

التشكيل الآلي للنص العربي

الحوسبة العربية (تعريب الحاسبات) – ٢٤

حسني المحتسب

+ محاضرات الطلاب ١

بسم الله الرحمن الرحيم

الحوسبة العربية (تعريب الحاسبات) – ٢٤  
التشكيل الآلي للنص العربي  
+ محاضرات الطلاب ١

حسني المحتسب

تَقْبِلُ اللَّهُ مِنَّا وَمِنْكُمْ

وَأَيُّهَا  
وَأَيُّهَا  
وَأَيُّهَا

# المسار

- التشكيل الآلي للنص العربي
- محاضرات الطلاب ١

# التشكيل الآلي للنص العربي

Form1

نص غير مشكّل

حدثنا يحيى بن يحيى التميمي،  
وأبو بكر بن أبي شيبة ومحمد بن  
العلاء الهمداني - واللفظ ليحيى  
- قال يحيى أخبرنا وقال  
الأخران، حدثنا أبو معاوية، عن  
الأعمش، عن أبي صالح، عن أبي

نص مشكّل

حدثنا يحيى بن يحيى التميمي،  
وأبو بكر بن أبي شيبة ومحمد بن  
العلاء الهمداني - واللفظ ليحيى  
- قال يحيى أخبرنا وقال  
الأخران، حدثنا أبو معاوية، عن  
الأعمش، عن أبي صالح، عن أبي

تتشكيل

إقرأ

خروج

ISDial.exe D:\DataContainer\Data\M8M\Tools\Aerobat505ME.zip

# التشكيل الآلي للنص العربي

- حدثنا يحيى بن يحيى التميمي، وأبو بكر بن أبي شيبة ومحمد بن العلاء الهمداني - واللفظ ليحيى - قال يحيى أخبرنا وقال الآخران، حدثنا أبو معاوية، عن الأعمش، عن أبي صالح، عن أبي هريرة، قال قال رسول الله صلى الله عليه وسلم: " من نفس عن مؤمن كربة من كرب الدنيا نفس الله عنه كربة من كرب يوم القيامة ومن يسر على معسر يسر الله عليه في الدنيا والآخرة ومن ستر مسلما ستره الله في الدنيا والآخرة والله في عون العبد ما كان العبد في عون أخيه ومن سلك طريقا يلتمس فيه علما سهل الله له به طريقا إلى الجنة وما اجتمع قوم في بيت من بيوت الله يتلون كتاب الله ويتدارسونه بينهم إلا نزلت عليهم السكينة وغشيتهم الرحمة وحفتهم الملائكة وذكرهم الله فيمن عنده ومن بطأ به عمله لم يسرع به نسبه"

# التشكيل الآلي للنص العربي

• حَدَّثَنَا يَحْيَى بْنُ يَحْيَى التَّمِيمِيُّ، وَأَبُو بَكْرِ بْنُ أَبِي شَيْبَةَ وَمُحَمَّدُ بْنُ الْعَلَاءِ  
الْهُمْدَانِيُّ - وَاللَّفْظُ لِيَحْيَى - قَالَ يَحْيَى أَخْبَرَنَا وَقَالَ الْآخِرَانِ، حَدَّثَنَا أَبُو  
مُعَاوِيَةَ، عَنِ الْأَعْمَشِ، عَنْ أَبِي صَالِحٍ، عَنْ أَبِي هُرَيْرَةَ، قَالَ قَالَ رَسُولُ اللَّهِ  
صَلَّى اللَّهُ عَلَيْهِ وَسَلَّمَ " مَنْ نَفَسَ عَنْ مُؤْمِنٍ كُرْبَةً مِنْ كُرْبِ الدُّنْيَا نَفَسَ اللَّهُ  
عَنْهُ كُرْبَةً مِنْ كُرْبِ يَوْمِ الْقِيَامَةِ وَمَنْ يَسَّرَ عَلَى مُعْسِرٍ يَسَّرَ اللَّهُ عَلَيْهِ فِي  
الدُّنْيَا وَالْآخِرَةِ وَمَنْ سَتَرَ مُسْلِمًا سَتَرَهُ اللَّهُ فِي الدُّنْيَا وَالْآخِرَةِ وَاللَّهُ فِي عَوْنِ  
الْعَبْدِ مَا كَانَ الْعَبْدُ فِي عَوْنِ أَخِيهِ وَمَنْ سَلَكَ طَرِيقًا يَلْتَمِسُ فِيهِ عِلْمًا سَهَّلَ اللَّهُ  
لَهُ بِهِ طَرِيقًا إِلَى الْجَنَّةِ وَمَا اجْتَمَعَ قَوْمٌ فِي بَيْتٍ مِنْ بُيُوتِ اللَّهِ يَتْلُونَ كِتَابَ اللَّهِ  
وَيَتَدَارَسُونَهُ بَيْنَهُمْ إِلَّا نَزَلَتْ عَلَيْهِمُ السَّكِينَةُ وَغَشِيَتْهُمُ الرَّحْمَةُ وَحَفَّتْهُمُ  
الْمَلَائِكَةُ وَذَكَرَهُمُ اللَّهُ فِيمَنْ عِنْدَهُ وَمَنْ بَطَأَ بِهِ عَمَلُهُ لَمْ يُسْرِعْ بِهِ نَسَبُهُ

## هل نحاول اقتراح الاحتمالات المختلفة للتشكيل؟

حدثنا	•	واللفظ	•	عن	•	"	•	ومن	•	والآخرة	•	سهل
يحيى	•	ليحيى	•	أبي	•	من	•	يسر	•	والله	•	الله
بن	•	-	•	صالح	•	نفس	•	على	•	في	•	له
يحيى	•	قال	•	،	•	عن	•	معسر	•	عون	•	به
التميمي	•	يحيى	•	عن	•	مؤمن	•	يسر	•	العبد	•	طريقا
،	•	أخبرنا	•	أبي	•	كربة	•	الله	•	ما	•	إلى
وأبو	•	وقال	•	هريرة	•	من	•	عليه	•	كان	•	الجنة
بكر	•	الآخران	•	،	•	كرب	•	في	•	العبد	•	وما
بن	•	،	•	قال	•	الدنيا	•	الدنيا	•	في	•	اجتمع
أبي	•	حدثنا	•	قال	•	نفس	•	والآخرة	•	عون	•	قوم
شبية	•	أبو	•	رسول	•	الله	•	ومن	•	أخيه	•	في
ومحمد	•	معاوية	•	الله	•	عنه	•	ستر	•	ومن	•	بيت
بن	•	،	•	صلى	•	كربة	•	مسلمًا	•	سلك	•	من
العلاء	•	عن	•	الله	•	من	•	ستره	•	طريقا	•	بيوت
الهمداني	•	الأعمش	•	عليه	•	كرب	•	الله	•	يلتمس	•	الله
-	•	،	•	وسلم	•	في	•	في	•	فيه	•	يتلون
٨	•	،	•	:	•	الدنيا	•	الدنيا	•	علما	•	كتاب



# هل نشكل هذه الكلمة؟

- درس
  - دَرَسَ
  - دَرَّسَ
  - دُرِسَ
  - دُرِّسَ
  - دَرَسٌ
- درس المعلم أحمد الدرس
  - درس اليوم ممتع
  - درس هذا الموضوع سابقا

# هل تشكل هذه الكلمات؟

- قال رسول الله صلى الله عليه وسلم
- علاقة الكلمات بما يجاورها

# علاقة الكلمات بما يجاورها

- قال رسول الله صلى الله عليه وسلم
- علاقة الكلمات بما يجاورها (احتمالية ظهور الكلمة بعد الكلمة السابقة لها)
- ثنائيات bi-gram
  - قال ← <بداية جملة>
  - رسول ← قال
  - الله ← رسول
  - صلى ← الله
  - الله ← صلى
  - عليه ← الله
  - وسلم ← عليه
- ثلاثيات tri-gram
- رباعيات quadra-gram
- خماسيات five-gram
- ....

# علاقة الكلمات بما يجاورها

- قال رسول الله صلى الله عليه وسلم
- احتمالية ظهور الكلمة بعد الكلمتين السابقتين لها

• ثلاثيات tri-gram

• رسول ← <بداية جملة> قال

• الله ← قال رسول

• صلى ← رسول الله

• الله ← الله صلى

• عليه ← صلى الله

• وسلم ← الله عليه

• رباعيات quadra-gram

• خماسيات five-gram

• .....

# علاقة الكلمات بما يجاورها

- قال رسول الله صلى الله عليه وسلم
- احتمالية ظهور الكلمة بعد الكلمات الثلاثة السابقة لها

• رباعيات quadra-gram

• الله ← <بداية جملة> قال رسول

• صلى ← قال رسول الله

• الله ← رسول الله صلى

• عليه ← الله صلى الله

• وسلم ← صلى الله عليه

• خماسيات five-gram

• .....

# علاقة الكلمات بما يجاورها

- قال رسول الله صلى الله عليه وسلم
- احتمالية ظهور الكلمة بعد الكلمات الأربعة السابقة لها
- خماسيات five-gram
- صلى ← <بداية جملة> قال رسول الله
- الله ← قال رسول الله صلى
- عليه ← رسول الله صلى الله
- وسلم ← الله صلى الله عليه
- .....

# كي نشكل كلمة

- هل نفترض أن الكلمة غير مشكلة؟
- هل نسمح بوجود كلمات مشكلة جزئياً؟
- هل نفترض أن الكلمات خالية من الأخطاء؟
- إذا كان التشكيل الجزئي مسموحاً، فهل نفترض الخلو من الأخطاء؟
- هل نعتبر الحركة على نهاية الكلمة؟
- ما تأثير ذلك على عملية التشكيل؟
- من يحدد الإجابة على هذه الأسئلة؟

# احصائيات

- النص المستخدم (مشكل)
  - ٦٦٣٣ جملة
  - ٤٠٥١١ كلمة
  - ٣٠٦٥٣٤ رمز عدا الفراغات
  - ٣٤٠٤٤٨ مع الفراغات
- الإحصائيات
  - الفتحة - : ٤٢٨٠٠
  - السكون - : ٢٣٣٢٧
  - الكسرة - : ٢١٨٦٤
  - الضمة - : ١٢٦١٣
- تابع الإحصائيات
  - الشدة - : ٣٠٢٩
  - تنوين الفتح - : ١٠٧٩
  - تنوين الكسر - : ١٠٦٩
  - تنوين الضم - : ٨٧٢
- إحصائيات أخرى
  - ال : ١٠٨٧٩
  - ا : ٣٤٤١٣ - ل : ٢٣٨٠٩
  - معدل الكلمات في الجملة ٦.١
  - معدل الرموز في الجملة ٤٦.٢ مع الحركات
  - معدل الرموز في الكلمة ٧.٦ مع الحركات



# احصائيات

حرف	تكرار	نسبة
ض	١٢٢٣	٠.٦
ز	٨٢٣	٠.٤
ث	٧٥٥	٠.٤
غ	٧٤١	٠.٤
؟	٧٢٣	٠.٤
ئ	٥٨٤	٠.٣
ء	٥٦٥	٠.٣
ظ	٤٢٩	٠.٢
ؤ	٢٠٤	٠.١
	١٩٠٧٩٧	١.٠

حرف	تكرار	نسبة
ف	٤٣٩٢	٢.٣
ك	٤١٧٢	٢.٢
ج	٤١٤٧	٢.٢
ه	٣٩٢٠	٢.١
ق	٣٣٨٧	١.٨
ح	٣٢٥٨	١.٧
ط	٢٦٠٨	١.٤
ص	١٩٦٧	١
خ	١٩٦٠	١
ى	١٥٩٩	٠.٨
ن	١٣١٦	٠.٧
ش	١٢٨١	٠.٧

حسني المحتسب

حرف	تكرار	نسبة
ا	٣٤٤١٣	١٨
ل	٢٣٨٠٩	١٢
ة	١٦٧٣٧	٨.٨
م	١٣٧٢٨	٧.٢
ي	١٢٩٨٢	٦.٨
ر	٧٩٧٤	٤.٢
ب	٧٧٣٠	٤.١
ن	٧٤١٧	٣.٩
ع	٧٤١٢	٣.٩
د	٦٨٨٤	٣.٦
و	٦٥٥٤	٣.٤
س	٥١٠٣	٢.٧

١٧

# نشاط: ما فائدة الإحصاءات؟

# خوارزمية عامة للتشكيل

- هل الكلمة من كلمات المجموعة المغلقة التي لها شكل واحد؟ شكلها
- هل الكلمة والكلمات الثلاثة التالية لها موجودة في مجموعات الرباعيات المترابطة (Quadra-gram)؟ شكلها وشكل رفيقاتها
  - هل نحتاج لاستخدام الخماسيات المترابطة؟
  - السداسيات المترابطة؟
  - السباعيات؟
- هل الكلمة والكلمتين التاليتين لها موجودة في مجموعات الثلاثيات المترابطة (Tri-gram)؟ شكلها وشكل رفيقاتها
- هل الكلمة والكلمة التالية لها موجودة في مجموعات الثنائيات المترابطة (Bigram)؟ شكلها وشكل رفيقتها. هل تشكل رفيقتها؟

# خوارزمية عامة للتشكيل (تابع)

- هل الكلمة صحيحة؟ هل نستخدم وحدة تصحيح آلية؟ صحح الكلمة وأعد الخطوات السابقة
- في حالة عدم تشكيل الكلمة في الخطوات السابقة تابع ما يلي حسب حروف الكلمة
- طبق القواعد الممكنة لتشكيل بعض حروف الكلمة
- عند إضافة أية حركة أعد تطبيق القواعد الممكنة
- إذا بقيت حروف غير مشكلة تابع الخطوات التالية

# خوارزمية عامة للتشكيل (تابع)

- هل الحرف والحروف والحركات الستة التالية له موجودة في مجموعات السباعيات المترابطة (seven-gram)؟ شكله وشكل رفقائه؟
  - هل نحتاج لاستخدام الثمانيات المترابطة؟
  - التساعيات المترابطة؟
  - العشريات؟
- هل الحرف والحروف والحركات الخمسة التالية لها موجودة في مجموعات السداسيات المترابطة (sex-gram)؟ شكله وشكل رفقائه
- هل الحرف والحروف والحركات الأربعة التالية له موجودة في مجموعات الخماسيات المترابطة (Five-gram)؟ شكله وشكل رفقائه

# خوارزمية عامة للتشكيل (تابع)

- هل الحرف والحروف والحركات الثلاثة التالية له موجودة في مجموعات الرباعيات المترابطة (Quadra-gram)؟ شكله وشكل رفقاه
- هل الحرف والرمزين التاليين له موجود في مجموعات الثلاثيات المترابطة (Tri-gram)؟ شكله وشكل رفيقه.
- هل الحرف والحرف التالي له موجود في مجموعات الثنائيات المترابطة (Bigram)؟ شكله وشكل رفيقه. هل تشكل رفيقه؟
- إذا لم يشكل الحرف حتى هذه الخطوة، اختر الحركة الأكثر احتمالية

# كلمات المجموعة المغلقة

- المجموعة المغلقة تحوي الكلمات التي لها شكل واحد
- نحتاج إلى تجميع هذه الكلمات
- الكلمات التي نتوقع أن يكون لها تشكيل فريد
  - كان
  - قال
  - ما
  - لا
  - طالما

# نشاط صفي : قواعد للتشكيل الآلي

- بعض القواعد المحتملة
  - الحرف قبل الألف مفتوح
  - الحرف قبل الواو الممدودة مضموم
  - الحرف قبل الياء الممدودة مكسور
- استخراج بعض القواعد



# متطلب ١٠ : شارك في إعداد الاختبار

• اطلع على المادة المغطاة واقترح سؤالين اثنين على الأقل ترى أن من المناسب أن يكونا جزءا من الاختبار الختامي للمادة مع ملاحظة ما يلي:

• نص كل سؤال بلغة عربية فصحي وبوضوح مع الصور والجداول عند الحاجة

• زمن الإجابة المتوقع لكل سؤال هو ١٠ دقائق

• بين ارتباط السؤال بأهداف المادة

• أكتب إجابة نموذجية لكل سؤال

• اجعل السؤال يتطلب نوعا من التفكير الناقد

وشكرا

سبحانك اللهم وبحمدك أشهد أن لا إله  
إلا أنت أستغفرك وأتوب إليك