

نظام حاسوبي لتشكيل النص العربي (المشكل الآلي)

الحوسبة العربية ICS484

صقر العريفي

فريق: ابداع



المسار

- مقدمة
- برامج التشكيل المتوفرة
- عيوبها
- طريقة عمل المشكل الآلي
- مثال
- مقارنة
- مميزات نظام التشكيل المستقل
- الخاتمة

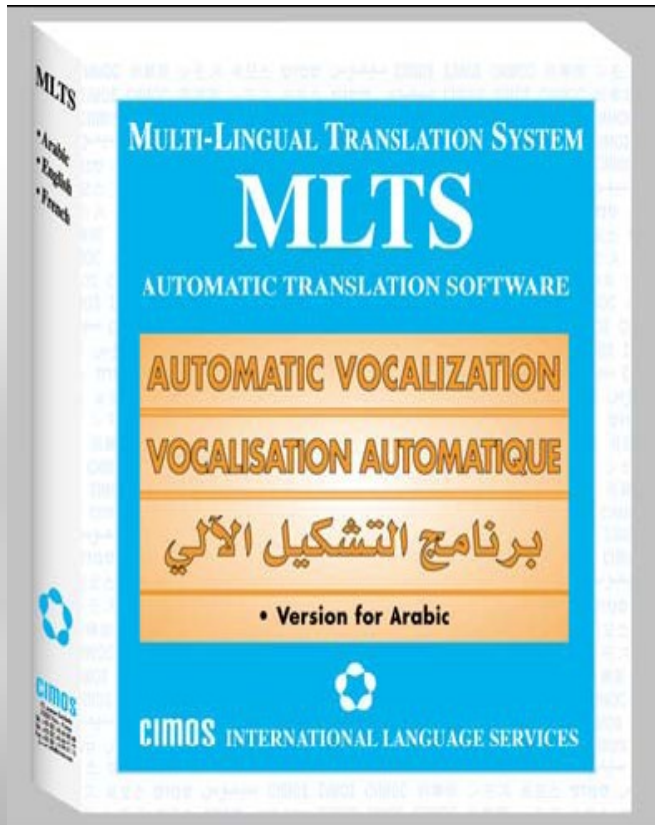
- يهدف المشروع إلى دراسة نظام التشكيل في اللغة العربية ومن ثم كتابة برمجيات تقوم بتشكيل النص العربي آليا بهدف دمج هذه البرمجيات في نظم حاسوبية أخرى كالناطق الآلي والتعرف الآلي على الكلام العربي ومحركات البحث وغيرها

مقدمة

- فريق العمل:
- د. منصور بن محمد الغامدي
- د. محمد خورشيد
- د. مصطفى الشافعي
- د. فايز الحرقان
- د. أبو أوس الشمسان
- د. محمد الكنهل
- سعد القحطاني
- سيد زيشان مظفر
- ياسر التويم
- عدنان يوسف
- حسني المحتسب

برامج التشكيل المتوفرة

- نسبة التشكيل الصحيح تساوي ٧٠ % تقريبا على مستوى الكلمة



برامج التشكيل المتوفرة

- الشركة العالمية
- سرعة عالية ودقة تصل إلى ٩٨ %

Diacritizer



برامج التشكيل المتوفرة

- قامت شركات أخرى بجهود في هذا المضمار ومن هذه الشركات : شركة أي بي إم، و إنفو آراب، وكولت

العيوب

- قليل من الشركات المذكورة توفر برنامج مشكلها الآلي مستقلا بذاته.
- المشكل الآلي المتاح مغلق المصدر.

طريقة عمل المشكل الآلي

• نظام التشكيل الآلي المستقل:

- استخلاص التسلسل الرباعي للحروف ذات أعلى احتمالية تشكيل وهي التسلسلات الرباعية الفاعلة في التشكيل فكان عددها ٦٨،٣٧٨ تسلسلا

الاحتمالية	التشكيل ٥		الحرف ٤	التشكيل ٣		الحرف ٢	التشكيل ٢		الحرف ١	التشكيل ١	الحرف ١
٠.١٨	٠	'	م	٠	٠	#	٠	'	د	٠	ج
١.٠٠	٠	'	هـ	٠	'	د	'	'	د	٠	ج
١.٠٠	٠	'	ء	٠	٠	ا	٠	'	د	٠	ج
٠.٩٢	٠	٠	#	٠	'	ر	٠	'	د	٠	ج

طريقة عمل المشكل الآلي

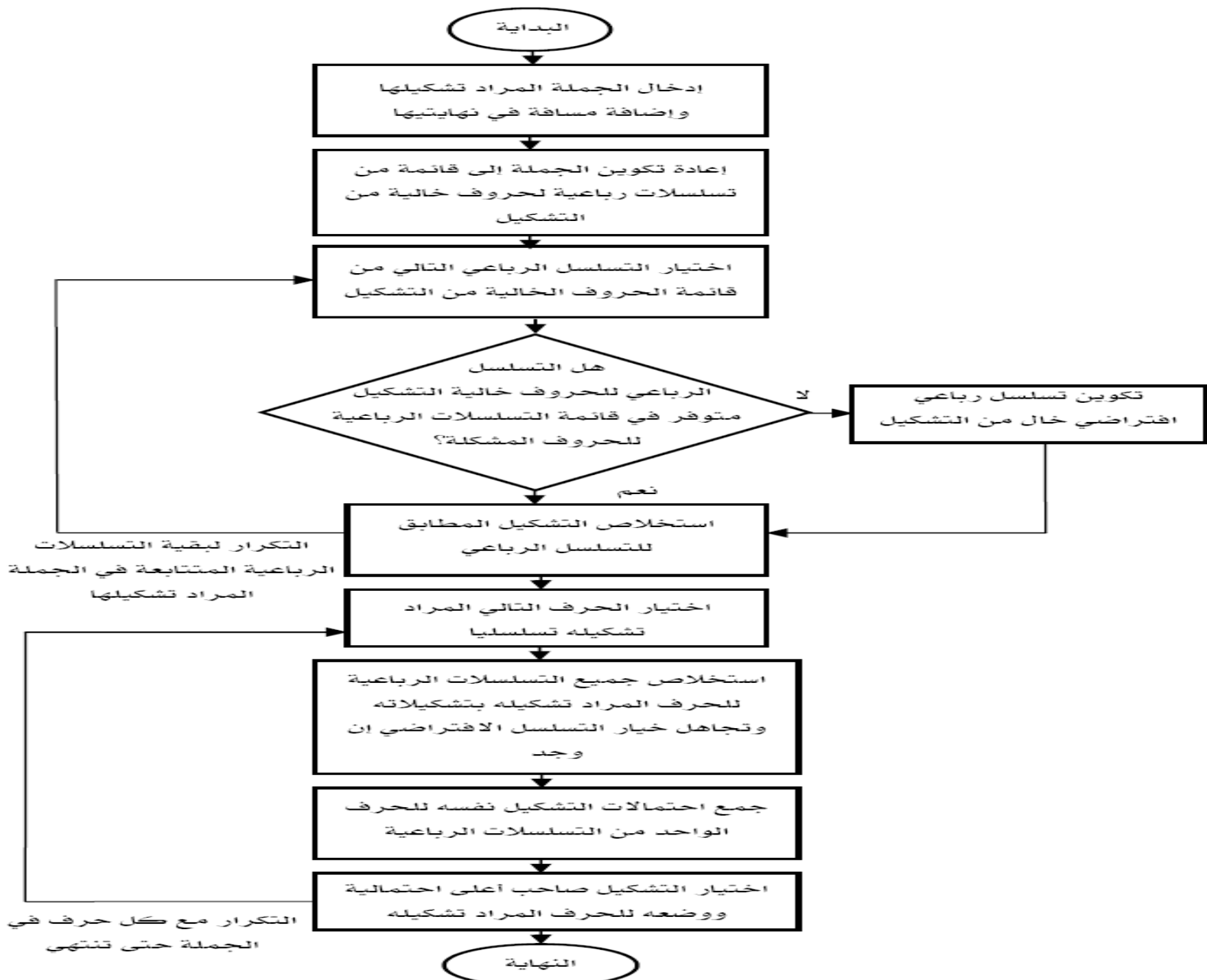
- تحليل النص غير المشكل إلى سلسلة من التسلسلات الرباعية للحروف المكونة له.
- استدعاء التسلسلات الرباعية للحروف المشكلة والتي تقابل التسلسلات الرباعية للحروف الغير مشكلة والمطلوب تشكيلها.
- الناتج:
- قائمة من التسلسلات الرباعية لحروف النص المراد تشكيلا مع تشاكيلا المختلفة واحتمالية ورود كل تسلسل.
- اختيار التشكيل الأعلى احتمالية، لكل حرف أربع احتمالات لتشكيله كحد أعلى. ثم تجمع الاحتمالات الموحدة وتدمج في احتمال واحد

مثال

• العبارة: المشكل الآلي للحرف العربي

٠،٤٩	٠	م	٠	ل	٠	٠	ا	٠	٠	#	#ال م
٠،٣٤	٠	ش	٠	م	٠	٠	ل	٠	٠	ا	ال م ش
٠،٨٩	٠	ك	٠	ش	٠	٠	م	٠	٠	ل	ل م ش ك
٠،٩٩	٠	ل	٠	ك	٠	٠	ش	٠	٠	م	م ش ك ل
٠،٤١	٠	#	٠	ل	٠	٠	ك	٠	٠	ش	ش ك ل #
٠،٣٧	٠	ا	٠	#	٠	٠	ل	٠	٠	ك	ك ل # ا
٠،٣٣	٠	ل	٠	ا	٠	٠	#	٠	٠	ل	ل # ا ل
٠،٩٢	٠	آ	٠	ل	٠	٠	ا	٠	٠	#	# ا ل آ
٠،٥٢	٠	ل	٠	آ	٠	٠	ل	٠	٠	ا	

المُشكِل الآلي للحرف العربي



مقارنة

• تم اخذ عينة من مقال بجريدة الرياض وتم تشكيله يدويا:

• وباستخدام احد انظمة التشكيل كانت

• النسبة ٧٤%

• بينما باستخدام هذا النظام

• كانت النسبة ٨٧%

مساعد المشكل - مقارنة النصوص المشكلة

Help Window Tools View Edit File

النص المشكل يدويا: فتح ملف

اختبار التطبيق

حذف التشكيل

مسح

النص المشكل آليا: فتح ملف

مسح

نتيجة المقارنة:

أخطاء التشكيل:

الخطأ	التكرار
— < —	7
— < —	4

النصان مختلفان قبل حذف التشكيل

النصان متطابقان بعد حذف التشكيل

نسبة الخطأ = 13.3333% نسبة مئوية

مميزات نظام التشكيل المستقل

- النظام الجديد مستقل تماما ويمكن برمجته بأية لغة برمجه والتصرف فيه بحثيا وتجاريا.
- يتعامل مع الحرف العربي مباشرة
- السرعة في الأداء (معدل فترة التشكيل 1.7 مليثانية) مقارنة بـ ٨٤.٢
- صغر الحجم مقارنة بالنظم الأخرى (٣ ميقا).
- ارتفاع نسبة التشكيل الصحيح
- النظام لا يستخدم أية قواعد أو قوانين نحوية أو صرفية

الخاتمة

- تم انتاج نظام للتشكيل الآلي المستقل يتميز سرعة ودقة في الأداء لا يحتوي على مكونات كالقواعد الصرفية أو النحوية

