي، تعرض معلومات مفصلة عن الفعل "ضرب".

نافذة التفاصيل:

- الفعل: ضَرَبَ
- الجذر: ضرب
- المعنى: ضَرَبَ لَهُ مَوْعِداً: حَدَّدَهُ وَعَيَّنَهُ.

نافذة التراكيب:

- جذر: ضرب
- تركيب: يضرب أخماساً لأسداس
- معنى التركيب: نو يضرب أخماساً لأسداس: يسعى في المكر والخديعة.

أزرار التنقل: < << >> >

زر العودة: عودة

الصفحات المشبهة و جمعها

مبالغة اسم الفاعل و جمعها

فعليل بمعنى فاعل

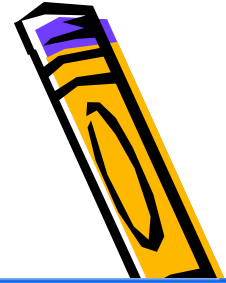
فعليل بمعنى مفعول

المصدر	الفعل
تَضْرَاب	ضَرَبَ

عود



واجهه المستثمر



تفاصيل

الكلمة: ضَرْبٌ

الجذر: ضرب

المعنى: الصَّنْفُ والنُّوعُ.

صيغة الجمع

الكلمة	جمع الكلمة	وزن الجمع	شيوخ الجمع
ضَرْبٌ	أَضْرَابٌ	أَفْعَالٌ	غير شائع
ضَرْبٌ	أَضْرُبٌ	أَفْعُلٌ	غير شائع

أمثلة الأسماء

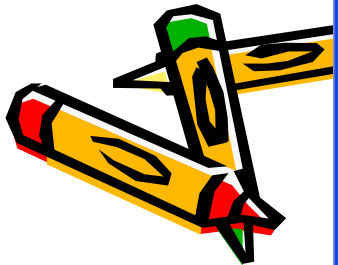
الفعل: ضَرْبٌ

المثال: يُقَالُ: هذا من ضَرْبِ ذلك، أي: من نَحْوِهِ وصِنْفِهِ. يُقَالُ:

التراكيب

عودة

عودة



استراتيجيات البحث

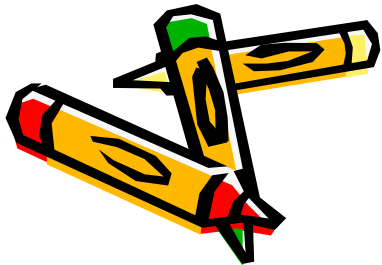


حسب الجذر

حذف الحركات

وضع الهمزة على السطر

البحث في جداول الأسماء والأفعال عن الكلمات التي جذرها
الكلمة المعطاة



استراتيجيات البحث



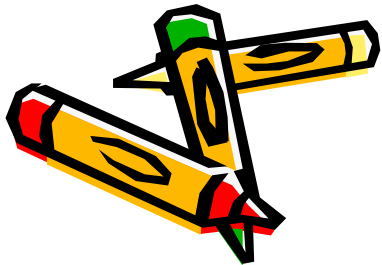
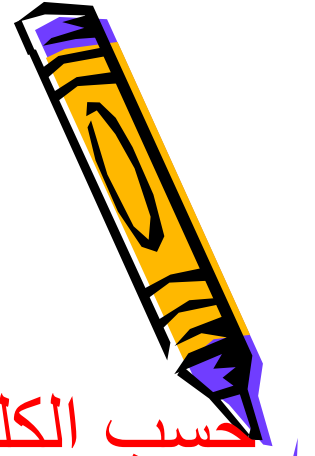
حسب الكلمة والمتشابهات

"like" instruction of SQL language,
giving the words "%x%y%z%", where
"%" means "anything."



استراتيجيات البحث

حسب الكلمة والجذر





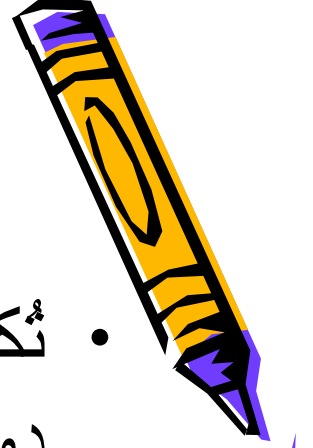
مادة (ضرب) في جدول الأفعال:

ملحق الجدول

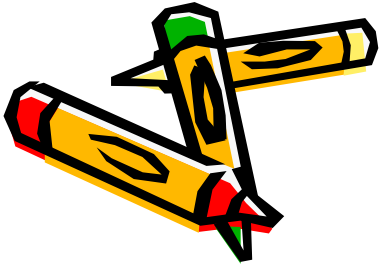
المعنى	المفعول به	حرف المدية	الفاعل	المدية أو التورم	وزن الفعل	الفعل	الجزء	شروع الفعل	رقم صفح الفعل
ضَرَبَ الشَّيْءُ: تحرك.			شيء	لازا	٢	ضَرَبَ	ضرب	ا	٢٦١٢٠
ضَرَبَ القَلْبَ: تَضَنَ.			قلب	لازا	٢	ضَرَبَ	ضرب	ا	٢٦١٢١
ضَرَبَ العِرْقَ: هاج دَمُهُ واحتلج.			عرق	لازا	٢	ضَرَبَ	ضرب	ا	٢٦١٢٢
ضَرَبَ الصَّرْسَ: أو نحوهُ: اغْتَدَّ وجعهُ وآلمهُ.			ضرس	لازا	٢	ضَرَبَ	ضرب	ا	٢٦١٢٣
ضَرَبَ الرَّجُلَ في الأَرْضِ: دَهَبَ وأبْعَدَ. وسَارَ في التَّغَاءِ الرَّجُلِ.	أرض	في	فلان	معد بحرف	٢	ضَرَبَ	ضرب	ا	٢٦١٢٤
ضَرَبَ الرَّجُلَ في الأَرْضِ: تَهَضَّ وأَسْرَعَ في السَّيْرِ.	أرض	في	فلان	معد بحرف	٢	ضَرَبَ	ضرب	ا	٢٦١٢٥
ضَرَبَ الرَّجُلَ في المَاءِ: سَجَّ.	ماء	في	فلان	معد بحرف	٢	ضَرَبَ	ضرب	ا	٢٦١٢٦
ضَرَبَ في الأمرِ بِسَهْمٍ ونحوهُ: شارك فيه.	أمر	في	فلان	معد إلى متعولين	٢	ضَرَبَ	ضرب	ا	٢٦١٢٧
ضَرَبَ عَنِ الأمرِ: كَفَّ وأَعْرَضَ.	أمر	عن	فلان	معد بحرف	٢	ضَرَبَ	ضرب	ا	٢٦١٢٨
ضَرَبَ اللُّونَ إلى اللونِ: مَالَ.	لون	إلى	لون	معد بحرف	٢	ضَرَبَ	ضرب	ا	٢٦١٢٩
ضَرَبَ يَدَهُ إلى كذا: أهُوَى.	يده	ب	فلان	معد إلى متعولين	٢	ضَرَبَ	ضرب	ا	٢٦١٣٠
ضَرَبَ إليه: أَفْخَرَ.	فلان	إلى	فلان	معد بحرف	٢	ضَرَبَ	ضرب	ا	٢٦١٣١
ضَرَبَ على المكتوبِ وغيره: حَتَمَ.	مكتوب	على	فلان	معد بحرف	٢	ضَرَبَ	ضرب	ا	٢٦١٣٢
ضَرَبَ النَّوْمَ على أذنه: عَلَّه.	أذنه	على	نوم	معد بحرف	٢	ضَرَبَ	ضرب	ا	٢٦١٣٣
ضَرَبَ فلانٌ على يدِ فلانٍ: أَمْسَكَ وقَبَضَ.	يد فلان	على	فلان	معد بحرف	٢	ضَرَبَ	ضرب	ا	٢٦١٣٤
ضَرَبَ فلانٌ على فلانٍ: أَفْسَدَ عليه أمرُهُ.	فلان	على	فلان	معد بحرف	٢	ضَرَبَ	ضرب	ا	٢٦١٣٥
ضَرَبَ القاضِي على يدِ فلانٍ: حَجَرَ عليه ومنعه التَّصَرُّفَ.	يد فلان	على	القاضي	معد بحرف	٢	ضَرَبَ	ضرب	ا	٢٦١٣٦



نظم تشكيل النصوص



- تُكتب النصوص باللغة العربية، في غالب الأحيان، من دون تشكيل.
- يستطيع القارئ المتمرس استعادة الحركات بناء على خبرته
- تحمل هذه الحركات معلومات لغوية هامة منها دلالية ومنها نحوية
- لإجراء أية عملية تحليل آلي على نص عربي لا بد أن يكون مشكولاً.



تشكيل النصوص العربية

بناء نواة مشكل آلي للنصوص العربية، يضيف ما
أمكن من الحركات والشدات إلى نص غير مشكول.

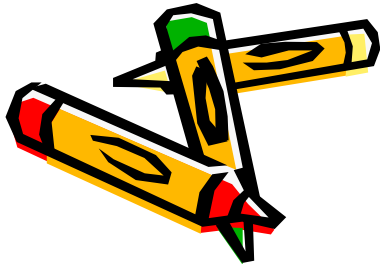
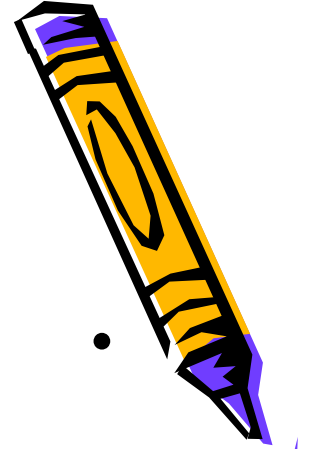
المراحل:

- التحليل الصرفي للكلمات Morphological Analysis

- تحديد أنواع الكلمات Part of Speech Tagging

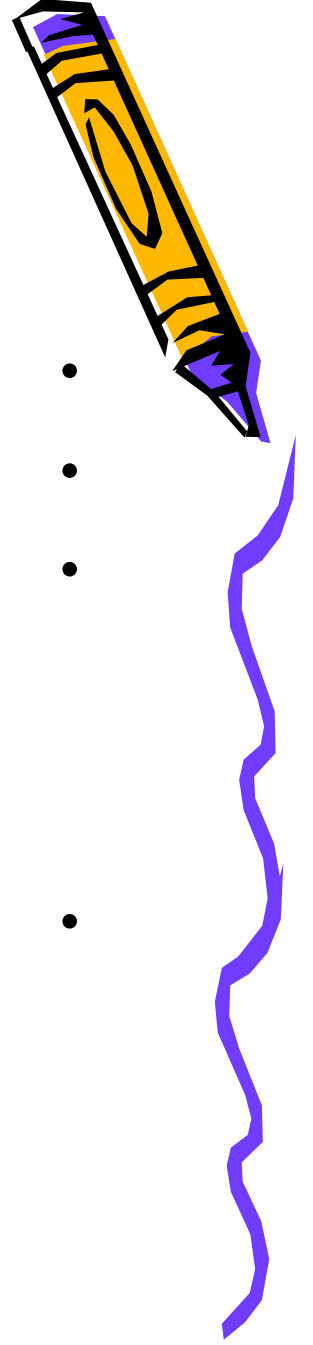
- تطبيق القواعد اللغوية التجريبية Heuristic

Linguistic Rules



التحليل الصرفي للكلمات

Morphological Analysis

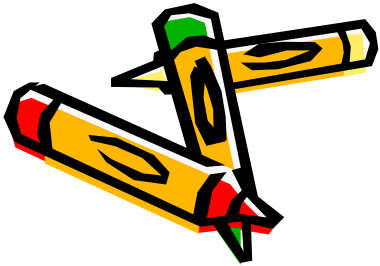


اعتماد محلل Buckwalter الصرفي
طريقة معاجم السوابق والجذوع واللواحق ولوائح التوافق.

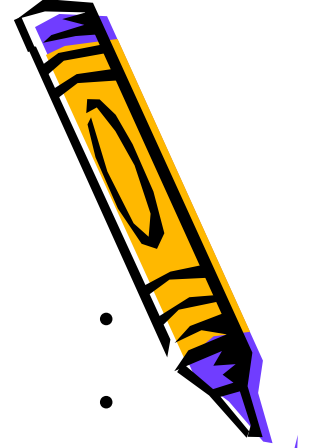
ثلاثة معاجم:

- معجم الجذوع اللغوية ويحوي الجذوع مع التحليلات الممكنة
 - معجم السوابق مع التحليلات الممكنة لكل سابقة
 - معجم اللواحق مع التحليلات الممكنة لكل لاحقة.
- لائحتان:

- لائحة توافق السوابق مع الجذوع
- لائحة توافق الجذوع مع اللواحق.



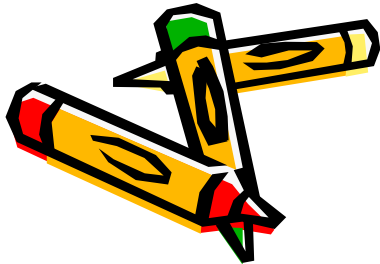
تحديد أنواع الكلمات Part of Speech Tagging



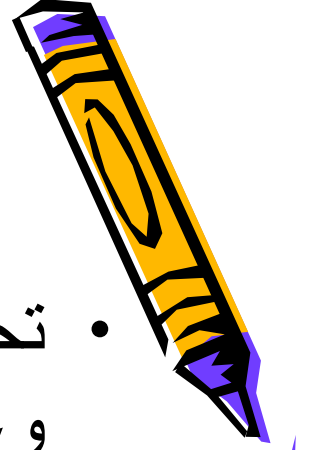
لكل كلمة عدة خيارات، كل خيار مرتبط بتشكيل معين لهذه الكلمة
تحديد نوع الكلمة POS من خلال بناء Tagger للغة العربية.
اعتماد طرق إحصائية تعرف بالتعلم غير المُشرف عليه
Unsupervised learning (لافتقارنا إلى مدونة مشكولة باللغة
العربية).

تعتمد هذه الطريقة على:

- إجراء التحليل الصرفي على كمٍ واسع من النصوص غير المشكولة واستنتاج قواعد إحصائية لتحديد أنواع الكلمات (من الحالات التي لا لبس فيها)
- ترتيب هذه القواعد بحسب علامات تأخذ بالاعتبار تكرار تطبيقها على النصوص.
- لدى تعليم نصوص جديدة، نطبق القواعد التي لها العلامات الأفضل، ثم نطبق قواعد تحويل لتحسين الأداء (إزالة الغموض أكثر) اعتماداً على الكلمة السابقة ونوعها والكلمة اللاحقة ونوعها.



تطبيق



- تحليل ٣ نصوص: عن الحاسب، وعن كوكب الأرض، وعن مدينة دمشق، يبلغ مجموعها قرابة ١٨٠٠٠ كلمة.
- نتائج التحليل:

- نسبة الكلمات غير الغامضة ٤٠% من النص.

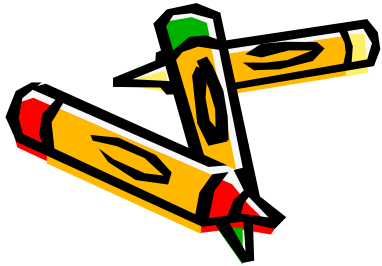
- ٣% من الكلمات لم يتعرف عليها المحلل.

- ٢٣% من الكلمات كان لها علامتان **Two tags**

- ١٤% من الكلمات كان لها ثلاثة علامات.

- ٧% من الكلمات كان لها أربع علامات.

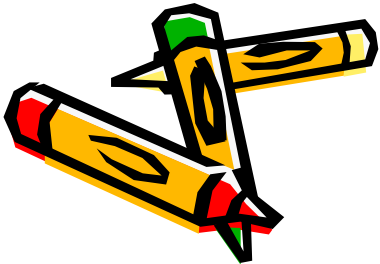
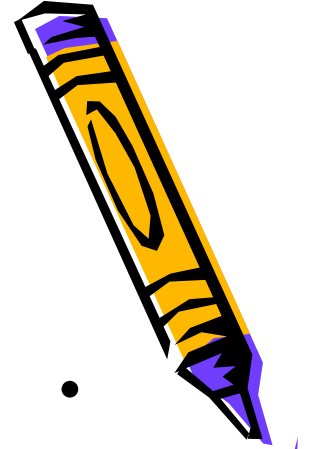
- ...



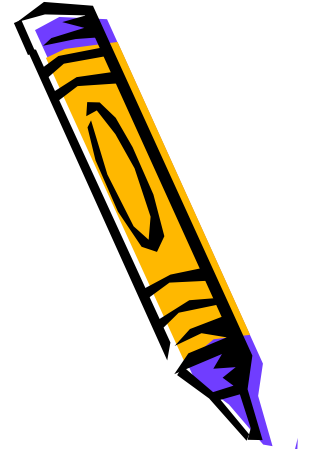
تطبيق القواعد اللغوية التجريبية

Heuristic Linguistic Rules

- القواعد المستنتجة آلياً تغفل أحياناً بعض الحالات السهلة المعالجة.
- إن القواعد محدودة بالقوالب المحددة، وبالتالي لا تغطي جميع الاحتمالات الممكنة.
- قمنا بإضافة قواعد لغوية تجريبية وبتجربة الأداء بعد إضافة كل قاعدة، بهدف إبقاء القواعد الجيدة وحذف السيئة منها.



عرض



المشكل الآلي

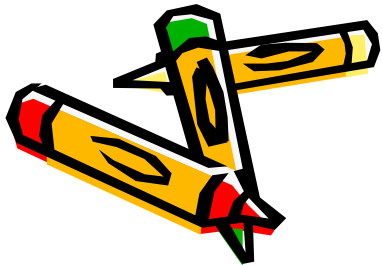
تحليل لغوي تشكييل جزئي تشكييل كامل

أنواع المقالة

إن ازدهار المقالة وانتشارها, واحتلالها مكان الصدارة بين سائر ألوان الكتابة, جعل كتابها يتناولون فيها موضوعات شتى يصعب حصرها, فما من جانب من جوانب الثقافة الإنسانية إلا تصدت له المقالة, وما من ميدان من ميادين المعرفة إلا خاضته وعالجت شأنًا

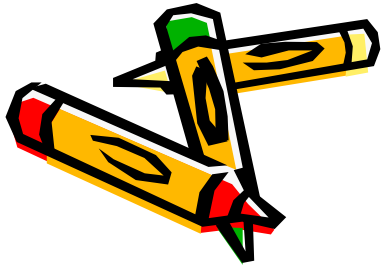
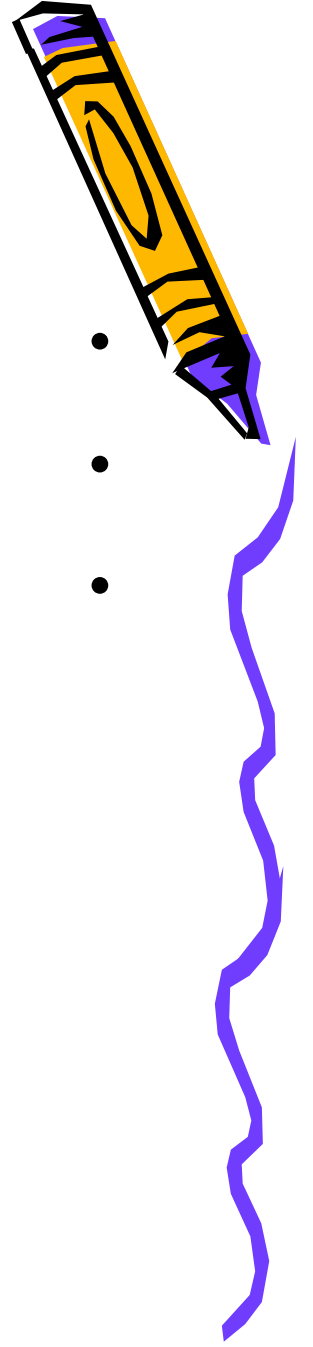
أنواع المقالة

إن ازدهار المقالة وانتشارها, واحتلالها مكان الصدارة بين سائر ألوان الكتابة, جعل كتابها يتناولون فيها موضوعات شتى يصعب حصرها, فما من جانب من جوانب الثقافة الإنسانية إلا تصدت له المقالة, وما من ميدان من ميادين المعرفة إلا خاضته وعالجت شأنًا من شؤونه, كذلك صارت مطية لعرض حاجات النفس

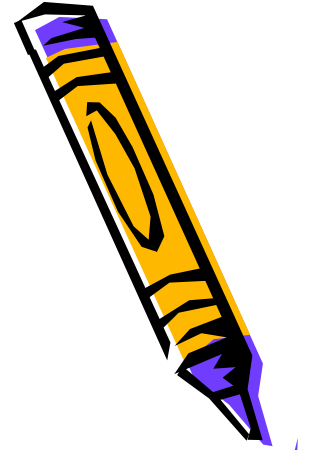


مركب كلام باستخدام الثنائيات الصوتية

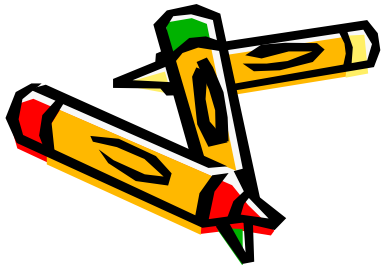
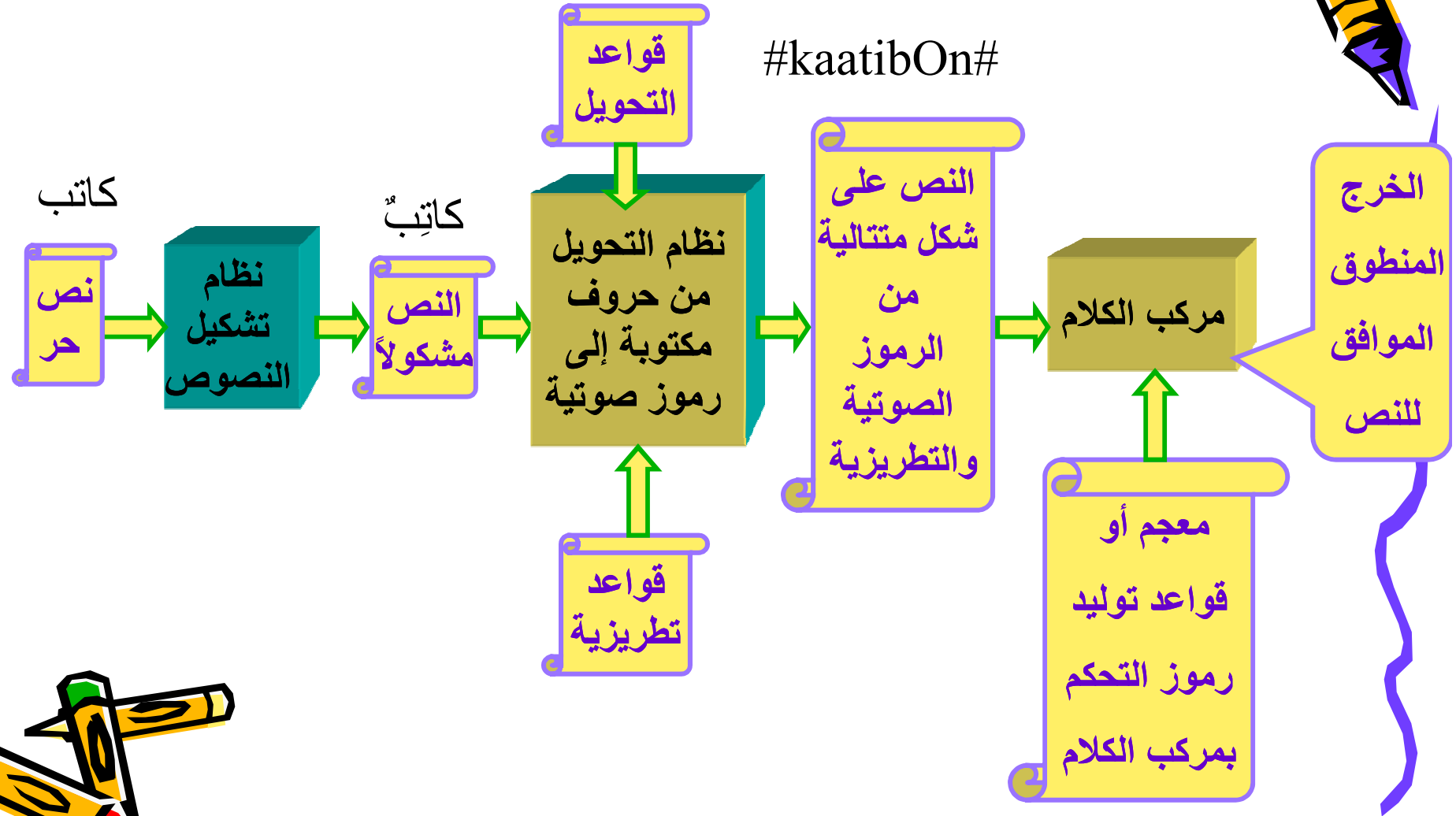
- كتلة التحويل من رموز كتابية إلى رموز صوتية.
- كتلة التحويل من رموز صوتية إلى أصوات منطوقة.
- كتلة تعديل أنماط النطق التطريزية *Prosody* لإعطاء صوت طبيعي ثم لإعطاء المشاعر المطلوبة.



مخطط نظام تركيب الكلام من النص



#kaatibOn#



نواة لمركب باستخدام أنصاف المقاطع

اللغة العربية لغة مقاطعية

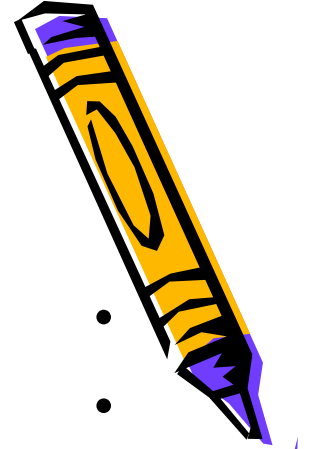
أنواع المقاطع الصوتية في اللغة العربية هي: V ، و CV و CVC و $CVCC$ (V هي الصوائت و C هي الصوامت) أنصاف المقاطع هي ما ينتج حين نقص المقطع الصوتي عند الصائت.

أنصاف المقاطع الناتجة: $CV\#$ و $VC\#$ و $VCC\#$ و VCV و $VCCV$ (يشير $\#$ إلى الصمت قبل الكلام أو بعده).

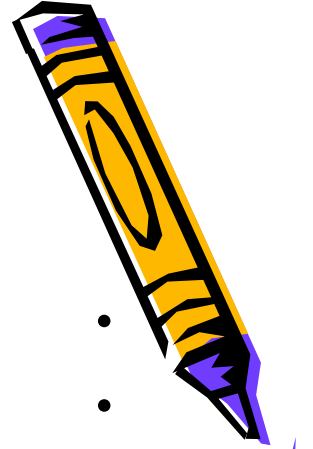
عدد الثنائيات الصوتية لا يتجاوز ١٥٠٠ (لـ ٣٨ صوتيم).

عدد أنصاف المقاطع، يتجاوز ١١٠٠٠ بعد استبعاد الصوامت المتنافرة.

انتهى تسجيل المدونة التي تتضمن أنصاف المقاطع بصوت أنثوي، وبقي تقطيعها وتحليلها.



نظام تعرف كلمات معزولة



التحكم الصوتي في ربوط (ذراع آلي)، موصول إلى حاسوب.
يتكون هذا الذراع من خمسة مفاصل: قاعدة وكتف ومرفق ومعصم و
قبضة.

كل من المفاصل الأربعة الأولى يمكن أن يتحرك نحو: اليمين أو اليسار
أو الأعلى أو الأسفل، أما القبضة فتكون حركتها بالفتح او الإغلاق.

يتكون الأمر من مجموعة كلمات هي على التالي:

- المفصل، جهة الحركة وكمية الحركة.

كمية الحركة هي عدد مقدر بالدرجات، يمثل زاوية الحركة.

يمكن لفظ كمية الحركة رقماً رقماً أو بالطريقة المعتادة (42 كلمة)

تقنية ماركوف المخفية HMM

نسبة التعرف 100% للشخص الذي درّب نظام التعرف، و 82%

للآخرين للتشكيلة المعتمدة (حجم النماذج الصوتية **book size**)

= 512، 15 متوسط سبستروم لكل نموذج مع مشتقيه الأول والثاني



نظام التعرف الضوئي على الخط العربي

• التعرف على الأحرف العربية :

- مختلف أشكالها إلا التزيينية منها

- مختلف الحجم

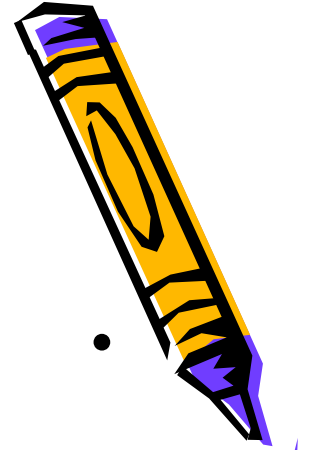
- يمكنه التغلب على ميلانات السطور

- يمكنه تقطيع الأحرف المتداخلة والفصل فيما بينها.

• دخل البرنامج ملف صورة من نوع **bmp** بلونين فقط.

والخرج ملف نصي يتضمن المحتوى النصي للكتابة

باللغة العربية



خاتمة

