

نظام حاسوبي لرومنة الأسماء العربية A Romanization System for Arabic Names

• فريق العمل:

- د: عبدالملك السلطان
- د: منصور الغامدي
- م: خالد الحقييل
- م: صالح الصبي

* هذا البحث مدعوم من قبل مدينة الملك عبدالعزيز للعلوم و التقنية

المقدمة

- ما هي الحاجة لرومنة الأسماء
- لماذا نحتاج إلى توحيد نتائج الرومنة

لمحات عن كتابة الأسماء العربية

- كتابة الاسم العربي باللغة العربية
 - النطق و الكتابة
 - الفونيم Phoneme
 - الغرافيم Grapheme
 - مكونات الاسم
 - الصوامت Constant
 - الصوائت Vowels
- كتابة الاسم العربي باللغات الأخرى
 - النقل الكتاب transliteration

لمحات عن أساليب الرومنة المستخدمة

- الرومنة بناء على السماع
 - + الإعتماد على السماع كوسيلة أساسية للترجمة
 - اختلاف النطق باختلاف المناطق
- الرومنة بناء على النقل الكتابي الإحصائي
 - + عمل دراسات إحصائية للحصول على الشكل الأكثر شيوعا للرومنة
 - النظام الإحصائي يقيس شيوع الإستخدام و لا يضمن صحة الرومنة
- الرومنة بناء على قواعد البيانات
 - + تعتمد على دراية مدخل البيانات بقواعد الرومنة و كذلك مهارته اللغوية في الرومنة و واللغة العربية
 - القدرات مختلفة من شخص لآخر مما يسبب تباينا في الرومنة

تصميم و برمجة نُظْم الرومنة (١)

- النقل الكتابي الدقيق للحروف العربية بإيجاد مقابل بالحرف الروماني لكل رمز عربي
مثال " سَمَر "

س = S ا = ʾ م = m ر = r

+ عن طريق معرفة الرومنة فقط يمكن الرجوع إلى الاسم العربي

- ٢١ رمزا في اللغة العربية لا يوجد لها نظير في الحرف الروماني،
وبالتالي يلزم استحداث رموز جديدة في الحرف الروماني

<i>iso</i>	<i>ini</i>	<i>med</i>	<i>fin</i>	DIN 31635 1982 (1.0)	ISO 233 1984 (2.0)	ISO/R 233 1961 (3.0)	UN 1972(4.0)	ALA-LC 1997(5.0)	EI 1960(6.0)
ا		ا	ا	ā	ʾ	— (3.1)(3.2)	— (4.1)	—	—
ء		ء	ء	ʾ	ʾ (2.1)	— (3.3)	ʾ (4.2)	— (5.1)	ʾ
ب	ب	ب	ب	b	b	b	b	b	b
ت	ت	ت	ت	t	t	t	t	t	t
ث	ث	ث	ث	ṭ	ṭ	ṭ	th	th	th
ج	ج	ج	ج	ǧ	ǧ	ǧ	j	j	dj
ح	ح	ح	ح	ḥ	ḥ	ḥ	ḥ	ḥ	ḥ
خ	خ	خ	خ	ḫ	ḫ	ḫ	kh	kh	kh

تصميم و برمجة نُظْم الرومنة (٢)

- دمج الرموز العربية لتقابل الرموز الرومانية: وذلك باستخدام رمز واحد بالحرف الروماني ليقابل أكثر من رمز في الحرف العربي، كما تحذف بعض الرموز عند الرومنة لعدم وجود مقابل لها

مثال الحروف "ص" و "س" تمثل بالحرف "s"

الحرف "ع" لا يقابله أي حرف في رومنة "عَلِي" = "Ali"

+ هذه الطريقة هي الأكثر شيوعا في الاستخدام بين الناس

+ سهولة التطبيق في النظم الحاسوبية المختلفة

- لا يمكن عمل التحويل العكسي من الرومنة إلى الاسم عربي

مكونات نظام الرومنة المقترح

- برنامج مصحح قائمة الأسماء
 - برنامج تهيئة قاعدة البيانات
 - برنامج الرومنة (محرك الرومنة)
 - واجهة رسومية للمستخدمين ضمن أنظمة التشغيل التي تدعم الواجهات الرسومية
- واجهة للعمل على نظام Windows

مصحح قائمة الأسماء

- تلخيص لفكرة عمل المصحح

– المدخلات

العمرى العُمريّ العُمريّ العَمريّ مثال

– فحص الأخطاء الكتابية

– تصحيح الأخطاء وإعادة الإدخال من جديد

– المخرجات

- ملف جاهز للإستخدام من الناحية اللغوية

برنامج تهيئة قاعدة البيانات (١)

- تلخيص لفكرة عمل برنامج التهيئة
 - الحركات يتم التعامل معها على شكل رموز بسبب أنها غير مدعومة في جميع الأنظمة
- المدخلات

- ملف نصي يحوي على الأسماء الأساسية مع التباكيل الممكنة لها يفصل بينها [TAB] وهي خاصية متوفرة في أغلب قواعد البيانات

– انشاء قاعدة للبيانات من قائمة الأسماء الموفرة في الملف النصي و تحت كل اسم توضع جميع التباكيل الممكنة له بالرموز المقابلة لها

برنامج تهيئة قاعدة البيانات (٢)

- مثال على المدخلات و المخرجات

```
<rawname name="      " count="2">  
<variations>00214100</variations>  
<variations>00130000</variations>  
</rawname>
```

محرك الرومنة (١)

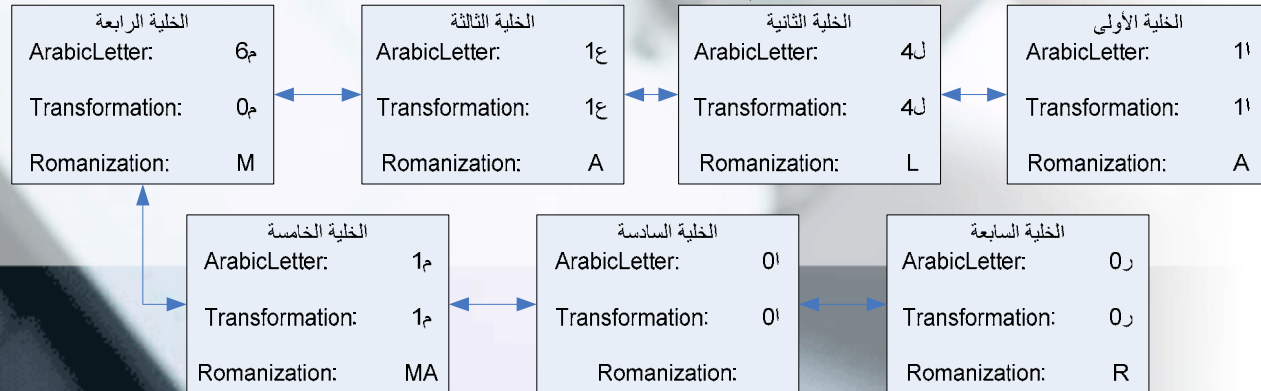
• تلخيص لفكرة عمل محرك الرومنة

- العمل يتم بناء على الترميز خاص الموافق لبيئة العمل التي يتم من خلالها استخدام المحرك
 - يتم العمل من خلال موجه الأوامر (Command Prompt) للحصول على قدر أكبر من الإستقلالية عن خصائص نظم التشغيل
- مثال

صالح C:\>Romanization.exe -encoding Windows.code

- تراكيب البيانات المستخدمة
- مثال

شكل تراكيب البيانات بعد رومنة الاسم "العَمَّار"



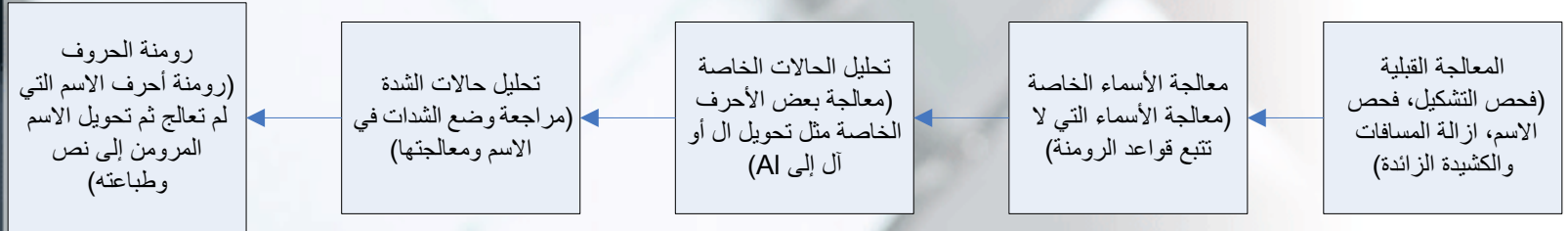
محرك الرومنة (٢)

• وظائف محرك الرومنة

- يستقبل اسم مشكلاً ومن ثم يعمل على تحليله وإخراج الرومنة الخاصة بهذا الاسم، و في حالة كون الاسم غير مشكلاً يعرض النظام التشاكيل الممكنة للاسم المراد رومنته
- القيام برومنة جميع الاسماء في قاعدة البيانات
- إضافة اسماء جديدة إلى قاعدة البيانات

• خطوات الرومنة

- خمس خطوات متتابعة في حالة فشل أحدها فإن الاسم يحتاج إلى تصحيح



محرك الرومنة (٣)

- تتم الرومنة بناء على قواعد اشترك مجموعة من المتخصصين في مجال اللغويات و الصوتيات لتحديد مقابلات الحروف العربية عند تحويلها للحرف الروماني
- يعامل الاسم المركب والكنية معاملة الاسم المفرد، وذلك بدمج الكلمات المكونة له في كلمة واحدة مرومنة
- تحذف الألف من ال التعريف في وسط الأسماء المركبة
- تعامل ال التعريف حسب كتابتها وليس حسب نطقها، و تختلف معاملة الحرف التالي بناء على كونها شمسية أو قمرية
- الحرف العربي له ثلاث حالات عند نقله للحرف الروماني
 - أن يكون المقابل له حرف روماني واحد
 - أن يكون المقابل له حرفان رومانيان
 - أن لا يكون له مقابل في الحرف الروماني
- الحروف المشددة يجب أن تلقى معاملة خاصة قبل البدء في عملية الرومنة و يختلف التعامل مع الحرف بناء على موقعة من الكلمة و حالة الحرف المشدد عن نقله للحرف الروماني

برنامج الواجهة

- هي عبارة عن واجهة للمستخدم لتسهيل عملية استخدام محرك الرومنة و الأدوات الأخرى المستخدمة في عملية الرومنة كإنشاء ملف ترميز جديد أو تعديل ملف موجود مسبقاً
- برنامج الواجهة يعمل كغطاء لمحرك الرومنة حيث بالإمكان تصميم واجهات مختلفة لنظام تشغيل واحد، كما يمكن عمل واجهة لكل نظام تشغيل
- الواجهة المصممة لنظام ويندوز
 - ثلاث خانات لإدخال الاسم الثلاثي كما في المعاملات الرسمية

الخاتمة

- الإختبارات و النتائج
- رومنة أكثر ٧٠,٠٠٠ اسم عربي مضبوط بالشكل
- المراجع و فرق العمل
- الخطوات المستقبلية

شكراً لكم على حسن الإستماع

...