

統一碼



Unicode

ترميز اللغة العربية في مواقع الإنترنت دراسة حالة للمواقع السعودية

د. هند بنت سليمان الخليفة

قسم تقنية المعلومات – كلية علوم الحاسب والمعلومات
جامعة الملك سعود

hend@ccis.edu.sa

المحاور الرئيسية

- مقدمة
- احصائيات
- ترميز اليونيكود: ماهيته وأهميته
- أهمية الدراسة وأهدافها
- منهج وعينة الدراسة
- نتائج الدراسة
- النتائج والتوصيات

統一碼



يونيكود
Unicode

مقدمة

• سابقاً:

- عرض النص العربي مر بكثير من المراحل في بدايات ظهوره على الإنترنت.
- صعوبة تمثيل النص العربي وعرضه بسهولة على جميع المتصفحات.

• حالياً:

- ظهور الرمز العالمي الموحد (Unicode).

統一碼



Unicode

احصائيات ٢-١

Script	Latin	Cyrillic	Arabic	Hanzi	Indic	Others*
Million users	2,238	451	462	1,085	807	129
% of Total	43.3%	8.7%	8.9%	21.0%	15.6%	2.5%
Key languages	Romance (European)	Russian Slavic (some)	Arabic Urdu Persian Pashtu	Chinese Japanese Korean	Hindi Tamil Bengali Punjabi Sanskrit Thai	Greek Hebrew Georgian Assyrian Armenian

إحصائية لتوزيع الكتابات المستخدمة على الإنترنت (Paolillo، 2005)

統一碼



Unicode

احصائيات ٢-٢

- نسبة المواقع العالمية التي تستخدم ترميز اليونيكود (utf-8) بلغت ٢٠,٤% مقارنة بترميز (iso-8859-1) للنصوص الإنجليزية، والذي شكلت نسبة انتشاره ٤٢,٨%.

– (Mark Davis) خبير الترميز من شركة قوقل ٢٠٠٦

統一碼



Unicode

ترميز اليونيكود: ماهيته



”... نظام ترميز عالمي، يُستخدم لتمثيل النصوص من أجل معالجتها على الكمبيوتر بحيث يخصص رقم وحيد لكل محرف في جميع اللغات العالمية، وذلك بغض النظر عن نوع الحاسوب أو البرامج المستخدمة... وتتضمن اليونيكود رموز اللغات الأوربية، ورموز اللغات التي تتم كتابتها من اليمين إلى اليسار، كاللغة العربية، ورموز اللغات الآسيوية. وتشمل مواصفة يونيكود أيضاً علامات التنقيط، والأحرف المميزة (diacritics)، والرموز الرياضية، والرموز التقنية، والأسهم.. إلخ”
– (يونيكود، ٢٠٠٦)

ترميز اليونيكود: أهميته (٢-١)

- التوجه العام لدى مطوري مواقع الإنترنت هو اعتماد هذا الترميز العالمي.
- يساعد الترميز العالمي المواقع متعددة اللغات على عرض غالبية اللغات إن لم يكن كلها في نفس الصفحة بشكل صحيح ومقروء.
- يتميز ترميز اليونيكود بأنه قابل للعرض على مختلف أنظمة التشغيل (cross-platform)، بسبب دعم معظم أنظمة التشغيل لهذا الترميز.

統一碼



يونيكود
Unicode

ترميز اليونيكود: أهميته (٢-٢)

• يحل ترميز اليونيكود مشكلة استعمال الرمز نفسه من قبل أكثر من لغة واحدة لأن كل حرف يمثل برمز خاص به، وبذلك يصبح لدينا حيزاً أكبر في تمثيل مختلف الحروف المكتوبة.

• تمثيل الأحرف بالترميز العالمي يفتح آفاقاً رحبة من الاستخدامات العالمية ويجعل المواقع العربية قابلة للتدويل (internationalization)، كما أنها ستساعد محركات البحث على فهرسة هذه المواقع بسهولة.

統一碼



Unicode

أهمية الدراسة وأهدافها

• أهميتها:

- بحث سبب ظهور بعض صفحات اللغة العربية في مواقع الإنترنت على شكل رموز أو علامات استفهام مما يضطره لتبديل لغة الواجهة من قائمة المتصفح.
- استخدام ترميز غير مدعم عالمياً يجعل من الصعوبة بمكان تمثيل أكثر من لغة في نفس الصفحة.

• أهدافها:

- قياس مدى انتشار استخدام ترميز اليونيكود في صفحات المواقع السعودية.
- اختبار العلاقة بين الترميز المستخدم ونوع الخادم النسيجي.

統一碼

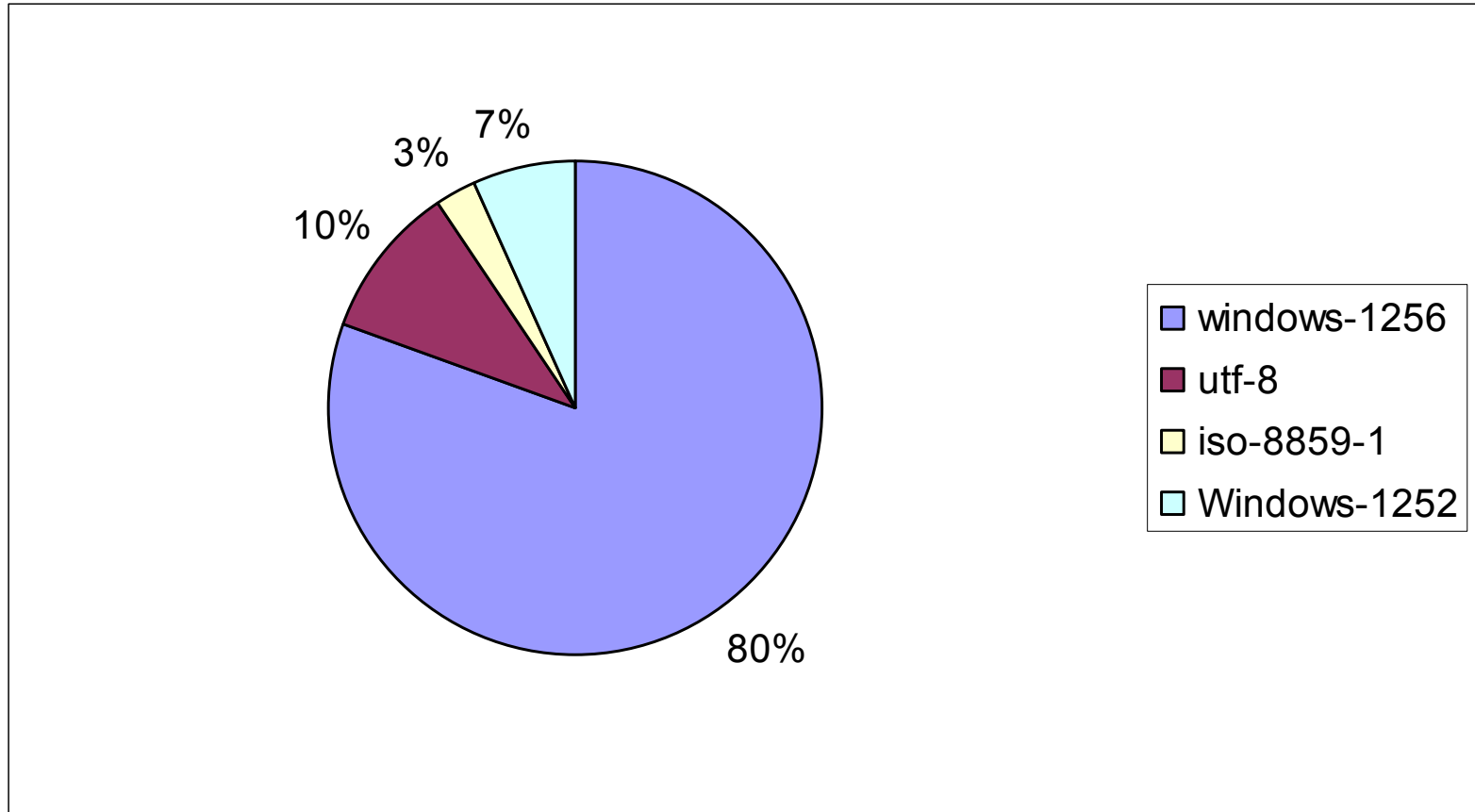


Unicode

منهج وعينة الدراسة

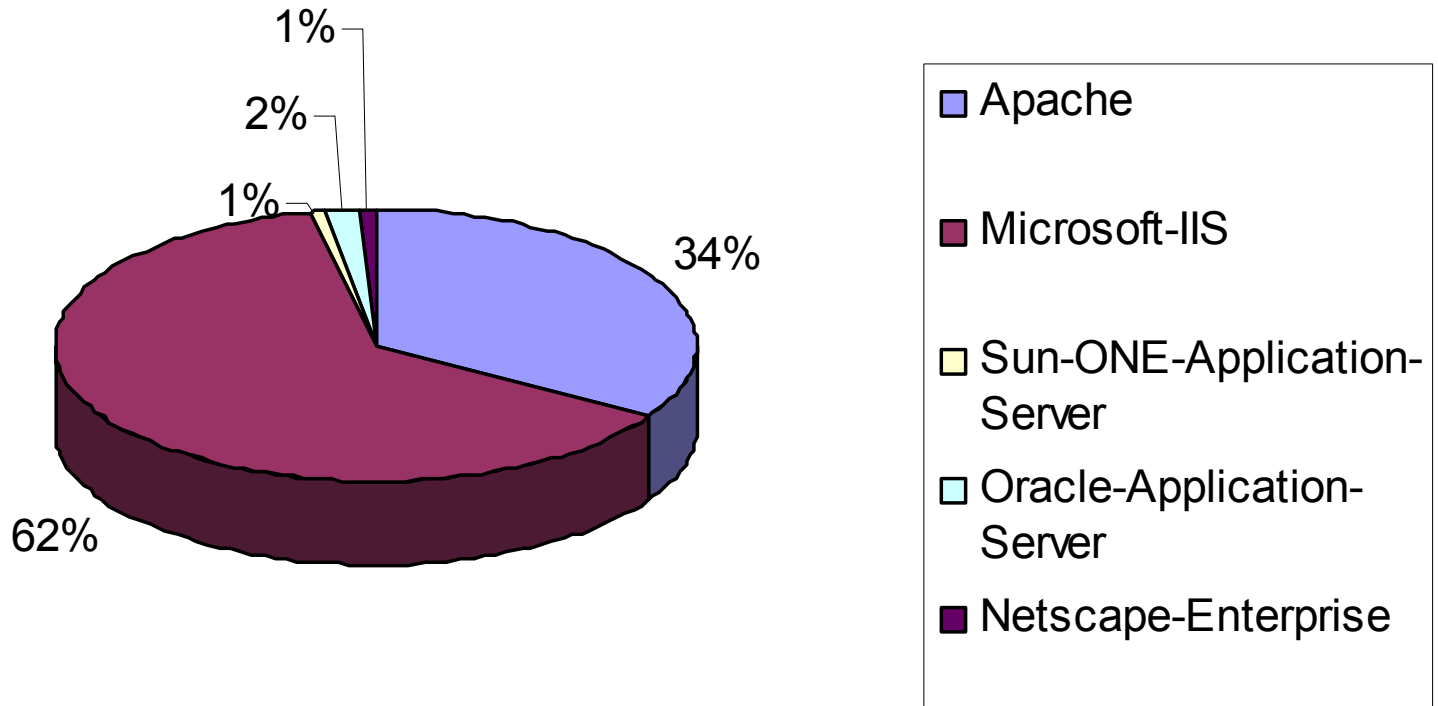
	()

نتائج الدراسة (الاحصاءات الوصفية ٢-١)



نسبة استخدام ترميز اليونيكود في المواقع السعودية بلغت ١٠% مقارنة بترميز ويندوز (windows-1256) والذي بلغ ٨٠% 11

نتائج الدراسة (الاحصاءات الوصفية ٢-٢)



نسبة توزيع أنواع الخوادم في العينة كانت ٦٢% لصالح خوادم شركة مايكروسوفت (Microsoft-IIS) مقارنة بـ ٣٤% لصالح خوادم أباتشي (Apache).

نتائج الدراسة (التحليل الاحصائي)

Encoding → WebServer↓	windows-1256	utf-8	iso-8859-1	windows-1252
Apache	33	3	1	3
Microsoft-IIS	61	7	1	5
Others	1	2	1	0

اختبار مربع كاي للاستقلالية أسفر عن وجود علاقة ذات دلالة إحصائية بين الترميز ونوع الخادم النسيجي عند مستوى معنوية $p \leq 0,025$

النتائج

- قلة استخدام المواقع السعودية لترميز اليونيكود واعتمادها بشكل كبير على ترميز شركة مايكروسوفت (windows-1256) سواءً كان الخادم النسيجي من نوع أباتشي أو مايكروسوفت.
- نسبة ٦% من المواقع وأغلبها حكومية، أهملت إدراج نوع الترميز في بداية الصفحة الرئيسية في موقعها.
- أوضحت النتائج الإحصائية أن ثمة علاقة بين نوع الخادم النسيجي والترميز المستخدم في إظهار النصوص العربية، حيث أن معظم خوادم شركة مايكروسوفت في عينة الدراسة استخدمت ترميز ويندوز بصيغتها المختلفتين (windows-1256,-1252).

統一碼



يونيكود
Unicode

التوصيات

- ينصح بإدراج نوع الترميز في صفحات الموقع علنياً لتجنب الزائر معمة التحويل. وذلك بإدراج النص التالي في رأس الصفحة:

```
<meta http-equiv="Content-Type"  
content="text/html; charset=UTF-8" />
```

- توعية الجهات المختلفة -سواءً كانت حكومية أو قطاع خاص- والتي لها مواقع على شبكة الإنترنت بأهمية استخدام الترميز العالمي (اليونيكود) في بناء مواقعها، وذلك حتى يمكن الوصول لها وقراءتها بسهولة وتصبح قابلة للتدويل (internationalization) وقابلة للتبادل (Interoperability) في حال تصدير بياناتها لتطبيقات أخرى.

統一碼



Unicode

شكراً لحسن استماعكم!

www.hend-alkhalifa.com

توحيد



يونيكود

Unicode