قضايا راهنة حول اللغة العربية والشابكة (الإنترنت)

محمد مراياتي
مستشار العلم والتكنولوجيا من أجل التنمية
الأمم المتحدة . إدارة الشؤون الاقتصادية والاجتماعية

مقدمـة :

لقد مضى على وجود الحاسوب أكثر من نصف قرن، كما مضى على وجود الشابكة (الإنترنت) أكثر من ربع قرن، استطاعت اللغة العربية فيها أن تدخل المجالات التطبيقية المختلفة لتقنية المعلومات والاتصالات مع المحافظة على خصائصها المميزة. إلا أن هناك العديد من القضايا لا تزال عالقة رغم معرفتنا بها، ورغم توصيات المؤتمرات الكثيرة بإيجاد حلول لها.

سأتطرق في هذه المقالة إلى عدد من القضايا التي لم تحل، على الرغم من مضي هذه المدة الطويلة، ولن تحل حلاً سليماً إلا بجهد عربي مشترك، كقيام مجلس وزراء الاتصالات العرب، مثلاً، باعتماد آلياتٍ محددة، وتعيين جهاتٍ مسؤولةٍ، ووضع التمويل اللازم، والخطط الزمنية لمعالجة هذه القضايا. ومن هذه الآليات قيام مختصين بالتعاون مع اتحاد مجامع اللغة العربية بتحديد القواعد المناسبة لغوياً لمراعاتها فيما سيُعتمد من حلول.

لقد آن الأوان لقيام الدول العربية رسمياً باعتماد المواصفات والمقاييس (أو المعايير) اللازمة، وكذلك التعاون في مشاريع بحث وتطوير، ما قبل التنافس Precompetitive، لتسهيل إيجاد الحلول الدائمة لقضايا اللغة العربية في التطبيقات المعلوماتية، وتمكين الشركات من إنتاج هذه التطبيقات وتسويقها.

من هذه القضايا العالقة ما يلي: توحيد مواصفة ترميز المحارف العربية، واعتماد مواصفة لأسماء النطاقات أو العنونة على الشابكة بالعربية، وإيجاد ونشر مواصفة "لرومنة" الأسماء العربية (كتابة الأسماء العربية بالحروف الرومية اللاتينية وبالعكس)، واعتماد مواصفات عربية للأرشفة (للكتب، والمجلات، والوثائق، والمتاحف وغيرها)، ومحرك بحث ومتصفح عربي

يعتمدان على مواصفات وميزات اللغة العربية، وبرامج قياسية لإدارة المحتوى، وعلامات الشكل أو استعمال المصوتات القصيرة والشدَّة، والترجمة الآلية، وأسس ومنهجية وضع المصطلح العلمي، وتطوير مدقق إملائي ومحلل صرفي ومدقق نحوي، ومعاجم ومكانز عربية رقمية.

أخيراً لا بد من تنسيق البحث والتطوير ما قبل التنافسي على المستوى العربي لتخفيف التكلفة على الجميع، ولضمان التقدم والنمو العربي على الساحة العالمية من جهة، ودعم التوجه نحو البرمجيات المفتوحة المصدر الداعمة للغة العربية Open Source من جهة أخرى.

الحاجة إلى مبادرة من المجامع بالتعاون مع مجلس وزارة الاتصالات العرب والقطاع الخاص

دخلت تقنية المعلومات والاتصالات جميع نواحي الحياة، كما أصبح معروفاً لدى الجميع. فقد بات الإنسان يعيش – إضافة إلى عيشه في الفضاء الفيزيائي المعهود – في "فضاء معلوماتي" جديد يغير الكثير من نشاطاته، وممارساته الحياتية كالتعليم والصحة، والتجارة e-com، والحكم e-gov، والكتاب e- book وهكذا.

إن تطور تطبيقات المعلوماتية هو تطور سريع جداً، والفائدة من إنتاجها واستخدامها جمةٌ، وأصبحت تؤثر في النمو وفي التنمية تأثيراً كبيراً. لكن الاستجابة العربية نحو ملائمة هذه التطبيقات للغة العربية هي استجابة بطيئة. فمثلاً حتى الآن وبعد مرور أكثر من نصف قرن على دخول الحاسوب إلى حياتنا لم تفرض الدول العربية مواصفة واحدة لمحارف اللغة العربية، في حين حصل هذا للغاتٍ هي أقل انتشاراً من اللغة العربية.

تأتي اللغة العربية في المرتبة الخامسة في العالم من حيث عدد المتكلمين الأصليين بها أو الأصليين والثانويين، فهي قبل الفرنسية والألمانية واليابانية والإيطالية، بَلْهَ الكورية وغيرها. ومع ذلك، فقد اعتُمدت مواصفات محارف هذه اللغات وفرضت رسمياً، ولم يتحقق ذلك عربياً. وهذه الحال هي نفسها في كل المواصفات الأخرى لاستعمال اللغة العربية في جميع التطبيقات المكتوبة والمحكية.

تعد رغبة القيادات العربية في التوجه بالدول العربية نحو "مجتمع المعلومات" ونحو "الاقتصاد القائم على المعرفة"، دافعاً لاتخاذ الإجراءات عربياً لحل هذه القضايا. ويُعد مجلس وزراء الاتصالات العرب مرجعاً هاماً لوضع المواصفات المرجوة واعتمادها وفرض

تطبيقها. ولن يحصل التوجه الاقتصادي والاجتماعي المذكور عبر الاهتمام بالبنية التحتية لتقنية المعلومات والاتصالات فقط، بل لا بد من الاهتمام بانتشار المحتوى الرقمي العربي، وهذا لن يحصل على مستوى الوطن العربي دون اعتماد وفرض المواصفات المطلوبة. أما مجامع اللغة العربية، فتقدم الأسس والقواعد الواجب مراعاتها في وضع هذه المواصفات بحيث تحافظ على خصائص اللغة وميزاتها.

اعتماد مواصفة من مواصفات محارف اللغة العربية الكثيرة :

من المعلوم أن إدخال المحتوى العربي وتخزينه في قواعد المعطيات أو البيانات، ومعالجته أو تناقله عبر الشابكة أو الجوال أو التلفزيون، كل هذا، يتطلب ترميزاً لمحارف اللغة.

وقد وُضعت – لهذه الغاية – في الاستعمال مواصفات يربو عددها على العشرين!! مثل ASMO 449 للترميز بسبع خانات (وهي تقابل مواصفة ASCll للغة الإنكليزية)، و Windows 1256 أو MSCP-1256 الذي اعتمدته شركة مايكروسوفت في التسعينيات، و ISO-8859-6 الذي أُقِرّ عام **1986**م وكان لي شرف ترأس الوفد العربي لاجتماع ISO في ستوكهولم الذي أقرت فيه هذه المواصفة (ثمانية خانات)، و DOS 864. ثم اعتمد ترميز المحارف العالمي Unicode universal character set (UTF16 أو UTF8).

ولكن هناك إشكالات كثيرة في استعمال هذه المواصفات، بسبب عدم توحيدها. ومن الضروري التحرك لفرض استعمالٍ محدد لها في مختلف التطبيقات على مستوى الوطن العربي. وقد نوقش هذا الموضوع كثيراً، كما هي الحال في مواصفة لوحة المفاتيح، ولا حاجة هنا للإعادة، بل المهم الآن هو اتخاذ موقف عربي موحد وفرضه رسمياً .

مواصفة لأسماء النطاقات أو العنونة العربية لمواقع الشابكة :

إن عنوان أي موقع على الشابكة هو رقمي، إلا أنه جرى اعتماد نظام للتسمية بحروف اللغة الإنكليزية قائم على استعمال ترميز المحارف بالمواصفة ASCll. والسبب في استعمال حروف اللغة الإنكليزية – وليس الأرقام – هو تسهيل حفظ عنوان الموقع، وإعطاؤه دلالة، وضمان سرعة نشر العنوان في المجتمع. ولتحقيق ذلك في المجتمع العربي ونشر التعامل مع الشابكة وتطبيقاتها المختلفة (http. Telenet. ftp…) من قبل الذين لا يتقنون الإنكليزية (وهم الأكثرية)، ونظراً لعدم قدرة نظام أسماء النطاقات الحالي – القائم على ASCll – على تمثيل المحارف العربية، ونظراً لازدياد الطلب على استعمال نظام النطاقات العربية (-A

IDN)، لا بد من اعتماد أسماء النطاقات أو عناوين المواقع باللغة العربية، ومثلها كتابة عنوان البريد الإلكتروني. إن ترك القرار لعوامل السوق كما هو الحال الآن سيؤدي إلى تقديم حلول ذات إشكالات مثل :

- تشتت المواقع العربية على الشابكة، كل في شبكه تعتمد نظام عنونة خاص بها.
- تكرار التسجيل لكل فئة من فئات المواقع.
- عدم مراعاة خصائص اللغة العربية ومميزاتها.
- عدم اعتراف ICANN أو IETF ببعض هذه الحلول أو كلها.

لهذا فإن قيام مجلس وزراء الاتصالات العرب بأخذ المبادرة والتنسيق مع كل الجهات العاملة في هذا المجال (ESCWA , IANA, MINC و AINC و ITU و KACST-Saudi NIC) هو الإجراء المطلوب. وكذلك فإن التداول مع المختصين باللغة العربية عبر مجامع اللغة مهم أيضاً بهدف اعتماد أفضل الحلول فيما يتعلق بمسائل مثل :

1. أفضل ترميز للمحارف العربية (مقترح Unicode).
2. أسماء النطاقات العليا بالعربية أي gTLDs و ccTLDs names (مقترح Unicode مع ما يقابله من Puny-Code).

ومن القضايا اللغوية التي يجب اتخاذ قرار في طريقة استعمالها في العناوين ما يلي :

أولاً : فيما يتعلق بترميز المحارف :

1. الشكل (الحركات)، هل يستعمل أم لا ؟.
2. الكشيدة (استطالة المحرف)، هل يسمح باستعمالها أم لا ؟.
3. التاء المربوطة، والهاء، هل يفرق بينهما ؟.
4. الهمزة كيف ترمز.
5. التمييز بين الألف المقصورة والياء.
6. الأرقام.
7. التمييز بين الصفر والنقطة.
8. دمج الكلمات.

الإطار رقم –1 –

مجمع . اللغة . العربية . سورية

الجامعة . السورية. سورية

موقع. السعودية

موقع. امارات

موقع. قطر

أرامكو . السعودية

أرامكو . شركة

الأمم . المتحدة. منظمة

الجمعية . العربية. شبكة

9. الفراغ، هل استعماله إلزامي بين الكلمات ؟
10. إقحام حروف لاتينية ضمن العنوان العربي، هل يسمح به ؟
11. استعمال بعض المحارف الخاصة.

ثانياً : فيما يتعلق بأسماء النطاقات العليا :

1. ما هي المبادئ التي ينتقى على أساسها كل من gTLD و ccTLD العربي.
2. لائحة بـ gTLD العربية وأخرى لـ ccTLD .

أمثلة لأحد الحلول المطروحة :

يبين الإطار رقم –1– أمثلة لإحدى المواصفات المقترحة للعناوين العربية على الشابكة.

الجهود القائمة في هذا المجال : ثمة العديد من الجهات التي تعمل للوصول إلى مواصفة عربية موحدة، ومن هذه الجهات:

1. مجموعة العمل العربية لدى MINC. ائتلاف أسماء الإنترنت المتعددة اللغات.
2. ائتلاف أسماء الإنترنت العربية AINC الذي بدأ عام **2001**م
3. فريق عمل أسماء النطاقات العربية ADNTF الذي بدأ عام **2003**م برعاية الإسكوا. وقد نجح بالتعاون مع MINC بإصدار RFC لدعم اللغة العربية لاستعمالها في أسماء النطاقات.
4. مجموعة عمل دول مجلس التعاون التي قامت برعاية من لجنة ITC لدى أمانة المجلس، وقد أطلقت مشروع التجربة العملية لأسماء النطاقات العربية.
5. فريق عمل أسماء النطاقات العربية التابع للجامعة العربية، الذي تبنى توسيع تجربة دول مجلس التعاون.

والمأمول الآن أن يسارع مجلس وزراء الاتصالات العرب بالتعاون مع القطاع الخاص إلى اتخاذ توجّهٍ لإقرار مواصفة عربية محددة.

مواصفة لكتابة الأسماء العربية بالأحرف اللاتينية وبالعكس :

توجد عدة تطبيقات نحتاج فيها إلى كتابة الأسماء أو الجمل العربية بالأحرف الأجنبية كالإنكليزية أو الفرنسية أو الألمانية وبالعكس. من هذه التطبيقات البحث في الشابكة عن أسماء أو مواضيع عربية، أو نقل معلومات عربية إلى أجانب لا يعرفون الحروف العربية، أو

ترجمة المخطوطات العربية إلى اللغات الأجنبية. أو أرشفة المحتوى الرقمي العربي في المكتبات والأرشيف العالمي، أو التداول بأسماءِ العَلم العربية في وسائل الإعلام الرقمية وخاصة الشابكة والهاتف الخلوي، وأخيراً – ومع الأسف – إرسال الرسائل العربية القصيرة SMS على الجوال أو الشابكة وما يسمى أيضاً بالدردشة. [انظر الجدول رقم 1]. ونقول مع الأسف لأن SMS والدردشة يمكن – بل يجب – أن تكون بالحروف العربية لو أن الدول العربية سارعت إلى فرض مواصفة لها متفق عليها.

هناك نوعان لتحويل الحروف من العربية إلى الأجنبية "الرومنة" وبالعكس "التعريب". النوع الأول تحويل المكتوب Transliteration، والنوع الثاني تحويل المنطوق Transcription أو التحويل حسب التهجي الصوتي. وفي كثير من الأحيان يدمج النوعان معاً في تحويل الحروف العربية إلى الأجنبية. والفرق بين النوعين يعود إلى أن الكتابة في العربية تهمل وضع المصوتات (الحركات) على الحروف في الأعم الأغلب، لذلك فإن تحويل النص حسب الحروف المكتوبة دون الشكل (الحركات) لا يؤدي الغرض. أما النوع الثاني فيحول الكلمات من العربية إلى الأجنبية لمنطوق الكلمة وليس لمكتوبها.

الجدول رقم .1 .

"الرومنة" المستعملة في الرسائل القصيرة (SMS) أو في "الدردشة"

Stand-alone	Initial	Medial	Final	Arabic Chat Alphabet	Phonetic Value (IPA)
ء	أ ؤ إ ئ ا ى, etc.			2	[ʔ]
ا	—		ـا	a	various, including [aː]
ب	بـ	ـبـ	ـب	b	[b]
ت	تـ	ـتـ	ـت	t	[t]
ث	ثـ	ـثـ	ـث	s / th	[θ]
ج	جـ	ـجـ	ـج	g / j	[ʤ] / [g]
ح	حـ	ـحـ	ـح	7	[ħ]
خ	خـ	ـخـ	ـخ	5 / 7' / kh	[x]
د	—		ـد	d	[d]
ذ	—		ـذ	z / th	[ð]
ر	—		ـر	r	[r]
ز	—		ـز	z	[z]
س	سـ	ـسـ	ـس	s	[s]
ش	شـ	ـشـ	ـش	sh	[ʃ]
ص	صـ	ـصـ	ـص	S / 9	[sˁ]
ض	ضـ	ـضـ	ـض	D / 9'	[dˁ]
ط	طـ	ـطـ	ـط	TH / 6	[tˁ]
ظ	ظـ	ـظـ	ـظ	Z / TH / 6'	[ðˁ] / [zˁ]
ع	عـ	ـعـ	ـع	3	[ʕ] / [ʔˁ]
غ	غـ	ـغـ	ـغ	gh / 3'	[ɣ] / [ʁ]
ف	فـ	ـفـ	ـف	f / ph	[f]
ق	قـ	ـقـ	ـق	q / 8	[q]
ك	كـ	ـكـ	ـك	k	[k]
ل	لـ	ـلـ	ـل	l	[l], [lˁ] (in *Allah* only)
م	مـ	ـمـ	ـم	m	[m]
ن	نـ	ـنـ	ـن	n	[n]
ه	هـ	ـهـ	ـه	h	[h]
و	—		ـو	w	[w] , [uː]
ي	يـ	ـيـ	ـي	i / y	[j] , [iː]

ونظراً لغياب مواصفة عربية لتحويل الحروف المكتوبة والمنطوقة، فإن العديد من الجهات في العالم اتخذت مواصفة خاصة بها. لذلك توجد الآن عشرات المواصفات ولا نزال نشهد مقترحات لمواصفات جديدة ! ! من أشهر هذه المواصفات ما يلي وقد أوردنا بعضاً منها في الجدول رقم (2):

1 . المواصفة العالمية ISO/R 233 الصادرة أول مرة عن ISO عام **1961**م، والمعدلة عام **1984**م لتصبح ISO 233، والمبسطة عام **1993**م لتصبح المواصفة ISO233-2.

2 . المواصفة التي اعتمدها المستشرقون عام **1936**م، وقاموس هانس وهر (Hans wehr) .

3 . المواصفة التي اعتمدها اتحاد المكتبات الأمريكية ALA-LC.

4 . المواصفة التي اعتمدها خبراء عرب في اجتماع بيروت عام **1971**م. وقد أثرت اللغة الإنكليزية في هذه المواصفة أكثر من الفرنسية؛ فاعتُمد مثلاً Sh الإنكليزية مقابلاً للحرف "ش" عوضاً عن Ch الفرنسية.

5 . المواصفة الإنكليزية BS4280 الموضوعة عام **1968**م.

6 . مواصفة الأمم المتحدة (للأسماء الجغرافية) UNGEGN.

7 . مواصفة شركة Xerox (وهي Transliteration)، وقد طوّرتها Buck walter عام **1990**م.

8 . المواصفة العسكرية الأمريكية SATTS المعتمدة عام **1970**م.

9 . المواصفة الألمانية DIN 31635 المعتمدة عام **1982**م.

10. مواصفة قلم Qalam المطروحة عام **1985**م ، والمبنية على التهجي أكثر من اللفظ.

11. المواصفة المعتمدة من المستشرقين الأسبان SAS.

12. مواصفة الموسوعة الإسلامية SM.

الإطار رقم . 2 .

Using the Google search engine, we checked for the most common spellings of some Arabic names ...

Muhammad	41%
Mohammed	32%
Mohamed	25%
Mahomet	3%
Gaddafi	72%
Qadhafi	16%
Gadafi	8%
Gadafy	2%
Qadhdhafi	1%
Quran	44%
Koran	37%
Qur'an	19%
al-Qaeda	61%
al-Qaida	36%
al-Qa'ida	2%
al-Qa'eda	1%
Mecca	85%
Makkah	14%
Mekkah	1%

ويدعي واضعو هذه المواصفات جميعها أنها وُضعتْ لعدم وجود مواصفة عربية معتمدة ومفروضة من الدول العربية. ولا شك في أن تعدد هذه المواصفات يؤدي إلى إشكالات في الشابكة، وفي البحث في قواعد المعطيات، وفي الإعلام.

الجدول رقم .2

بعض المواصفات المستعملة لكتابة الحروف العربية بمحارف أجنبية

Letter	Name	SATTS	UNGEGN	ALA-LC	DIN-31635	ISO 233	ISO/R 233	Qalam	SAS	SM	IPA
ء	hamza	E	’, —	—, ’	’	’, ‘	—, ’	’	’ (zero word-initially)	’ (disappears after ’al-’ and where alif waṣl is.	[ʔ]
ا	’alif	A			ā	’	ā	aa	a, i, u (syllable-initial) ā (lengthening)	aa	various, including [aː]
ب	bāʾ	B	b	b	b	b	b	b	b	b	[b]
ت	tāʾ	T	t	t	t	t	t	t	t	t	[t]
ث	ṯāʾ	C	th	th	ṯ	ṯ	ṯ	th	ṯ	ç	[θ]
ج	ǧīm, jīm, gīm	J	j	j	ǧ	ǧ	ǧ	j	ŷ	j	[ʤ] / [ʒ] / [g] / [j]
ح	ḥāʾ	H	ḥ	ḥ	ḥ	ḥ	ḥ	H	ḥ	ḥ	[ħ]
خ	ḫāʾ	O	kh	kh	ḫ	ḫ	ḫ	kh	j	x	[x]
د	dāl	D	d	d	d	d	d	d	d	d	[d]
ذ	ḏāl	Z	dh	dh	ḏ	ḏ	ḏ	dh	ḏ	đ	[ð]
ر	rāʾ	R	r	r	r	r	r	r	r	r	[r]
ز	zāy	;	z	z	z	z	z	z	z	z	[z]
س	sīn	S	s	s	s	s	s	s	s	s	[s]
ش	šīn	:	sh	sh	š	š	š	sh	š	ś	[ʃ]
ص	ṣād	X	ṣ	ṣ	ṣ	ṣ	ṣ	S	ṣ	ṣ	[sˁ]
ض	ḍād	V	ḍ	ḍ	ḍ	ḍ	ḍ	D	ḍ	ḍ	[dˁ]
ط	tāʾ	U	t	t	t	t	t	T	t	t	[tˁ]
ظ	ẓāʾ	Y	ẓ	ẓ	ẓ	ẓ	ẓ	Z	ẓ	ḍ	[ðˁ] / [zˁ]
ع	ʿayn	`	‘	‘	‘	‘	‘	`	‘	ř	[ʕ] / [ʔˁ]
غ	ġayn	G	gh	gh	ġ	ġ	ḡ	gh	g	ğ	[ɣ] / [ʁ]
ف	fāʾ	F	f	f	f	f	f	f	f	f	[f]
ق	qāf	Q	q	q	q	q	q	q	q	q	[q]
ك	kāf	K	k	k	k	k	k	k	k	k	[k]
ل	lām	L	l	l	l	l	l	l	l	l	[l], [lˁ] (in *Allah* only)
م	mīm	M	m	m	m	m	m	m	m	m	[m]
ن	nūn	N	n	n	n	n	n	n	n	n	[n]
ه	hāʾ	~	h	h	h	h	h	h	h	h	[h]
و	wāw	W	w	w	w	w	w	w	w (consonantal) ū (lengthening)	w (consonantal) o (lengthening)	[w] , [uː]
ي	yāʾ	I	y	y	y	y	y	y	y (consonantal) ī (lengthening)	y (consonantal) e (lengthening)	[j] , [iː]
آ	’alif mamdūda	AEA	ā	ā, ’ā	’ā	’å	ā, ’ā		ā	’aa	[ʔaː]
ة	tāʾ marbūṭa	@	h, t	h, t	h, t	ẗ	h, t	h, t	t (zero when in absolute state)	t	[a], [at]
ى	’alif maqṣūra	/	y	y	ā	ŷ		ae	à	à	[aː]
لا	lām ’alif	LA	lā	lā	lā	la’	lā	la	l’ (with hamza) lā (with lengthening alif)	treated as laam then alif usually: laa	[laː]
ال	’alif lām	AL	al-	al-	al-	’’al	al-	al	al-	al- When assimilation occurs: ál-	

فمثلاً هناك من عدَّ 32 طريقة "رومنة" لكتابة اسم الرئيس القذافي في الإعلام العالمي (أي 32 طريقة كتب بها الاسم بالحروف الأجنبية). يبين الإطار رقم . 2 . أمثلة حول رومنة بعض الأسماء الهامة.

مواصفات الأرشفة :

نتيجة للتوجه نحو مجتمع المعلومات والاقتصاد القائم على المعرفة، عظم دور اللغة بوصفها وعاء للمعرفة، وعظم دور الأرشيف الذي يمثل تراكم المعرفة وخاصة العلمية والتقنية. ومع وجود الشابكة والهاتف الخلوي أصبح النفاذ إلى الأرشيف متيسراً ومبذولاً عالمياً، من أي مكان وفي أي وقت. وقد جرى تطوير نظم لتحقيق ذلك منها "نظام أرشيف المعلومات المفتوح" .Open Archival Information System (OAIS)

وأصبح النفاذ إلى المكتبات، وإلى قواعد معلومات الوثائق المختلفة، والمخطوطات، وإلى قواعد المعلومات العلمية والتقنية في متناول الباحثين والطلاب والفعاليات الاقتصادية والعامة. وحتى يكون ذلك ناجعاً باللغة العربية لا بد من الاتفاق عربياً على اعتماد بعض المواصفات وفرضها في التطبيق. فمن شأن ذلك أن يجعل النفاذ إلى كل المكتبات العربية الرقمية ممكناً وغير مكلف من أي مكان في الوطن العربي. ومن شأنه أيضاً أن يجعل عملية الرقمنة وعملية إدارة المكتبات والوثائق أقل تكلفة على الجميع وطنياً وعربياً. فمثلاً إذا جرى اعتماد مواصفة موحدة للفهرسة Centralized Union Catalogue، فإن إدخال الكتاب أو المرجع أو الوثيقة سينفذ مرة واحدة وطنياً، وربما عربياً!! بدل إدخاله عشرات المرات وبتوصيف مختلف (Metadata) في كل مرة !!. وهكذا فإن اعتماد مواصفات مثل MARC في معلومات التوصيف (Format for metadata) ومواصفات الترميز والتعرّف Coding, Identification وما يقابل MASTER لتوصيف المخطوطات، وغيرها من المواصفات (مثل ISAD , SPECTRUM, DUBLIN Core , ...) أصبح غاية في الأهمية عربياً.

يخضع العمل الأرشيفي حالياً لتغيرات جذرية، ويشهد دخول مجالات لم تكن موجودة قبلاً. وقد أصبحت هذه التغيرات هامة وملحة لكي تجاري متطلبات تحول المجتمع في الدول العربية نحو "مجتمع المعلومات"، وكذلك تحول الاقتصاد نحو "الاقتصاد المبني على المعرفة". وهذا بدوره يجعل هذه التغيرات ذات تأثير اقتصادي واجتماعي، وذات قيمة مضافة كبيرة.

إن سبب هذه التغيرات هو – طبعاً – التطور الذي حدث في تقانة المعلومات والاتصالات، والذي أدى إلى ترميز الوثائق والمعلومات بكل أنواعها، كنص أو صورة أو مخطط أو فلم أو كلام منطوق، بالترميز الرقمي الذي وحّدته "البت". وقد رافق هذا التطور تطور كبير في خوارزميات معالجة الوثائق والمعلومات بكل أنواعها، وفي بروتوكولات نقلها على الشبكات السلكية واللاسلكية.

نورد فيما يلي بعض الأمثلة عن هذه المجالات الجديدة في العمل الأرشيفي :

1- رقمنة الأرشيف السابق (going digital) لدى كل الجهات وفي كل المجالات.
2- أرشفة المعلومات الرقمية الحالية digital archiving : أدواتها ووسائطها.
3- التطور نحو البحث الآلي في الأرشيف، ليس على مستوى الكلمة أو الجملة فقط، ولكن على مستوى الدلالة أو المعنى.
4- الأرشفة والبحث عبر النصوص المترابطة Hyper text وآلياتهما.
5- الأرشفة المتعددة الوسائط multimedia.
6- إدارة المعرفة Knowledge Management، والتنقيب في النصوص Text Mining.
7- المسائل الجديدة في أمن الأرشيف
8- الطرق والآليات والوسائط الجديدة في حفظ أو خزن الأرشيف والتغير الجذري في طبيعة مستودعات الأرشيف وشروطها.
9- تكنولوجيات جديدة في صيانة الأرشيف وعمره.
10- النظم البرمجية الجديدة كالأرشفة (مفتوحة – مملوكة، المعايير الدولية الجديدة).
11- تغيرات في حقوق الملكية والتعامل معها.

إن هذه التغيرات الجذرية في حقل الأرشفة وتأثيراتها الاقتصادية والاجتماعية والثقافية، واتساع حقل هذه التأثيرات ليشمل كل فعاليات الإنتاج والخدمات في المجتمع؛ إضافة إلى الفعاليات الثقافية والتعليمية والاجتماعية، يتطلب قيام العديد من الدول ومن مؤسساتها بوضع سياسات وخطط لرقمنة الأرشيف وتعليمه ووضع المواصفات الناظمة للعمل فيه.

محرك بحث ومتصفح عربيان:

متصفح أو مستعرض الشابكة Browser **هو تطبيق برمجي يمكِّن المستخدمَ من عرض نصوص أو صور أو أفلام موجودة في صفحة من صفحات موقع على الشابكة، كما يمكّنه من التعامل مع هذه المعلومات. أهم المتصفحات المتوفرة في الأسواق:** internet explorer **من شركة مايكروسوفت، ثم** Mozilla Firefox**، ثم** Safari **من شركة** Apple**، ثم** Navigater **وتعديلاته من شركة** Netscape**، ثم** Opera.

تختلف المتصفحات بعضها عن بعض في دعمها للغات العالمية وفق مواصفات محددة. ومن الجدير بالذكر وجود العديد من المتصفحات المفتوحة المصدر مثل Netscape Mozilla. **وتتوفر حالياً كثير من المتصفحات تدعم عرض اللغة العربية ضمن أنظمة التشغيل مثل:**

آ- Explorer **لمايكروسوفت، و** Netscape **على نظام** Windows **سواء أكان نظام التشغيل يدعم اللغة العربية أم لا. وهناك عدة خيارات لمجموعات المحارف العربية مثل :** Windows 1256 **و** ASMO 708 **و** ISO-8859-6 **و** DOS 869 **و** UTF 8

ب. المتصفح Opera **يدعم اللغة العربية على كل أنظمة التشغيل** Windows **و** Linux **و** UNIX **و** MAC.

ج. المتصفح Safari **من شركة "آبل" على حواسيب ماكنتوش يدعم اللغة العربية.**

د. المتصفح Mozilla **على نظام التشغيل** Linux **يدعم اللغة العربية.**

أما محرك البحث (Search engine) **فهو برنامج خدمة للحصول على معلومات مطلوبة بالبحث عنها في حاسوب، أو في شبكة حواسيب كالشابكة. يقوم المحرك بثلاث مهام هي: مسح مواقع الشابكة، ثم تجميع ومواءمة أو تحيين فهارس بالمعلومات الموجودة على الشابكة، ثم تلبية طلب المستخدم بإجراء البحث. وعادة ما يطلب من المحرك البحث عن كلمة، أو عدة كلمات، أو شبه جملة، أو جملة باللغة العربية.**

أهم محركات البحث الموجودة حالياً في السوق العالمية هي حسب معدل الانتشار عام 2006: Google **ثم** Yahoo **ثم** MSN Search **من شركة مايكروسوفت، ثم** Ask.Com **وغيرها. من جهة أخرى يوجد أيضاً محركات بحث مفتوحة المصدر (مثل** ht://Dig, Nutch, Senas **و ...)**

وحتى يكون البحث باللغة العربية شاملاً، لا بد من القيام بتحليل صرفي سواءً في تصميم الفهارس، أو في تحديد الكلمات التي يبحث عنها. وإن عدم وجود محرك بحث عربي متقدم في استعماله للتحليل الصرفي يقلل من فاعلية وجدوى استخدام الشابكة، ويحرم المستخدم من الوصول إلى الكثير من المراجع، كما أنه يطيل الزمن اللازم لإجراء البحث المطلوب والحصول على المحتوى المنشود.

هناك العديد من مراكز البحوث قامت بدراسات على التحليل والتركيب الصرفي باستعمال الحاسوب، لكن الشركة الرائدة في تطبيقه في مجال محركات البحث هي شركة صخر، إلا أنها تستعمله في منتجاتها (موقع عجيب Ajeeb.Com، وموقع al-islam.com ,...) وهو غير متاح عامة للبحث على الشابكة.

وقد أُعلن عن قيام مشروع تطوير محرك بحث بالتعاون بين شركة Mitsco السعودية وشركة Seekport الألمانية ومقر العمل في دبي، ومحرك البحث يسمى صوافي "Sawafi" ومن المتوقع طرحه في الأسواق نهاية عام **2006**م أو بداية **2007**م ولكن من غير المعروف بعد مدى فعالية استعمال "صوافي" للتحليل الصرفي في تصميمه. من جهة أخرى فإن تطوير محرك بحث عربي هو من ضمن المشاريع المطروحة من قبل كل من ESCWA والجامعة العربية في لائحة خطة العمل المعتمدة في "القمة العالمية لمجتمع المعلومات" الذي عقد في تونس **2005**م.

إدارة المحتوى :

من أجل تنمية المحتوى الرقمي العربي، برقمنة القديم وإدخال الجديد رقمياً، لا بد من العمل على حسن إدارة هذا المحتوى. إن نظم إدارة المحتوى Content Management System (CMS) هامة كي ينتشر استعمال الشابكة. وإن توفر هذه النظم باللغة العربية، وبأسعارٍ معقولة، وعملٍ فعال صحيح، هو مطلبٌ رئيسي. والمرجو قيام مراكز البحث والباحثين في الجامعات بدراسة أفضل الحلول لاستعمال اللغة العربية استعمالاً سليماً في هذه النظم. وإذا ما اتخذ وزراء الاتصالات العرب قراراً بالطلب من هذه المراكز والجامعات العربية القيام بذلك (كما فعل مجلس وزراء الخارجية العرب في دورته الأخيرة في 7 أيلول **2006**م في القاهرة في موضوع الدراسات والبحث في الطاقة النووية). فمن المأمول أن يستجاب لهذا القرار وأن تدعمه القيادات العليا والقطاع الخاص.

من نظم ادارة المحتوى :

- **إدارة محتوى الشابكة.**
- **إدارة المداولات في التجارة الإلكترونية.**
- **نظم إدارة المحتوى المتكاملة بغية إدارة معلومات المؤسسات والشركات والمنظمات المختلفة.**
- **إدارة دورة حياة المعلومات المتعددة الوسائط.**
- **إدارة النشر الرقمي.**
- **إدارة المحتوى التعليمي.**
- **إدارة الصور والأرشيف المصور.**
- **.....**

المصطلح على الشابكة وفي مجتمع المعلومات :

مع بداية ثمانينيات القرن الماضي تسارَع تطوير إدارة المصطلح أفقياً وعمودياً. ومن أهداف إدارة المصطلح:

- **وضعُ سياسةٍ لإدارة المصطلح على المستوياتِ الوطنيةِ والمؤسساتية، وحتى في شركاتِ الإنتاج والخِدْمات؛**
- **تنفيذُ هذه السياسة باتخاذ إجراءاتٍ عمليةٍ قانونية وإدارية ومالية وبشرية؛**
- **تنشيطُ التعاون والتنسيق الإقليمي والعالمي في مختلِفِ نشاطات إدارة المصطلح؛**
- **إيجادُ بنيةٍ أفقيةٍ وعمودية لإدارة المصطلح وتطويرُها؛**

أما البنيةُ الأفقية لإدارة المصطلح فتشتمل على مؤسساتٍ عامةٍ أو شبهِ عامة بغية:

1- دعمِ النشاطاتِ المصطلحية وتنظيمِها وتنسيقها لخدمة جميع أفراد المجتمع؛

2- نشرِ المعلومات المتعلقة بالمؤسسات العاملة في حقول المصطلح، وبالمنشورات والخِدْمات المتوفرة فيه؛

3- دعمِ وتنسيق الجهود الوطنية لإيجاد الحلول للمسائل المصطلحية؛

وأما البنيةُ العمودية لإدارة المصطلح فتشتمل على هياكلَ تنظيميةٍ عامةٍ وخاصة ترعى مجموعاتِ العملِ المتخصصةَ في كل حقلٍ من حقول المعرفة. وهذه المجموعاتُ تقوم

بتنظيم المصطلحات وتنسيقِها واعتمادِها معيارياً. ويوجد حالياً أكثرُ من مئة مجموعةِ عملٍ في كلِّ لغة من اللغات العالمية. تختص هذه المجموعاتُ بوضع المصطلحات الرسمية، ويصل هذا العدد إلى أكثرَ من خمسةِ آلافِ مجموعةِ عملٍ في العالم حالياً في كل اللغات. أما إذا أخذنا بالحسبان جميع مجموعات العمل واللجان التي تعمل في شؤون المصطلح، فإن هذا العدد سيصل إلى أكثرَ من خمسينَ ألفِ مجموعةٍ أو لجنة؛

وترعى البنيةُ العموديةُ أيضاً المعاجمَ واللوائحَ المعيارية للمصطلحات. وهذه المعاجمُ واللوائح منها ما يَصدر في الجريدة الرسمية في الدولة، ومنها ما يَصدر عن مؤسساتٍ ذات سلطة رسمية، أو تتمتع بسلطة معنوية، أو تتمتع بسلطة في مجال التنسيق؛ وتختلف هذه المعاجمُ واللوائحُ بحسب مجموعة العمل التي قامت بوضعها. فمن وجوه الاختلاف:

- المنهجيةُ المعتمَدة: (هل هي رسمية: كالمنهجية الموضوعة من قبل المنظمة الدولية للتقييس ISO؛ أو غير رسمية؟)
- ومن وجوه الاختلاف: توثيقُ العمل: (مصادرُه، والمراجعُ المعتمَدة، والخبراءُ العاملون فيه... الخ)؛
- ومن وجوه الاختلاف: الجهةُ القائمة بالعمل: (شخصٌ أو مجموعةٌ...)، ومستوى معرفتهم بالاختصاص أو الحقل، ومعرفتِهم بعلم المصطلح؛
- ومن وجوه الاختلاف أيضاً: تطبيقُ معايير ضمان الجودة وإدارةِ الجودة في وضع هذه المصطلحات.

وتنقسم المعايير في حقل المصطلح إلى نوعين:

يشتمل النوعُ الأول على معاييرِ أسسِ وطرائقِ ومنهجياتِ وضْعِ المصطلح وتنسيقِه ونشره. ويَقود هذا التقييسَ عالمياً لجنةُ ISO/TC37،

أما النوع الثاني فيشتمل على معاييرَ بالمصطلحاتِ نفسِها؛ أي: على لوائحِ مصطلحاتٍ في كلِّ حقل من حقول المعرفة. وكما ذكرنا فهناك أكثرُ من خمسِ مئةِ مجموعةِ عملٍ في هذا المجال عالمياً.

ومن أمثلةِ المعايير العالمية في حقل المصطلح ما تقوم به المنظماتُ العالمية المعروفة مثل: ISO في لجنتِها المتخصصة ISO-TC37 واللجانِ الوطنية المتعاونة معها في شتى دول العالم، ومثل IEC، وITU وغيرها.

ومن الجدير بالذكر أن هناك تسعاً وسبعين **79** دولةً تشارك في اللجنة ISO-TC37، منها أربعُ دولٍ عربيةٍ بصفة مراقب هي: مصر وتونس وسورية والسعودية، ولكن حضورَها مازال ضعيفاً.

أدوات في إدارة المصطلح

لم يَعُدِ العملُ في حقل (المصطلح) و(المصطلح الجديد) ممكناً بدون استعمال الأدوات التي تقدمها تقانةُ المعلومات والاتصالات. فهذه الأدوات تعمل على تطبيقِ مبادئِ ونظرياتِ وآلياتِ المعالجةِ الآلية للغات في هذين الحقلين، ولاسيما فيما يتعلق برصدِ و"تفقد" المصطلح في النصوص، وقياسِ تواترِ وروده، وجَمْعِ تصاريفه، وكذلك مراقبةِ استعماله. من هذه الأدوات (أي أدوات إدارة المصطلح):

- النظمُ البرمجيةُ لرصد "المصطلحات الجديدة" ومراقبتِها؛
- نظمُ إدارةِ المصطلح وتصنيفِه؛
- نظمُ المساعدة في توليد المصطلح باستعمال الحاسوب؛
- النظمُ المساعدة في جمع "المصطلح الجديد" من الإنترنت باستعمال "محركات البحث"، ومجموعات الأخبار، ومنتديات النقاش المتخصصة؛
- برمجياتُ المكانز اللغوية وإدارتِها؛
- المعاجمُ الإلكترونية للمصطلحات العامة والمتخصصة؛

تساعد الأدواتُ المعلوماتيةُ في حقل المصطلح كذلك على القيام بمختلِف أنواع العمل المصطلحي؛ كالإبداعِ في توليد مصطلحاتٍ جديدة، ورصدِ المصطلحات الجديدة وتفقدِها في جميع الاختصاصات العلمية، ووصْفِ المصطلحات وشرحِها، ووضعِ المعاجم ولوائحِ المصطلحات الاختصاصية، وتنسيقِ المصطلحات وتقييسِها.

علامات الشكل :

تعد الحركات الثلاث والشدة محارف في اللغات الأخرى وهي تكتب حكماً في تلك اللغات على شكل مصوتات أو حرف مضاعف. إن الزيادة الهائلة في وجود المحتوى الرقمي على الشابكة والدور الكبير للغة في حياتنا الحالية والمستقبلية، وضعف أفراد المجتمع العربي عامة في إتقان لغتهم، تستدعي استعمال الشكل في الكتابة العربية، ولو جزئياً، كما تستدعي إيجاد برمجيات تعالج علامات الشكل إما لوضعها آلياً، أو لتعرفها أو لمعالجتها في شتى التطبيقات. ومن هذه التطبيقات تحويل النص المكتوب إلى محكي، والتدقيق الإملائي والصرفي والنحوي، وهذا يتطلب اتفاقاً عربياً على أساسيات هذه الحالات، ومن ثم تشجيع مراكز البحوث ثم الشركات على تطوير المنتجات اللازمة لذلك وتسويقها، مع فرض استعمال المواصفات المعتمدة في ذلك.

الإطار رقم . 3 .

بعض مواقع الترجمة الآلية على الشابكة

www.almisbar.Com
targim.ajeeb.Com
www.systranet.Com
www.babylom.Com
www.worldlingo.Com
www.wordreference.Com
www.freetranstation.Com

الموقعان الأخيران لا يترجمان إلى العربية

التشكيل : Vocalization, Vowelization, Diacritization

الترجمة الآلية :

يعتمد نمو المحتوى الرقمي العربي وخاصة في المجال العلمي والتقني على الترجمة إلى اللغة العربية. إن الترجمة الآلية (أو الترجمة بمساعدة الحاسوب التي هي أقل طموحاً من الترجمة الآلية) هي الأداة الحديثة لتحقيق هذا النمو. يوجد الآن على الشابكة عدة مواقع تقدم الترجمة الآلية والآنية من الإنكليزية إلى العربية. من هذه المواقع موقع "المسبار" (almisbar) الذي أصبح يحظى بأكثر من 200 ألف زائر يومياً، وموقع "عجيب" (ajeeb) لشركة صخر الذي يترجم كل صفحات الشابكة والنصوص.

إن رفع أداء الترجمة الآلية من العربية وإليها بحاجة إلى مزيد من البحث والتطوير في معالجة اللغة العربية حاسوبياً، وخاصة في مجالات الصرف والنحو والدلالة. ومجلس وزراء الاتصالات العرب والقطاع الخاص مدعوان لتشجيع مراكز البحوث والجامعات لإقامة

معاهد أو فرق أو مخابر بحث وتطوير في هذه الحقول. كما أن هناك حاجة إلى تعميق منهجية الترجمة إلى العربية ووضع مواصفات عربية فيها. يبين الإطار رقم . 3 . المرفق أهم مواقع الترجمة على الشابكة.

دعم البحث والتطوير في قضايا معالجة اللغة العربية معلوماتياً :

مع أن العديد من معاهد البحوث والجامعات في الدول العربية وفي العالم تقوم بإجراء بحوث في مجالات معالجة اللغة العربية (NLP)، فإن كمية ونوعية هذه الأعمال لا تزال قليلة وبطيئة، ولا توازي حجم اللغة العربية عالمياً. لذلك نؤكد ضرورة إيلاء هذا المجال من البحث والتطور الاهتمام اللازم، والطلب من الجامعات ومعاهد البحوث والقطاع الخاص القيام بإحداث برامج في المجالات التي وردت في هذا البحث، وكذلك في مجالات هامة أخرى مثل: تعرف الحروف العربية OCR، والمدقق الإملائي، والمدقق الصرفي، والمدقق النحوي، والتحليل والتركيب الدلالي، وتوليد النص الآلي، والترجمة بين اللغات، وتركيب وتعرف الكلام (ASR, TTS)، والفهم الآلي للنص.

النتيجـة :

نرى أن يتولى مجلس وزراء الاتصالات العرب إحداث الهيكلية اللازمة لمتابعة وضع واعتماد هذه المواصفات، بالتعاون مع الجهات المختصة عربياً وعالمياً وبالتشاور مع الشركات المنتجة، وهي عادة مستعدة للتعاون (كما جرى سابقاً بينها وبين المنظمة العربية للمواصفات والمقاييس ASMO).

من هذه الجهات مجموعة عمل الإنترنت الهندسية (IETF)، ونقابة الإنترنت للأسماء والأرقام المخصصة (ICANN)، ومركز المواصفات والمقاييس العربية التابع لـ AIDMO المنظمة العربية للتنمية الصناعية، والمنظمة العالمية للتقييس ISO، وائتلاف أسماء الإنترنت المتعدد اللغات Multilingual Internet Names Consortium (MINC)، والائتلاف العربي لأسماء الإنترنت (AINC).

من جهة أخرى نقترح أن يقوم اتحاد مجامع اللغة العربية والمجامع نفسها، بالتعاون مع جمعيات الحاسوب المختلفة في الدول العربية، وكذلك بالتعاون مع القطاع الخاص وبالتحديد شركات الاتصالات وتقنية المعلومات، بإنشاء جمعيات تهتم بأمور "اللغة العربية والشابكة" في كل دولة من الدول العربية.

مختصرات

ADN	Arabic domain name
AIDMO	Arab Industrial Development and Mining Organization
AIDMO-CSM	AIDMO-Center of Standardization and Metrology
AIDNA	Arab Internet and Domain Name Association
AINC	Arabic Internet Names Consortium
AKMS	Arab Knowledge Management Society
ALECSO	Arab League Educational, Cultural and Scientific Organization
ALC	Arabic Linguistic Committee
ALUG	Arab Linux User Group
Arabcin	Arab Club for Information
ASIP	Arab Society for Intellectual Property
ASMO	Arab Standards and Metrology Organization
ASR	Automatic speech recognition
CCS	coded character set
ccTLD	country code top-level domain
DNS	Domain Name System
DNSO	Domain Name Supporting Organization
DOS	Disk Operating System
EAD	encoded archival description
ECMA-ATF	European Computer Manufacturing Association, Arabic Task Force
EDI	electronic data interchange
EFI	electronic form interchange
ESMTP	extended simple mail transfer protocol
ESCWA	Economic and Social Commission for Western Asia
ETSI	European Telecommunications Standards Institute
FIPS	Federal Information Processing Standards
GCC	Gulf Cooperation Council
gTLDs	generic TLDs
HTML	hypertext markup language
HTTP	hypertext transfer protocol
IANA	Internet assigned number authority
ICANN	Internet Corporation for Assigned Names and Numbers
ICT	information and communications technology
iDNS	internationalized DNS
IEC	International Electrotechnical Commission
IETF	Internet Engineering Task Force
IFAS	International Federation for the Application of Standards
IKM	information and knowledge management
ISAD	International Standard for Archive Description
IP	Internet Protocol
ISO	International Organization for Standardization
ISOC	Internet Society
ISP	Internet service provider
IT	information technology
ITU	International Telecommunication Union
LDC	Linguistic Data Consortium

MARC	machine-readable cataloguing
MASTER	Manuscript Access through STanderds for Electronic Records
MHTML	Multilingual-HTML
MIME	Multipurpose Internet mail extensions
MINC	Multilingual Internet Names Consortium
NLP	natural language processing
OAIS	Open Archival Information System
OCR	optical character recognition
OS	operating system
OSS	open source software
PC	personal computer
R&D	research and development
RFC	Requests for Comments
RITSEC	Regional Information Technology and Software Engineering Center
SaudiNIC	Saudi Network Information Center
SGML	standard generalized markup language
SMTP	simple mail transfer protocol
SPECTRUM	Museum documentation standard
TC-8	Technical Committee 8 (AIDMO-CSM)
TLD	top-level domain
TTS	text-to-speech
UCS	Universal Multiple-Octet Coded Character Set (Unicode)
URI	uniform resource identifier
URL	uniform resource locator
UTF	Unicode transfer format
WWW	World Wide Web
XML	EXtensible markup language