

---

<http://www.m22b.com>  
<http://www.al-islam.com/arb>  
<http://www.snqc.org/>  
<https://addons.mozilla.org/firefox/2036/>  
<http://news.netcraft.com/>  
[http://net-square.com/httpprint/httpprint\\_\\_paper.html](http://net-square.com/httpprint/httpprint__paper.html)  
[http://news.netcraft.com/archives/web\\_\\_server\\_\\_survey.html](http://news.netcraft.com/archives/web__server__survey.html)

ترميز اللغة  
العربية  
في مواقع  
الإنترنت

ويتم ذلك بوضع شفرة برمجية في بداية الملف الرئيس في الموقع والتي من نوع HTML or XHTML لبيان نوع الترميز، وفي المثال التالي حددنا نوع الترميز باليونيكود:

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

أما إذا كان الملف الرئيس من نوع XML، فيستخدم الأمر التالي:

```
<?xml version="1.0" encoding="utf-8" ?>
```

وهناك طرق أخرى لإدراج نوع الترميز يمكن الحصول عليها من (Teruhiko، ٢٠٠٢).

2

٢. 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. وهذا التمثيل لا يشمل فقط محتوى الصفحة بل أيضاً إعدادات الخادم النسيجي وقواعد البيانات بالإضافة لعرض الصفحة نفسها.

وأخيراً، لا بد من الإشارة أن صغر حجم العينة قد يكون عاملاً رئيساً في جوهرية العلاقة بين المتغيرات محل الدراسة، ونرى أنه لا يمكن تعميم هذه النتيجة على جميع المواقع السعودية أو حتى العربية إلا بعد عمل دراسة مشابهه ولكن على عينة أكبر وذلك لتعميم النتيجة.

شكر

تشكر الباحثة كل من الأخت أثير الخليفة والأستاذة أمل الابراهيم والأستاذة شهناز المقبل لمساهمتهن في إنجاح هذا البحث.

## المراجع

- يونيكود. ما هي الشفرة الموحدة "يونيكود"؟ (٢٠٠٦). متاح على الإنترنت <http://www.unicode.org/standard/translations/arabic.html>
- محمد زكي محمد خضر. الحرف العربي والحوسبة. (٢٠٠١). الموسم الثقافي لمجمع اللغة العربية. عمان، الأردن. متاح على الإنترنت. [http://www.al-mishkat.com/khedher/Papers/paper\\_v.htm](http://www.al-mishkat.com/khedher/Papers/paper_v.htm)
- محمد أنس طويلة. البرمجيات الحرة ومفتوحة المصدر واللغة العربية. (٢٠٠٥). المجلة العربية للعلوم والتقنية. متاح على الإنترنت. <http://www.linux.org.sy/node/51>
- Adam Stone. "Internationalizing the Internet." IEEE Internet Computing, vol. 07, no. 3, pp. 11-12. May/June, 2003.
- John Paolillo. "Language Diversity on the Internet." pp. 43-89. in John Paolillo, Daniel Pimienta, Daniel Prado, et al. Measuring Linguistic Diversity on the Internet. UNESCO Publications for the World Summit on the Information Society 2005. 113 pp. Available online. [http://www.uis.unesco.org/template/pdf/cscl/MeasuringLinguisticDiversity\\_\\_En.pdf](http://www.uis.unesco.org/template/pdf/cscl/MeasuringLinguisticDiversity__En.pdf)
- KUROSAKA Teruhiko. The Reality of Web Encoding Identification or Lack Thereof. (2002). 22nd International Unicode Conference. San Jose, California. Available online: <http://www.sonic.net/~kuro/IUC22-A5-320.pdf>
- Mark Needleman. The Unicode Standard. Serials Review. Volume 26, Issue 2, August 2000. Pages 51-54.

## هوامش

- <http://global-reach.biz/globstats/index.php3>
- <http://www.w3.org/International/questions/qa-who-uses-unicode>. Unicode at Google. (2006-11).
- <http://macchiato.com/>
- [http://en.wikipedia.org/wiki/Character\\_\\_encoding](http://en.wikipedia.org/wiki/Character__encoding)
- <http://www.raddadi.com/> تمت الاستعانة بدليل الرادادي لتصنيف المواقع السعودية، و يعتبر موقع الرادادي واحداً من أقدم وأشمل الأدلة العربية على شبكة الإنترنت

كما يوضح الجدول ٤ تفصيل توزيع الخادمتان النسيجية حسب تصنيف الموقع. ويلاحظ من الجدول أن تصنيف إمارات المناطق والصحف والمجلات هي من أكثر التصنيفات استخداماً لخادم الأباتشي، بينما تعتبر تصنيفات الوزارات والمديريات والمجالس والمصالح والمؤسسات والشركات من أكثر المواقع اعتماداً على خادمتان مايكروسوفت.

كما يبين الجدول استخدام نسبة لا بأس بها من مواقع الوزارات والمديريات والجامعات والمؤسسات والشركات لخادمتان نسيجية أخرى مثل خادم (Oracle-Application-Server) وخادم (Netscape-Enterprise) وخادم (Sun-ONE-Application-Server).

التصنيف	خادم أباتشي (Apache)	خادم مايكروسوفت (Microsoft-IIS)	خادمتان أخرى
وزارات	25%	68.75%	6.25%
إمارات مناطق	60%	40%	-
مديريات وهيئات	17.6%	76.4%	5.9%
جمعيات ولجان	50%	50%	-
مجالس ومصالح	33.3%	66.6%	-
إدارات أمنية	50%	50%	-
جامعات وكليات ومعاهد	45%	50%	5%
مؤسسات وشركات	15.4%	77%	7.6%
صحف ومجلات	53%	47%	-
أخرى (متاحف وسياحة)	22.22%	77.77%	-

جدول ٤ :  
توزيع نوع الخادم  
النسيجي حسب  
تصنيف المواقع

#### - التحليل الإحصائي

تم استخدام اختبار مربع كاي للاستقلالية (Chi Square Test of Independence) لمعرفة الارتباط بين المتغيرين (bivariate) محل الدراسة وهما: نوع الترميز (Encoding) المستخدم ونوع الخادم النسيجي (Web Server) كما هو مبين في الجدول ٥.

وأُسفرت النتيجة عن وجود علاقة ذات دلالة إحصائية بين الترميز ونوع الخادم النسيجي عند مستوى معنوية  $p \leq 0.025$ .

#### مناقشة النتائج والتوصيات

كشفت نتائج الدراسة الحالية عن قلة استخدام المواقع السعودية لترميز اليونيكود واعتمادها بشكل كبير على ترميز شركة مايكروسوفت (windows-١٢٥٦) سواء كان الخادم النسيجي من نوع أباتشي أو مايكروسوفت.

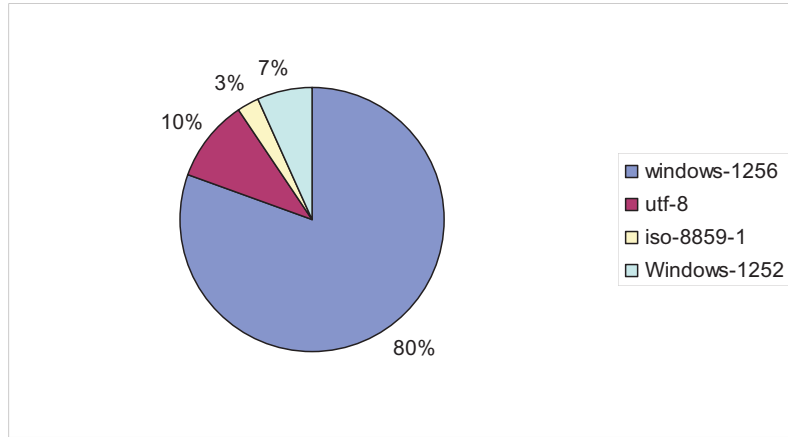
أما في الحالات التي استخدم فيها ترميز اليونيكود UTF-8 سواء في خادم الأباتشي أو مايكروسوفت تبين أن ظهور هذا الترميز مع خادمتان مايكروسوفت كانت أكثر من ظهورها مع خادمتان الأباتشي، مما يعني وجود محاولات متواضعة لاستخدام هذا الترميز. كما أنه من الملاحظات التي لفتت انتباهنا عند تحليل المواقع أن نسبة ٦٪ منها وأغلبها حكومية، أهملت إدراج نوع الترميز في بداية الصفحة الرئيسية في موقعها.

كما أوضحت النتائج الإحصائية أن ثمة علاقة بين نوع الخادم النسيجي والترميز المستخدم في إظهار النصوص العربية، حيث أن معظم خادمتان شركة مايكروسوفت في عينة الدراسة استخدمت ترميز ويندوز بصيغتيها المختلفتين (windows-١٢٥٦، -١٢٥٢). ونعتقد أن سبب استخدام هذه الترميزات هو أنها تأتي تلقائياً عند تركيب الخادم النسيجي أو عند إعداد عرض اللغة العربية، مما يعني أن مطور الموقع لم يعمل على تغييره ليستخدم الترميز العالمي أو قد يكون السبب من أن المشرف لا يدرك أهمية هذا الترميز.

ومن خلال استعراضنا للنتائج السابقة خلصنا إلى عدد من التوصيات وهي كالتالي:

١. من الأسباب التي تجعل زائر موقع ما يعمل على تغيير الترميز عندما يواجه صفحة غير مقروءة باللغة العربية، هو أن ترميز اللغة لم يذكر بشكل صريح في رأس الملف الرئيسي للموقع، وهذا ما شاهدناه في بعض المواقع الحكومية. لذا ينصح بإدراج نوع الترميز في صفحات الموقع علنياً لتجنب الزائر معمة التحويل.

شكل ١ :  
توزيع نسب  
الترميز حسب  
استخدامها في  
عينة الدراسة



4

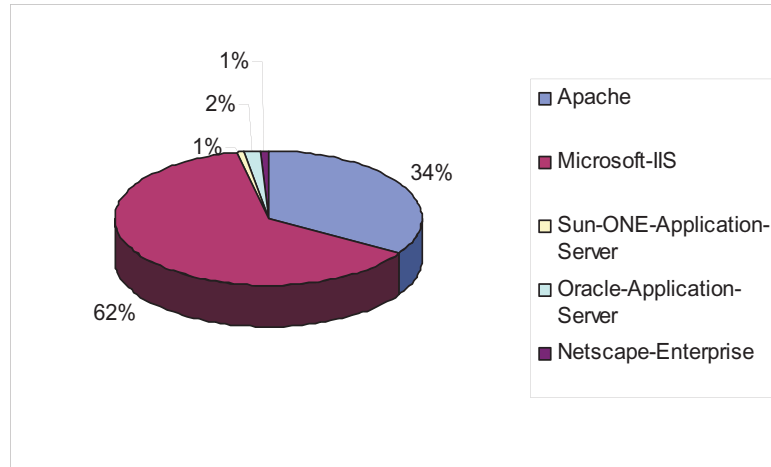
الترميزات الأخرى	ترميز ويندوز (windows-1256)	ترميز اليونيكود	التصنيف
%12.5	%62.5	%25	وزارات
%20	%80	-	إمارات مناطق
%6.25	%81.25	%12.5	مديريات وهيئات
%40	%60	-	جمعيات ولجان
%25	%75	-	مجالس ومصالح
%25	%66.6	%8.3	إدارات أمنية
-	%95	%5	جامعات وكليات ومعاهد
%23.07	%69.23	%7.7	مؤسسات وشركات
%11.75	%76.47	%11.76	صحف ومجلات
%10	%80	%10	أخرى (متاحف وسياحة)

جدول ٣ :  
توزيع الترميز  
حسب تصنيف  
المواقع

كما بينت نتائج الدراسة أن نسبة توزيع أنواع الخوادم في العينة (صورة ٢) كانت ٦٢٪ لصالح خوادم شركة مايكروسوفت (Microsoft-IIS) مقارنة بـ ٣٤٪ لصالح خوادم أباتشي (Apache).

وهذه النتيجة مغايرة للاتجاه العالمي في استخدام الخوادم مفتوحة المصدر، ففي إحصائية أجرتها شركة Netcraft في يناير من هذا العام (٢٠٠٧) على أكثر من ١٠٦ مليون موقعاً على الشبكة العنكبوتية، تبين أن أكثر من ٦٠٪ من جميع المواقع التي تم مسحها تعمل على خادمات أباتشي المفتوحة المصدر مقابل ٣٠,٧٨٪ لخادمت شركة مايكروسوفت.

شكل ٢ : توزيع  
نسب خوادم  
الاستضافة في  
العينة محل  
الدراسة



كما تمت الاستعانة بمواقع تقدم خدمة قراءة رؤوس الطلب، حيث يمكن معرفة معلومات عن ترميز موقع ما ونظام الخادم عن طريق قراءة رؤوس الطلب (HTTP Header). وتعمل مثل هذه الخدمات على طلب معلومات من الخادم واسترجاع نتيجة هذا الطلب أو إظهار أي معلومات يقوم الخادم النسيجي بالرد به. من هذه المواقع موقع (web-sniffer.net) وموقع (www.softswot.com/) (http-header.php) وخدمة (Webserver Search) المقدمة من موقع (netcraft).

أما في بعض الحالات والتي تشتمل فيها الخدمات السابقة في استرجاع نوع الترميز لموقع ما بسبب استخدام الموقع لأسلوب الإطارات (Frames) في تصميمه، نضطر أحياناً لزيارة الموقع المراد تحليله وعرض مصدر الصفحة الرئيسية للبحث عن نوع الترميز.

وبما أن هناك احتمالية اتخاذ بعضاً من المواقع السعودية عدداً من الإجراءات الأمنية لحماية خادماتها من الاختراقات عند إرسال مثل هذه البيانات مثل استخدام طريقة "التعميم" أو "التمويه"، فقد تم التأكد من صحة المعلومات التي أرجعتها تلك الخدمات السابقة باستخدام أداة (httpprint) وهو عبارة عن محرك لكشف بصمات الخادم النسيجي (fingerprinting engine) يستخدم طرق التحليل الإحصائي مع تقنيات المنطق المبهم (fuzzy logic) لتحديد نوع الخادم النسيجي وذلك في حالة إذا ما كانت البيانات المرجعة مموهة. وقد أكدت الأداة نفس النتائج التي توصلنا إليها باستخدام الطرق السابقة الذكر.

التصنيف	عدد المواقع
وزارات	16
إمارات مناطق	5
مديريات وهيئات	16
جمعيات ولجان	6
مجالس ومصالح	12
إدارات أمنية	11
جامعات وكليات ومعاهد	20
مؤسسات وشركات	13
صحف ومجلات	17
أخرى (متاحف وسياحة)	10
<b>المجموع</b>	<b>126</b>

جدول ٢ :  
توزيع المواقع في  
عينة الدراسة  
حسب تصنيفه

#### نتائج الدراسة

سنعرض فيما يلي نتائج الدراسة وفقاً لطريقة تحليل البيانات التي اتبعناها، وهي كالتالي:

##### ١- الإحصاءات الوصفية

وهنا تمت الاستعانة بأدوات الإحصاء الوصفي لحساب النسب المئوية للبيانات محل الدراسة ورسومها بيانياً. وقد أظهرت النتائج كما في صورة ١، أن إجمالي نسبة استخدام ترميز اليونيكود في المواقع السعودية بلغت ١٠٪ مقارنة بترميز ويندوز (windows-١٢٥٦) والذي بلغ ٨٠٪.

كما أظهرت النتائج أن هناك مواقع استخدمت ترميز تابع لشركة مايكروسوفت غير مخصص للغة العربية مثل ترميز (windows-١٢٥٢) المخصص للغة اللاتينية في أنظمة تشغيل ويندوز باللغة الانجليزية. كما بينت النتائج أن بعض المواقع استخدمت ترميز (windows-١٢٥٢) المخصص للغة اللاتينية أيضاً.

الجدول ٢ يوضح تفصيل توزيع الترميزات حسب تصنيف الموقع. ففي الجدول الموضح يلاحظ غياب استخدام ترميز اليونيكود في مواقع إمارات المناطق والجمعيات واللجان والمجالس، بينما نجد مواقع الوزارات من أكثر التصنيفات استخداماً لترميز اليونيكود. في المقابل نلاحظ أن الترميزات الأخرى، ونعني بها إما ترميز (windows-١٢٥٢) أو ترميز (windows-١٢٥٢)، قد غابت فقط عن تصنيف الجامعات والكليات والمعاهد.

كما توضح النتائج أن أكثر تصنيف استخداماً لترميز ويندوز (windows-١٢٥٦) هي مواقع المديريات والهيئات، بينما نجد أن مواقع الجمعيات واللجان هي أقل التصنيفات استخداماً لترميز اليونيكود ولكنها في ذات الوقت استخدمت ترميزات موازية لترميز اليونيكود مثل ترميز (windows-١٢٥٢) و ترميز (windows-١٢٥٢).

١. التوجه العام لدى مطوري مواقع الإنترنت هو اعتماد هذا الترميز العالمي.

٢. يساعد الترميز العالمي المواقع متعددة اللغات على عرض غالبية اللغات إن لم يكن كلها في نفس الصفحة بشكل صحيح ومقروء.

٣. يتميز ترميز اليونيكود بأنه قابل للعرض على مختلف أنظمة التشغيل (cross-platform)، بسبب دعم معظم أنظمة التشغيل لهذا الترميز (Stone، ٢٠٠٣).

٤. 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 (خضر، ٢٠٠١).

٥. 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.

### أهمية الدراسة وأهدافها وفرضياتها

كثيراً ما يعاني المتصفح العربي من مشكلة ظهور بعض صفحات اللغة العربية في مواقع الإنترنت على شكل رموز أو علامات استفهام مما يضطره لتبديل لغة الواجهة من قائمة المتصفح. وقد يسبب مثل هذا الموقف لبعض المواقع التجارية أو الحكومية إلى فقدان زوارها، علاوة على أن مشاكل اختلاف الترميز يسبب في ظهور محتوى تلك المواقع بشكل غير صحيح.

كما أن استخدام ترميز غير مدعم عالمياً يجعل من الصعوبة بمكان تمثيل أكثر من لغة في نفس الصفحة. لذا هدفت هذه الدراسة لما يلي:

١. قياس مدى انتشار استخدام ترميز اليونيكود في صفحات المواقع السعودية.

٢. اختبار العلاقة بين الترميز المستخدم ونوع الخادم النسيجي.

وعلى ضوء أهداف البحث قمنا بصياغة فرضية للتأكد من صحتها، وهي كالتالي: توجد علاقة جوهرية بين الترميز ونوع الخادم النسيجي.

وبناءً على هذه الفرضية ستكون متغيرات الدراسة ما يلي: الترميز كمتغير تابع، ونوع الخادم النسيجي كمتغير مستقل.

### منهج وعينة الدراسة

تم استخدام منهج دراسة الحالة في تصميم هذه الدراسة وهي إحدى المناهج المستخدمة في البحوث التحليلية، حيث اعتمدنا على جمع البيانات من عينة مختارة من المواقع السعودية اختيرت اختياراً عشوائياً لتمثيل المجتمع المراد دراسته.

وقد حاولنا أن تكون العينة المختارة ممثلة للمواقع السعودية التي تقدم خدمات لزوارها، ومما لاحظناه خلال مرحلة جمع البيانات غيابٌ عددي في تصنيفي إمارات المناطق والجمعيات واللجان مقارنة ببقية التصنيفات، كما وجدنا أن بعضاً منها لم يعد يعمل مما أثر سلباً على نسبة المواقع الممثلة لهذا التصنيف.

وأخيراً لا بد من الإشارة هنا أن عناوين العينة المختارة ليست من الضرورة أن تنتهي باللاحقة (sa)، حيث يوجد العديد من المواقع السعودية في عينة الدراسة مما لا تنتهي بهذه اللاحقة مثل موقع الرئاسة العامة لشؤون الحرمين ووزارة الشؤون الإسلامية واللجنة الوطنية السعودية للجودة وبعض مواقع الصحف والمجلات وغيرها.

وبناءً على ما سبق، فإن عينة الدراسة قد شملت على أكثر من ١٢٠ موقعاً لمؤسسات حكومية وتجارية في المملكة العربية السعودية. ويوضح جدول رقم ٢ توزيع المواقع حسب تصنيفها.

### إجراءات جمع البيانات

تم استخدام أكثر من أداة لاستخلاص معلومات عن نوع الترميز والخادم النسيجي للمواقع محل الدراسة. فقد استعنا بإضافة (Plug-in) تدعى (Server Spy) لمتصفح الفايروفوكس (Firefox) تعمل على عرض نوع الخادم النسيجي في شريط يظهر أسفل الموقع الذي تمت زيارته.

ومع تطور التقنيات وتنوعها ظهر الرمز العالمي الموحد أو ما يعرف اختصاراً بترميز اليونيكود (Unicode) لحل مشكلة تمثيل الأحرف غير اللاتينية، وبدأت بعضاً من المواقع العربية استخدامها كونها الحل الأمثل لعرض صفحات الموقع بصورة أوسع خصوصاً مع دعم معظم أنظمة التشغيل لهذا الترميز.

ومع ذلك لم تمتثل معظم المواقع العربية لترميز اليونيكود المعروف بـ UTF-8 وبدأت باستخدام أنواع أخرى للترميز حسب ما يعلي عليهم الخادم النسيجي ونظام التشغيل، فاستخدم ترميز ويندوز ١٢٥٦-Windows وترميز ISO-٨٨٥٩-٦ كبدلين لترميز اليونيكود.

والسؤال المطروح هو، ما مدى وعي المواقع العربية بأهمية استخدام ترميز اليونيكود؟ خصوصاً أن كثيراً من المواقع العالمية اتجهت لاستخدام هذا الترميز العالمي للحروف. فعلى سبيل المثال مواقع الشركات الكبرى مثل قوقل ومايكروسوفت والبريد الإلكتروني مثل بريد ياهو وقوقل وبعض المواقع الخدمية كمواقع المفضلات الاجتماعية والمدونات وموسوعة ويكيبيديا والأدلة تستخدم هذا الترميز وتعتمد عليه.

وفي إحصائية حديثة لنسبة انتشار ترميز اليونيكود على شبكة الإنترنت لخبير الترميز والباحث في شركة قوقل (Mark Davis) تبين أن نسبة المواقع التي تستخدم ترميز اليونيكود (utf-8) كانت ٤, ٢٠٪ مقارنة بترميز (iso-٨٨٥٩-١) للنصوص الإنجليزية، والتي شكلت نسبة انتشارها ٨, ٤٢٪. وهذه المعلومات جمعت بناءً على معطيات الـ (META) المأخوذة من رأس صفحة المواقع المخزنة في فهرس محرك البحث قوقل.

ونظراً لأهمية استخدام ترميز اليونيكود في المواقع العربية عامة والسعودية منها خاصة، سنعمل في هذه الدراسة على اختبار مجموعة من المواقع التي تستخدم اللغة العربية في المملكة العربية السعودية وذلك لرؤية مدى انتشار استخدام ترميز اليونيكود في هذه المواقع، وأيضاً محاولة تفسير العلاقة بين استخدام ترميز ما ونوع الخادم النسيجي. والسبب في حصر التركيز في هذه الدراسة على المواقع السعودية، ذلك لكون المملكة مقبلة على مرحلة جديدة من التعاملات الإلكترونية متمثلة بالحكومة الإلكترونية والتجارة الإلكترونية وأيضاً انضمام المملكة العربية السعودية مؤخراً لمنظمة التجارة الدولية سيجعل مواقع مؤسساتها الحكومية والأهلية محط اهتمام دول العالم.

وعلى ضوء ما سبق سنتناول هذه الورقة في أجزاءها الأربع ما يلي: في الجزء الأول سنتحدث عن ماهو ترميز اليونيكود وأهميته، في الجزء الثاني سنوضح المغزى من هذه الدراسة وفرضيات البحث. أما في الجزء الثالث فسنبين الإجراءات التي قمنا بها لعمل الدراسة مع توضيح للعينة المستخدمة. وأخيراً سنختم هذه الدراسة بمناقشة النتائج مع ذكر بعضاً من التوصيات.

ترميز اليونيكود: ماهيته وأهميته

تعرف موسوعة ويكيبيديا الترميز (Encoding) على أنه "رموز تعمل على ربط مجموعة من الحروف بتمثيل رقمي لها، بغرض تخزينها وإمكانية تبادلها".

وتختلف أنواع الترميز حسب نظام الحاسب المستخدم، فهناك ترميز الآسكي (ASCII) والذي ظهر مع بداية ظهور الحاسبات عام ١٩٦٢م، بعدها ظهر ترميز منظمة ISO والتي يرمز لسلسلة ترميزاتها بالرمز (ISO ٨٨٥٩) ومن ثم ترميزات شركة مايكروسوفت في نظامها الشهير ويندوز، وأخيراً ظهور الرمز العالمي الموحد (اليونيكود).

يعرف الرمز العالمي الموحد على أنه " ... 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).

وتعرف مواصفة اليونيكود صيغتين للترميز تمثلان هياكل التحويل في مواصفة (ISO-١٠٦٤٦) وهما UTF-8 و UTF-١٦ (Needleman, ٢٠٠٠). حيث تتميز صيغة UTF-8 بأنها تمثل بثمانية بت (Bit) وتصبح قابلة للتوسع لتمثيل حروف أكبر حجماً عند الحاجة إليها، بينما صيغة UTF-١٦ فتأخذ ١٦ بتاً لتمثيلها. والصيغة الأكثر انتشاراً بينهما هي صيغة UTF-8.

لا يقتصر نفع ترميز اليونيكود على مواقع الإنترنت، بل أيضاً يمتد ليشمل مجال البرمجيات الحرة ومفتوحة المصدر، حيث وفر دعم أنظمة التشغيل المفتوحة لليونيكود إمكانية تعريب البرامج فور إصدار نسخة من النظام (طويلة، ٢٠٠٥).

ويعتبر ترميز اليونيكود أحد أهم الترميزات المستخدمة لتمثيل الحروف المكتوبة، وتكمن أهمية استخدامه في التالي:

# ترميز اللغة العربية في مواقع الإنترنت: دراسة حالة للمواقع السعودية

هند بنت سليمان الخليفة

جامعة ساوثهمبتون، ساوثهمبتون، المملكة المتحدة

الكلمات المفتاحية: يونيكود، مواقع سعودية، ترميز عالمي، UTF-8، لغة عربية.

8

**ملخص:** استهدفت هذه الدراسة القيام بتحليل عدد من المواقع السعودية لمعرفة مدى انتشار استخدام الرمز العالمي الموحد (اليونيكود) في عرض محتوياتها، كما استهدفت أيضاً إلى معرفة العلاقة بين استخدام أنواع معينة من الترميز (Encoding) ونوع الخادمتان النسيجية (web servers).

وقد تم اختيار أكثر من ١٢٠ موقعاً سعودياً كعينة للدراسة، وتم استخلاص المعلومات المطلوبة من الموقع وهي: الترميز ونوع الخادم النسيجي، ومن ثم تحليلها باستخدام اختبار مربع كاي للاستقلالية (Chi Square Test of Independence) لإيجاد الارتباط بين المتغيرات محل الدراسة.

وقد أظهرت النتائج أن ١٠٪ فقط من المواقع السعودية تستخدم ترميز اليونيكود، كما أوضحت الدراسة أنه هناك ارتباط جوهري بين نوع الترميز المستخدم ونوع الخادم النسيجي.

## مقدمة

تمثل الحقبة التي نعيشها مع وجود شبكة الإنترنت واحدة من أهم العصور المعلوماتية التي مرت على بني البشر. وتتجلى هذه الأهمية في المعلومات الهائلة المتجددة على الشبكة العنكبوتية والمتاحة للعموم.

وفي آخر إحصائية لموقع Global Research لعام ٢٠٠٤ عن نسبة انتشار اللغات المستخدمة في مواقع الإنترنت، أوضحت الإحصائية أن ٢٥٪ تقريباً من المعلومات الموجودة على الشبكة العنكبوتية هي باللغة الانجليزية مقابل ٤، ١٪ للغة العربية.

وقد نشرت منظمة اليونسكو أيضاً إحصائية حديثة عن توزيع اللغات المستخدمة في الإنترنت. حيث أوضحت هذه الإحصائية أن الكتابة العربية تشكل ٩، ٨ بالمائة من إجمالي الكتابات المستخدمة على الإنترنت، كما هو موضح في الجدول ١ (Paolillo، ٢٠٠٥).

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

جدول ١ :  
إحصائية  
لتوزيع الكتابات  
المستخدمة على  
الإنترنت

وبما أن الإنترنت قد صممت في الأصل لاستخدام اللغة اللاتينية، فنجد أن عرض النص العربي قد مر بكثير من المراحل في بدايات ظهوره على الإنترنت. ففي بداية دخول الإنترنت في العالم العربي بدأت الكثير من المواقع العربية على قتلها بعرض النص العربي على شكل صورة وذلك لضمان قابلية قراءة النص لجميع المتصفحين، وأيضاً بسبب صعوبة تمثيل النص العربي وعرضه بسهولة على جميع المتصفحات، ولكن هذه الطريقة لم تكن حلاً مقنعاً آنذاك.