مؤتمر المجمع الثاني
اللغة العربية في مواجهة المخاطر
دمشق، ٢٠- ٢٣ /١٠/ ٢٠٠٣

المحتوى العربي في الشبكات الحاسوبية
دوره وأهميته

The Arabic Language Content Gap

د. محمد مراياتي
المستشار الإقليمي للعلم والتكنولوجيا
اللجنة الاقتصادية والاجتماعية لغربي آسيا
بيروت، ص.ب. ٨٥٧٥-١١
mrayati@un.org

مستخلص

تزداد أهمية اللغة العربية على الشبكات الحاسوبية المختلفة مع توجه المجتمع نحو مجتمع المعلومات وتوجه الاقتصاد نحو الاقتصاد المبني على المعرفة. يصطلح عالمياً على هذا الوجود بكلمة المحتوى (content). ويعبر المحتوى عن وجود المعرفة بشكل رقمي (digital) على الحواسيب وعلى الشبكات الداخلية (intranet) والعالمية (extranet) و الإنترنت (internet). ويغطي المحتوى مجالات شتى مثل النشر، والأعمال، والمكتبات، والإدارة الحكومية (e-gov)، والعلم والتكنولوجيا، والصحة والثقافة، والتراث، والسياحة والتسلية، ومعلومات عن المنظمات غير الحكومية والإقليمية وغير ذلك من المجالات. ويتطور المحتوى عالمياً لكل لغة من اللغات الأساسية في العالم وخاصة المحتوى الموضوع على الإنترنت (online content). وسنعرض في هذا البحث وضع محتوى هذه اللغات. تصنف المعلومات الموجودة كمحتوى للغة من اللغات في أصناف مثل معرفة ماذا، ومعرفة كيف، ومعرفة من، ومعرفة لماذا. من جهة أخرى يمر المحتوى بمراحل محددة في علمية إيجاده وهي مراحل توليده، ثم تحويله إلى الشكل الرقمي، وحيازته وخزنه ضمن الحواسيب (acquisition)، ثم معالجته (processing)، وبعدهما عرضه أو طباعته (display)، وأخيراً نشره واستعماله أو استخدامه في عملية التنمية الاقتصادية والاجتماعية. ولنمو المحتوى العربي فوائد وعائدات على المجتمع والاقتصاد العربي، كما يؤثر على الثقافة العربية واللغة العربية والتقارب العربي.

هناك مؤشرات لقياس المحتوى في لغة من اللغات، وهذه المؤشرات تتبلور عالمياً بغية تقويم التوجه نحو مجتمع المعلومات لأمة من الأمم، وبغية مقارنة التقدم النسبي لمختلف الأمم في هذا المجال. من هذه المؤشرات عدد الصفحات باللغة العربية على الإنترنت (pages)، وعدد المواقع باللغة العربية (sites)، ومدى استعمال هذه المواقع والمعلومات (hits)، وكذلك تقييس استعمال هذه اللغة (standards). وكذلك وجود محركات بحث (search engines) وأدلة (directories) باللغة العربية، ومن المؤشرات أيضاً عائدات الدعاية باللغة العربية في المواقع العربية.

ستتعرض الدراسة أيضاً إلى آليات زيادة المحتوى العربي سواءً كانت offline content أو online content. وهذه الآليات تقسم إلى قسمين: تقنية وغير تقنية، ونذكر منها نظم تعرف الحروف العربية (OCR) أو الترجمة الآلية من الإنكليزية إلى العربية وبالعكس، وبرمجيات النشر على الشبكة باللغة العربية (web publishing)، وإصدار التشريعات المحفزة لزيادة المحتوى العربي، والحاضنات لشركات المعرفة العربية incubators، وتعليم المعلوميات والاتصالات بالعربية وخاصة في برمجيات الشبكات (web ware)، والبحث والتطوير في مجالات معالجة اللغات الطبيعية (Natural Language Processing) [NLP]، وتطوير الأدوات البرمجية لتسهيل التعامل مع اللغة العربية في كافة مراحل إيجاد المحتوى من التوليد حتى الاستعمال مروراً بالحيازة والمعالجة والعرض والنشر.

في الختام، يقدم البحث توصية لزيادة نمو المحتوى العربي.

المحتويات

مقدمة ١

المحتوى على الإنترنت ومستعمليه ((content and users ١

مجالات وجود المحتوى العربي ٧

أنواع المحتوى ٨

مراحل تداول المحتوى العربي ٨

فوائد نمو المحتوى العربي ١٠

مؤشرات قياس المحتوى العربي على الإنترنت ١٠

آليات زيادة المحتوى العربي ١٠

العقبات الفنية أمام توليد المحتوى العربي ١١

خاتمة وتوصية ١١

المراجـــع ١٣

مقدمة

لقد أدخلت التطورات المتسارعة في تكنولوجيا المعلومات والاتصالات تغيرات هامة في عملية التنمية الاقتصادية والاجتماعية، فكانت عاملاً أساسياً في نمو الاقتصاد نحو ما يسمى بالاقتصاد المبني على المعرفة أو الاقتصاد الجديد، وعاملاً أساسياً في توجه المجتمع نحو ما يسمى بمجتمع المعلومات. إن المعرفة أخذت تلعب دوراً أكبر في حياتنا اليوم، وأصبحت تتجسد بشكل رقمي (digital) في الحواسيب ضمن قواعد المعطيات وقواعد المعرفة (database knowledge) وعلى الشبكات وأهمها شبكة الإنترنت. إن وعاء المعرفة هو اللغة، وبالتالي فإن المعلومات تجسدت بشكل رقمي وفق لغات العالم المختلفة. إن مجمل ما يوجد من معلومات في لغة ما بشكل رقمي مخزون في الحواسيب (offline) أو موضوع على الشبكات الحاسوبية (online) هو ما نصطلح عليه بالمحتوى. فمحتوى اللغة العربية الموجود على الإنترنت يمثل كل المعلومات المتوفرة على الإنترنت بشكل رقمي وفي شتى مجالات المعرفة والحياة. يزداد المحتوى بكل اللغات بتسارع هائل سنوياً وبمعدلات أسية، فقد قدرت إحدى الدراسات مجمل محتوى كل اللغات على الإنترنت عام ١٩٩٧ بما يزيد عن مائة مليون صفحة، ليزيد عام ١٩٩٩ عن البليون صفحة. تنبع أهمية المحتوى في لغةٍ ما من فوائدها وعائداتها على مجمع تلك اللغة. فهناك فوائد إدارية وخدماتية من المحتوى المتعلق بالحكومة الإلكترونية، وفوائد اقتصادية من المحتوى المتعلق بالتجارة الإلكترونية وخاصة فوائد التجارة الداخلية على مستوى كل دولة والتجارة البينية بين الدول العربية في حالة اللغة العربية على سبيل المثال، وهناك فوائد تتعلق بتكوين الأطر البشرية في حالة المحتوى التعليمي والتدريبي أو ما يسمى e-learning، وهناك فوائد علمية وتكنولوجية من وجود المحتوى العلمي والتكنولوجي، وفوائد ثقافية من وجود المحتوى الثقافي والتراثي وهكذا.

تزداد أهمية المحتوى وعائداته مع ازدياد المستخدمين للإنترنت وللحاسوب، وتقاس هذه الفائدة بعدد هؤلاء المستخدمين المتكلمين للغة المحتوى المعني، فكلما ازداد عدد متكلمي لغة المحتوى المدروس كلما ازدادت عائدات ذلك المحتوى، ويعرف هذا المبدأ لدى الاقتصاديين بمبدأ "عائدات التشبيك" (network effects) أو (network externalities).

لقد وصل عدد مستعملي الإنترنت في تموز/يوليو ٢٠٠٢ إلى ٥٦٠ مليون مستخدم في العالم تصل نسبة متكلمي غير اللغة الإنكليزية فيهم إلى ٥٩.٨% أي أن نسبة المستعملين ممن تعتبر الإنكليزية لغة الأم لهم هي ٤٠.٢% وهذا يدل على تزايد وأهمية إيجاد المحتوى بغير اللغة الإنكليزية. إن إيجاد المحتوى باللغات الوطنية يعدُّ من العوامل المقللة لما يسمى بظاهرة الهوة الرقمية (digital divide).

تتسارع مختلف الأمم لزيادة المحتوى بلغتها والأمة العربية بدأت تسعى بهذا الاتجاه.

المحتوى على الإنترنت ومستعمليه (content and users)

يزداد المحتوى على الإنترنت بمعدلات أسية وأحد مؤشرات قياس المحتوى هو عدد الصفحات على الإنترنت (web page). لقد وصل عدد الصفحات في تموز/يوليو ٢٠٠٢ إلى حوالي ٣١٣ بليون صفحة في كل اللغات، وكان قد زاد هذا العدد من حوالي البليون صفحة عام ١٩٩٩، أي بمعدل ٣٠٠ مرة في السنوات الثلاث الأخيرة. تتصدر اللغة الإنكليزية اللغات الأخرى حيث تشكل ما نسبته ٦٨.٤% من الصفحات، تليها اللغة اليابانية فالألمانية فالصينية. يبين الجدول رقم (١) اللغات العشر ذات المحتوى الأعلى على الإنترنت، وليست اللغة العربية بين هذه اللغات العشر رغم أنها تأتي عالمياً ضمن اللغات الست الأولى من حيث عدد متكلميها، كما أنها لغةٌ رسمية من لغات الأمم المتحدة الرسمية الستة.

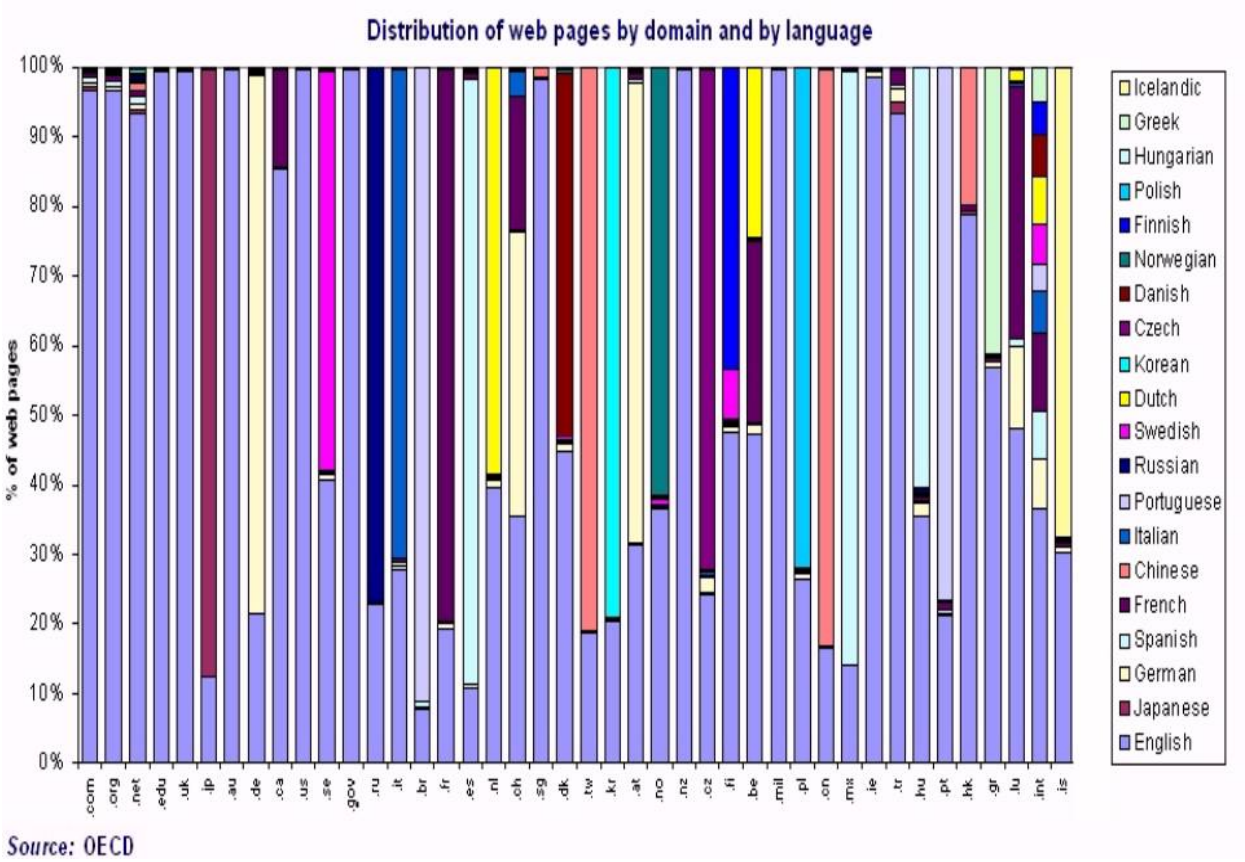
الجدول رقم -١-

المحتوى على الإنترنت (WEB) حسب اللغات	
الإنكليزية	%٦٨.٤
اليابانية	%٥.٩
الألمانية	%٥.٨
الصينية	%٣.٩
الفرنسية	%٣.٠
الإسبانية	%٢.٤
الروسية	%١.٩
الإيطالية	%١.٦
البرتغالية	%١.٤
الكورية	%١.٣
باقي اللغات	%٤.٦
مجموع صفحات WEB	**٣١٣ بليون صفحة**

Source: Vilaweb.com, as quoted by emarketer

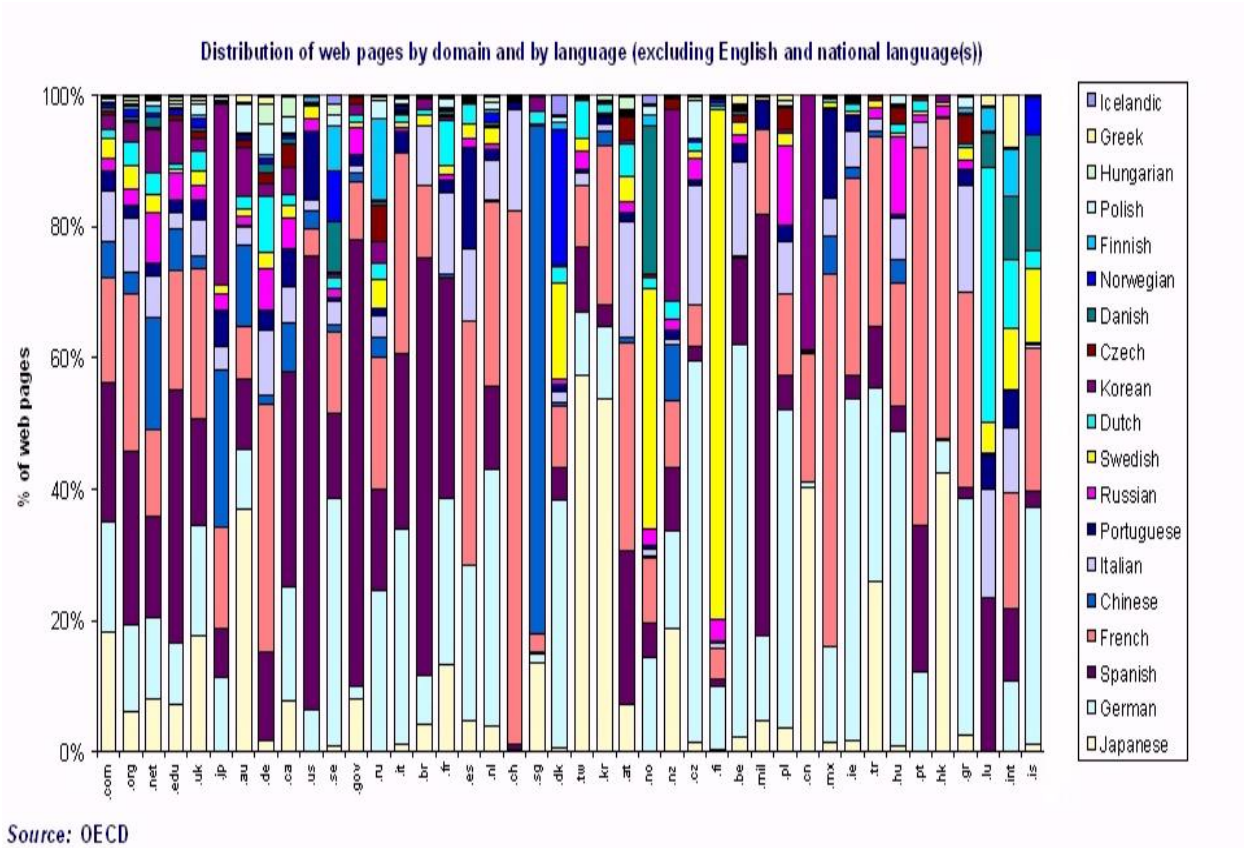
يبين **الشكل رقم –١-** توزع الصفحات web pages حسب المسجلين تحت حقل كل دولة أو حقلٍ عام (domain) وذلك عام ١٩٩٩، وكما يبين هذا الشكل فقد كانت الإنكليزية مستعملة من قبل المسجلين تحت حقول الدول والحقول العامة (الحقول العامة هي .com، .org، .net، .edu، .gov، .mil). نلاحظ أن اليابانيين والبرازيليين والفرنسيين والمكسيكيين والرومان والألمان والإيطاليين والأسبان والتايوانيين وغيرهم يضعون أكثر من ٧٥% من صفحاتهم بلغاتهم الأم. يبين **الشكل رقم –٢-** توزيع صفحات الـ web حسب اللغات وحسب حقول الدول والحقول العامة (domain) بعد استثناء اللغة الإنكليزية واللغات الوطنية، وهذا الشكل يدلل على أن كافة المواقع تسعى لوضع محتوى بلغات عديدة حسب مصالحها ولنشر ثقافتها ولتسويق منتجاتها.

الشكل رقم -١-



Source: OECD

الشكل رقم -٢-



Source: OECD

كما ذكرنا في المقدمة، فإن فائدة المحتوى بلغة ما تزداد مع ازدياد عدد مستخدمي الإنترنت من متكلمي تلك اللغة. ويبين **الجدول رقم -٢-** المؤرخ في تموز/يوليو ٢٠٠٢ توزع مستخدمي الإنترنت في العالم وعددهم ٥٦٠ مليون شخص حسب اللغات، ونجد أن عدد مستخدمي الإنترنت من متكلمي اللغة العربية هو حوالي ٤.٤ مليون مستخدم في العالم العربي والمهجر، طبعاً لا يتضمن هذا الرقم العارفين للغة العربية في العالم الإسلامي، الذي بلغ تعداد سكانه عام ٢٠٠٢ ما يزيد عن مليار ومائتي مليون شخص. كما نرى من الجدول أن نسبة مستخدمي الإنترنت من متكلمي اللغة العربية يشكلون حوالي ٠.٨٩% من مجمل مستخدمي الإنترنت في حين أن سكان العالم العربي يشكلون ما نسبته ٥% من مجموع سكان العالم. نستنتج من هذا أن نسبة المحتوى العربي ونسبة المستخدمين العرب على الإنترنت لا تزال أقل مما يجب.

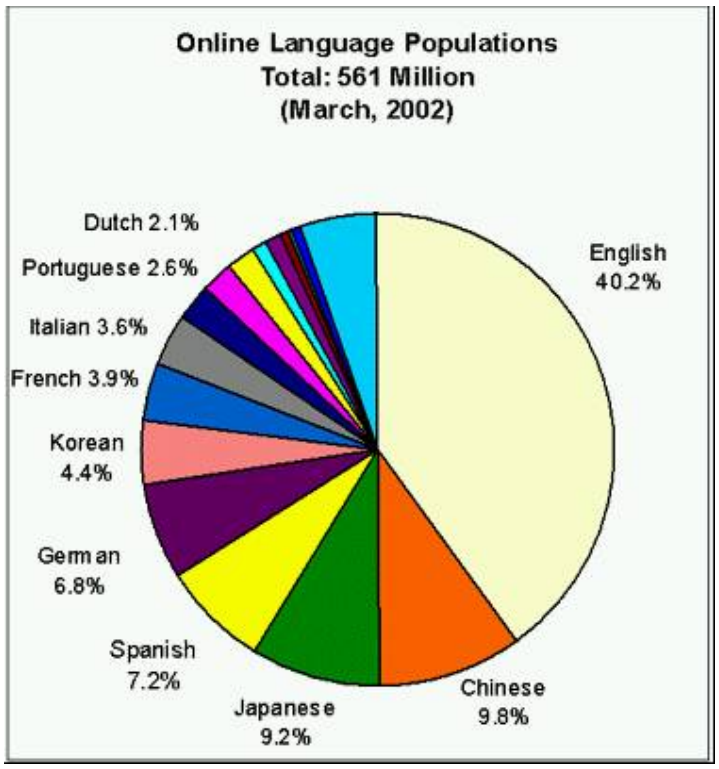
الجدول رقم -٢-

	Internet access (M)	%'age world online pop.	2003 (est. in M)	Total pop. (M)	GDP ($B)	%'age of world economy	GDP per capita (K)	Net Hosts
English	**228**	**40.2%**	**270**	567	$13,812	33.4%		
Non-English	**339**	**59.8%**	**510**	5633	$27,590	66.6%		
European Languages (non-English)	**192.3**	**33.9%**	**259.3**	1,218	$12,550	30.3%		
Catalan	1.9 [2]		2.2	6.6				
Czech	2.2 [3]		3	12	$53		$5.1	214
Dutch	11.8 [4]	2.1%	13	23.6	$570		$24.2	2485
Finnish	2.1 [5]		3.5	6	$127		$24.4	945
French	22.0 [6]	3.9%	28	77	$1734	4.2%	$21.5	2388
German	38.6 [7]	6.8%	49	100	$2421	5.8%	$24.9	3784
Greek	1.6 [8]		3	12	$184		$16.9	182
Hungarian	1.3 [9]		3	14.5	$96		$9.4	211
Italian	20.2 [10]	3.6%	27	62	$1471	3.6%	$24.7	2313
Polish	6.7 [11]		8.5	44	$306		$7.8	654
Portuguese	14.9 [12]	2.6%	26	176	$1472	3.6%	$8.34	1909
Romanian	0.8 [13]		1.2	26	$98		$4.4	69
Russian	11.5 [14]	2.0%	15	167	$730	1.8%	$5.0	415
Danish	3.2 [15]			5.4	$176		$32.9	707
Icelandic	0.2 [15]			.3	$6		$23.5	47
Norwegian	2.5 [15]			5	$126		$27.7	630
Swedish	6.2 [15]			9	$223		$22.3	1330
Scandinavian languages (total)	12.0 [15]	2.1%	13	19.7	$525	1.3%	26.0	2714
Slovak	0.7 [16]		1.5	5.6	$47		$8.7	69
Slovenian	0.6 [17]		1	2	$22.9		$10.9	26
Spanish	40.8 [18]	7.2%	53	350	$3684	8.9%	$11.0	3241
Turkish	3.9 [19]		7	67.4	$454		$6.7	140
Ukranian	0.8 [20]		2	47	$115		$2.3	56
TOTAL EUROPEAN LANGUAGES (excl. English)	**192.3**	**33.9%**	**259**	1,218	$14,112	33.9%		24,529
ASIAN LANGUAGES								
Arabic	4.4 [20]	0.8%	6	300	$678	1.6%	$4.2	95
Chinese	55.5 [21]	9.8%	125	874	$5370	13.0%	$5.4	2388
Hebrew	1.9 [22]		2.5	5.2	$132		$21.0	223
Japanese	52.1 [23]	9.2%	75	125	$3,315	8.0%	$26.1	7118
Korean	25.2 [24]	4.4%	35	78	$835	2.0	$17.3	440
Malay	4.8 [25]		7	229	$835	2.0%	$3.7	95
Thai	2.3 [26]		3	46	$453		$7.3	81
TOTAL ASIAN LANGUAGES	**146.2**	**26.1%**	**254**					10,440
TOTAL WORLD	**560**		**762**	**6,200**	**$41,400**			
Source: Global Reach (global-reach.biz/globstats)								

يبين **الشكل رقم –٣-** توزع مستخدمي الإنترنت حسب لغاتهم الأم بتاريخ آذار/مارس ٢٠٠٢، أما **الشكل رقم –٤-** فيبين تطور أعداد المستخدمين حسب لغاتهم الأم مع الزمن.

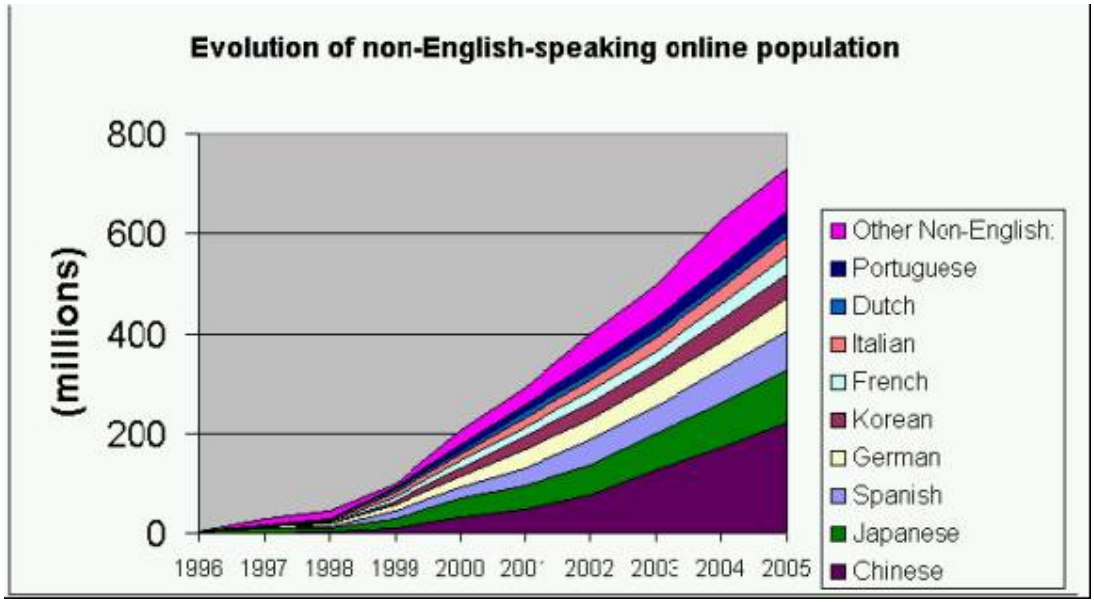
يخزن المحتوى في مخدماتٍ (servers) حاسوبية تدعى المضيفات (host). ويعد خزن المعلومات في مضيفات ضمن الدولة من مكاسب هذه الدولة وله عائدات اقتصادية واجتماعية وأمنية. تسعى الدول لزيادة عدد المضيفات لديها، وهذا لا يتحقق إلا إذا كان تشغيل المضيفات اقتصادياً بالنسبة لواضعي المحتوى، أو بتعبير آخر إن نجاح دولة ما في زيادة عدد المضيفات لديها يتطلب أن تكون إدارة هذه المضيفات اقتصادية ومنافسة وموثوقة وآمنة بالمقارنة مع المضيفات المعروضة من قبل الدول الأخرى.

الشكل رقم –٣-



Source: Global Reach ٢٠٠٢, www.globalreach.com

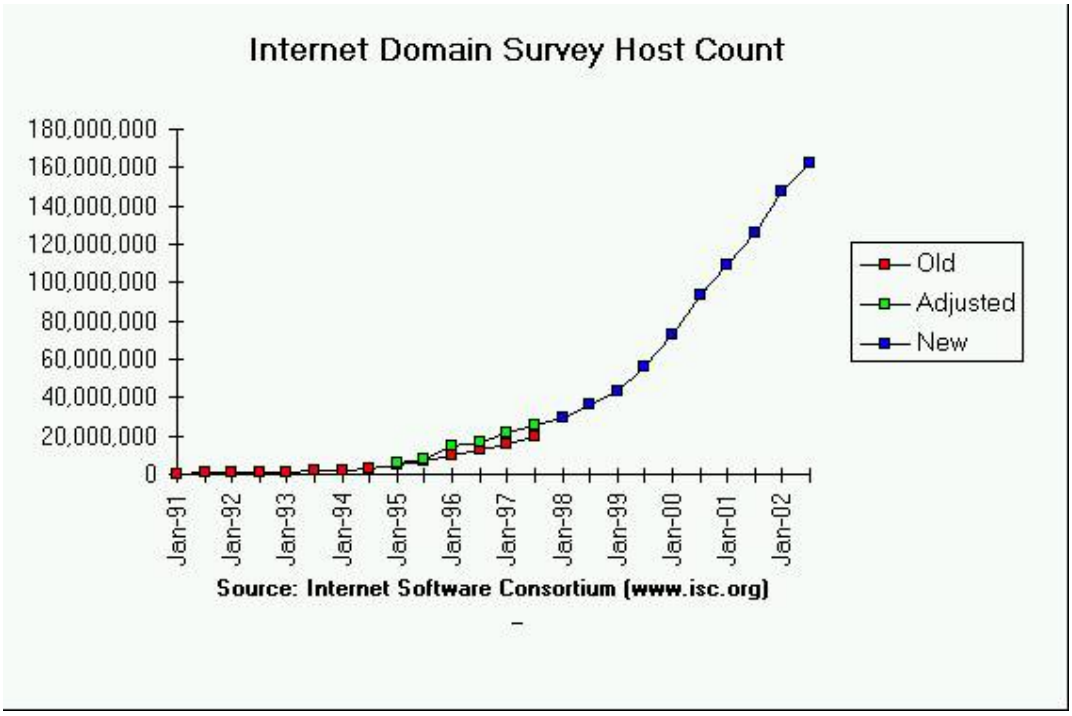
الشكل رقم -٤-



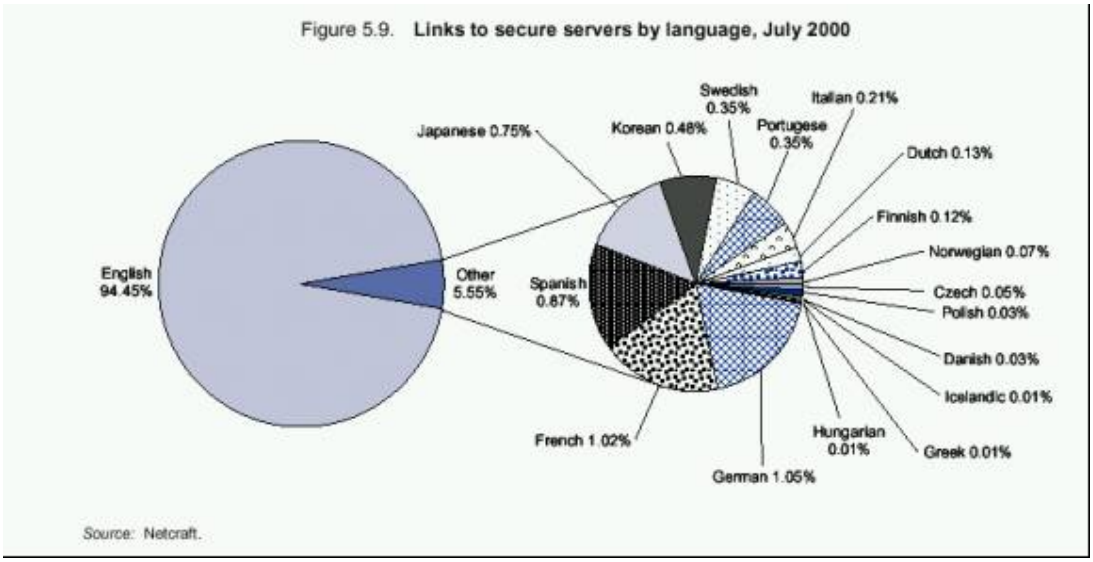
Source: Global Reach ٢٠٠٢

والعالم العربي مدعو للاهتمام بزيادة عدد المضيفات لديه، وخفض أسعار استضافة المحتوى العربي، وضمان أمن هذا المحتوى من العبث. يبين **الشكل رقم -٥-** تطور عدد المضيفات في العالم إذ زاد عن ١٥٠ مليون مضيف عام ٢٠٠٢، منها في العالم العربي ١١٢.٥٢٢ مضيف عام ٢٠٠١. هذا العدد من المضيفات يمثل مجمل المضيفات آمنة (secure) أو غير آمنة. ويبين **الشكل رقم -٦-** أن معظم الوصلات للمخدمات الآمنة هي باللغة الإنكليزية (٩٤.٤٥% من المخدمات) أما الوصلات باللغات الأخرى لمخدمات آمنة فلا تشكل أكثر من ٥.٥٥%. ولا تظهر العربية بين اللغات الـ ١٧ الأول في هذا المجال.

الشكل رقم -٥-



الشكل رقم –٦-



مجالات وجود المحتوى العربي

ينتشر المحتوى العربي على الإنترنت كانتشار أي لغة أخرى على مختلف مجالات الاقتصاد والاجتماع والثقافة وغيرها. وندرج أدناه أمثلة حول بعض هذه المجالات، التي يأمل أن يزداد المحتوى العربي فيها:

- الأعمال: مواقع الشركات، دليل الشركات، دليل المصدرين، دليل المصانع، البنوك، صناديق المال،...
- النشر: الجرائد، المجلات، الدوريات العلمية، الإذاعات، التلفزيونات.....
- الحكومة الإلكترونية: البوابة الحكومية، مواقع الوزارات والمؤسسات العامة،
- العلم والتكنولوجيا: الجامعات، مراكز البحوث، الجامعات الافتراضية، ....
- المكتبات: نص، صوت، صور، فلم، الكتاب الإلكتروني، ....
- الصحة: العيادات، المستشفيات، الطبابة عن بعد،...
- المنظمات: غير الحكومية، الإقليمية، الدولية،...
- الثقافة: المتاحف، بوابات الثقافة والفكر والموسيقى والأدب، والرسم و...
- التراث: التراث العربي، التراث الإسلامي، (adabwafan) و (Alwaraq)...
- السياحة: المواقع السياحية التاريخية والطبيعية، والفنادق، والمطاعم، والنقل،...
- التسلية: ألعاب أطفال، أفلام،...
- .....

إن تنشيط زيادة المحتوى في كل من هذه المجالات يحتاج لمبادراتٍ من قبل الحكومات العربية، تستهدف كل مبادرة منها زيادة المحتوى في مجال من المجالات. وهذه المبادرات يجب أن تشتمل على النواحي القانونية والمؤسساتية والبشرية والمالية والدعائية والتنظيمية والإدارية.

أنواع المحتوى

إن المحتوى الذي يغطي المجالات المذكورة أعلاه يمكن أن يصنف في أربعة أنواع عامة هي: معرفة ماذا، معرفة لماذا،معرفة كيف، معرفة مَن. تعكس المعرفة مدى السيطرة على الأشكال المختلفة للمعلومات ويمكن تسمية تصنيف المعرفة هذا في أربعة تسميات هي أ- معرفة المعلومة ب- معرفة العله ج- معرفة الكيفية د- معرفة أهل الاختصاص

وتعمل تكنولوجيا المعلومات الآن على ترميز هذه الأنواع من المعرفة وبالتالي تحويلها إلى سلع تؤثر(بشكل أكثر مباشرة مما مضى) في الاقتصاد والمال والمنعة الوطنية.

أ- "معرفة المعلومة" أو "معرفة ماذا- know what - تشتمل على معرفة الحقائق وهي أقرب ما تكون إلى معرفة المعلومات التقليدية ، كمعرفة الحقائق الطبية من قبل الطبيب أو معرفة القوانين والشرائع من قبل المحامي وأمثالها.

ب- "معرفة العلة" أو " معرفة ماذا" - know why - وتشتمل على معرفة الأسباب وراء ظواهر الطبيعة واستثمارها لخدمة الإنسان ، وتكمن هذه المعرفة وراء التقدم العلمي والتكنولوجي و وراء الصناعة وإنتاج السلع المختلفة, وتتركز مصادر هذه المعرفة في وحدات التعليم والبحث والتطوير العام والخاص.

ج- "معرفة الكيفية" أو "معرفة كيف" – know how – وتشير هذه المعرفة إلى الخبرة في تنفيذ الأشياء سواء كانت هذه الأشياء هي إدارة الأفراد أو تشغيل العمليات أو تشغيل الأجهزة والآلات أو استخدامات التكنولوجيا المختلفة، وعادة ما تكون هذه المعرفة ملكاً للشركات والمؤسسات ويحتاج الحصول على بعضها إلى آليات مختلفة ومعقدة ومكلفة.

د - "معرفة أهل الاختصاص" أو" معرفة من" - know who - وتزداد حالياً أهمية هذه المعرفة، معرفة من يستطيع عمل شيء ما لا بد منها لتنفيذ هذا العمل بشكل سليم واقتصادي. وتفعيل الاقتصاد حالياً يحتاج لهذه المعرفة حاجة كبيرة . كما تسرع هذه المعرفة تنفيذ المشاريع تسريعاً أكيداً وسليماً.

إن تعليم السيطرة على هذه الأنواع الأربعة من المعرفة يتم عبر وسائط مختلفة . " فمعرفة المعلومة" " ومعرفة العلة" تؤخذان من الكتب والمؤسسات التعليمية والتدريبية ومن قواعد المعلومات . أما النوعان الآخران فلا يؤخذان كاملاً إلا بالممارسة .

لكن توفير المعرفة وتحويلها إلى معلومات جعل من تكنولوجيا المعلومات IT أداة هائلة في وضع المعرفة في متناول العالم ، خاصة وأن شبكات المعلومات مثل الانترنت وغيرها تجعل المسافات قصيرة والزمن مختصر والتكلفة بسيطة والتداول سهلاً . إن هذا الترميز للمعرفة وتخزينها رقمياً انطلاقاً من توفرها كمعلومات على شكل كتب ومجلات وأوراق عمل ومراجع وفهارس وصور وصوت وأفلام ورسومات ، إضافة لتسهيل نقلها عبر الشبكات الرقمية العالمية يجعلها أداة للتنمية الاقتصادية والثقافية والأمنية ذات دور فعال للغاية ، وهذا ما يقر بنا من " مجتمع المعلومات" الذي يولد وينقل ويستعمل المعرفة لخدمته في كافة المجالات .

إن توفير المعرفة وتحويلها إلى معلومات رقمية يجعلها تتحول الى سلعة تزداد أنواعها يوماً بيوم ويزداد دورها في الاقتصاد العالمي الذي يتجه نحو " اقتصاد المعرفة " .

مراحل تداول المحتوى العربي

هناك مراحل محددة للتداول مع المحتوى العربي بدءاً من توليده وانتهاءً باستعماله أو استثماره، وكل مرحلة من هذه المراحل تحتاج لجهود وأدوات ومشاريع خاصة بها. وسنذكر باختصار هذه المراحل فيما يلي:

(١) مرحلة توليد المحتوى الجديد: وهي مرحلة إبداعية تعكس نشاط الأمة وإنتاجها الفكري والثقافي والعلمي والتكنولوجي. يولد المحتوى الجديد في لغة من اللغات من نشاطات الأمة في البحث والدراسة والتطوير. وتعتمد عملية توليد المحتوى العربي على عوامل عدة منها تمويل البحث

العلمي، وحرية الفكر والتعبير عنه، ووجود الطلب على الإبداع والتجديد. إن التعبير أصبح في عصرنا الحالي يجري بشكل رقمي (digital) يخزن في الحاسوب مباشرةً، فالكتب والدراسات والبحوث والثقافة والأدب تكتب حالياً مباشرة بشكل رقمي قابلٍ للنشر الرقمي ولوضعه على الإنترنت مباشرة. وهذا ما بدأت تمارسه الكثير من دور النشر العربي حالياً.

(٢) مرحلة تحويل المحتوى الموجود أو القديم إلى الشكل الرقمي (go digital): وهي مرحلة تتطلب إدخال معارف الأمة السابقة من كتب ووثائق وفن ومعلومات وغير ذلك إلى الشكل الرقمي، وخزنها في الحاسوب أو على وسائط الخزن الرقمية كالقرص الليزري CD Rom، أو القرص المغناطيسي diskette، ويشهد العالم العربي الكثير من المشاريع في هذا المجال مثل: www.Alwaraq.com.

(٣) مرحلة تخزين المحتوى وتبويبه ومعالجته: يجري في هذه المرحلة الهامة تبويب المعلومات المخزنة ضمن أشكال تسهل البحث فيها، وتسهل استخلاص المعلومات اللازمة منها. تستعمل في ذلك العديد من الأدوات البرمجية كقواعد المعطيات Databases، وقواعد المعرفة Knowledge base، والنظم الخبيرة expert system، وبرمجيات الفهرسة الآلية (indexing)، وبرمجيات معالجات اللغات الطبيعية Natural Language Processing (NLP)، مثل: برمجيات البحث search عن الكلمة أو الوثيقة أو المعنى، ومثل: برمجيات فهم النصوص، وبرمجيات تحليل النصوص، وبرمجيات الترجمة الآلية. ويندرج هذا أيضاً على الكلام العربي مثل برمجيات تعرف الكلام speech recognition وبرمجيات فهم الكلام وترجمته، وبرمجيات تركيب الكلام speech synthesis وغيرها.

إن التعامل باللغة العربية مع كل هذه الأدوات هو تعامل كثيف اللغة language intensive، ويعتمد على اللغة وخواصها، والكثير منها لا بد من تطويره من جديد من أجل اللغة العربية خاصة. وللنجاح في هذه المرحلة لا بد للدول العربية من اعتماد مبادرات على مستوى الدولة والقطاع الخاص لدعم البحوث والتطوير فيها، ولتشجيع قيام الشركات الخاصة بها، ولتوفير البيئة المناسبة لنموها. لقد قامت في العالم العربي العديد من النشاطات الخاصة في هذه المرحلة، من أهمها نشاطات شركة صخر ونشاطات شركة مايكروسوفت، ولكن المطلوب لا يزال هائلاً والجهود المبذولة في العالم العربي لا تزال مشتتة، رغم أهمية هذه النشاطات وعائداتها الاقتصادية الكبيرة الواعدة والتي بدأت تشكل ما يسمى بـ language intensive industry الصناعات كثيفة اللغة.

(٤) مرحلة عرض المحتوى أو طباعته: وهي مرحلة تتعلق بالتعامل مع الحرف العربي وأشكال طباعته أو إظهاره أو نقله عبر شبكات الحواسيب وعبر الإنترنت، وهي مرحلة تحتاج لجهد في تقيس استعمال حروف اللغة العربي (standards) وقد قامت جهود عربية عديدة في هذه المجالات ولا تزال قائمة ولكنها بطيئة وضعيفة: نشاطات المنظمة العربية للتقييس سابقاً ASMO، والآن AIDMO، و ALECSO.

(٥) مرحلة نشر المحتوى العربي: أهم ما في هذه المرحلة هو وضع المحتوى العربي الرقمي (digital) لكل المجالات التي أتينا على ذكرها على الإنترنت، وفهرسته (indexing) في محركات البحث على الإنترنت (search engines). النشاط العربي في هذا المجال لا يزال ضعيفاً وخاصة في مجال اللغة العربية ومحركات البحث.

(٦) مرحلة استخدام واستعمال المحتوى: يعتمد نجاح هذه المرحلة على جودة المحتوى وعلاقته وفائدته للمستثمر (relevant)، كما يعتمد على زيادة معدل النفاذ العربي للإنترنت (access) بما في ذلك نفاذ الصناعيين والتجار والطلاب والمواطنين والمثقفين، أي كافة شرائح المجتمع. كما يعتمد نجاح هذه المرحلة على أسعار الحواسيب وأسعار الاشتراك بالإنترنت والهاتف وتوافرها أي على البنية التحتية لتكنولوجيا المعلومات والاتصالات.

فوائد نمو المحتوى العربي

إن زيادة المحتوى العربي سيعود بفوائد كبيرة سواءً من الناحية الاقتصادية أو الاجتماعية أو الثقافية، فهو أولاً سيسمح بالانتقال في النفاذ الى المعرفة من نخبة صغيرة في المجتمع تتكلم الإنكليزية أو الفرنسية إلى قوى المجتمع العاملة بأسرها، وهو عامل هام جداً في الاقتصاد الجديد وفي مجتمع المعلومات القادم. والفائدة الثانية ستكون في تنشيط تعليم وتعلم وتدريب المجتمع العربي لاستعمال التقانات الجديدة الفعالة كالإنترنت عبر e-learning وكالكتاب الإلكتروني e-book، والتدريب عن بُعد وغيره. وفائدة ثالثة ستتحقق عبر زيادة التجارة المحلية e-commerce والتجارة البينية العربية. أما الفائدة الرابعة فستتأتى عبر تطبيق الحكومة الإلكترونية e-government التي يتوخى أن تؤدي إلى الإدارة الشفافة والفعالة والحكم الرشيد. والفائدة الخامسة ستتحقق من خلال التشبيك لمختلف النشاطات الاقتصادية والاجتماعية العربية (networking) كتشبيك مراكز البحوث والجامعات والصناعات عبر بوابات عربية تنشأ لهذا الغرض (portals). ولسنا هنا في صدد ذكر كل الفوائد ولكن أتينا على بعضها، ولكن هذه الفوائد لن تتحقق إذا لم ينمى الطلب عليها وفق مبادرات وطنية وعربية من الحكومات والقطاع الخاص والمنظمات غير الحكومية والمجتمع المدني (demand pull).

مؤشرات قياس المحتوى العربي على الإنترنت

لقد بدأت تعتمد عالمياً بعض المؤشرات لقياس ومقارنة المحتوى لمختلف اللغات، وسنذكر بعض هذه المؤشرات بشكل مختصر، علماً بأن تحصيل قيم هذه المؤشرات ومتابعتها يُعدّ عملاً جديداً في العالم العربي ويحتاج لجهود ومبادرات من الحكومة والقطاع الخاص ومن المنظمات الإقليمية والدولية حتى يجري اعتمادها وقياسها. من هذه المؤشرات ما يلي:

- عدد الصفحات باللغة العربية
- عدد المواقع باللغة العربية وفي الدول العربية Web sites
- فائدة المحتوى للدول العربية local relevance
- مدى تغطية المجتمعات النائية communities coverage
- مدى تقييس استعمال اللغة على الشبكات standards
- وجود محركات بحث وأدلة باللغة العربية directories and search engines
- عائدات الدعاية باللغة العربية في المواقع العربية

لقد ذكرنا في بداية هذه الورقة قيم بعض هذه المؤشرات كعدد الصفحات web pages في كل اللغات والذي يساوي ٣١٣ بليون صفحة، وعدد الصفحات للغات العشر الأولى على الإنترنت.

آليات زيادة المحتوى العربي

كما ذكرنا فإن من المحتوى ما هو على الشبكة online content ومنه ما هو ليس على الشبكة offline content، وزيادة المحتوى أياً كان يحتاج لتوفير أو دعم آليات ضرورية لتحقيق هذه الزيادة سنعددها فيما يلي:

- تعرف الحروف العربية OCR
- Machine Translation من الإنكليزية إلى العربية
- التحول نحو الرقمية go digital (مبادرات وطنية)
- النشر على الإنترنت Web publishing (web ware)
- تقييس استعمالات اللغة العربية في المعلوميات والاتصالات
- إصدار تشريعات محفزة لقيام صناعات في المحتوى العربي (Arabic content industries)
- حاضنات شركات المعرفة العربية (language intensive industries)
- تعليم المعلوميات والاتصالات (بالعربية) وخاصة web ware
- دعم مالي لمشاريع زيادة المحتوى العربي (حوافز ضريبية، صناديق...)
- بحث وتطوير في مجالات NLP، speech، لسانيات حسابية
- تطوير الأدوات البرمجية لتسهيل التعامل مع اللغة العربية في كافة مراحل تداول المحتوى من التوليد إلى الاستثمار أو الاستخدام.

العقبات الفنية أمام توليد المحتوى العربي

يوجد عدد من العقبات الفنية التي لا تسهل وجود المحتوى العربي، وهي تحتاج لورقة خاصة بها حيث تعالج أموراً فنية تقنية يمكن إيجاد حلول لها إذا قامت الجهات العربية المختصة بالتعاون مع الجهات الدولية المعنية بها لإيجاد الحلول، نذكر منها:

- معايير حول تقييس استعمال اللغة العربية على الإنترنت: مثل وجود التعليمات RFC باللغة العربية، ومثل استعمال اللغة العربية في أسماء الحقوق وللعناوين على الإنترنت domain names، واستعمال اللغة العربية في البريد الإلكتروني، وفي البروتوكول FTP وغيرها وفي usenet news، وفي خدمات الأدلة repertoires، وفي قاعدة المعطيات whois II

- استعمال اللغة العربية في لغات البرمجة الخاصة بالإنترنت كبروتوكول HTTP ولغة HTML وخاصة المؤشرات في هذه اللغة (tags)

- تطبيقات الإنترنت كاستعمال اللغة العربية في محركات البحث وفي الدردشة (chat) وفي إدارة القوائم (list-serve)، وفي برمجيات التعمية أو التشفير.

- إشكالات تتعلق بأنظمة التشغيل Unix و windows و linx و x-window، حيث لا تزال بعض الاشكالات المتعلقة باللغة العربية وغيرها من اللغات. ونلاحظ جهوداً تبذل في هذا الاتجاه من قبل بعض الشركات الخاصة ومن قبل المنظمات الإقليمية والدولية المعنية، ولكن هذه الجهود بطيئة وضعيفة.

خاتمة وتوصية

إن المحتوى العربي أصبح ضرورة ملحة مع التوجه نحو الاقتصاد المبني على المعرفة ونحو مجتمع المعلومات، ونوصي بتبني مبادرات على مستوى كل دولة عربية وعلى المستوى القومي، تهدف هذه المبادرات لتحقيق الآليات التي أتينا على ذكرها في هذه الدراسة، وإلى إيجاد الحلول للعقبات التقنية التي تعترض نمو هذا المحتوى.

المراجـــع

- Robert M. Solow, ***"Growth theory"*** ٢nd edition, Oxford University Press, ٢٠٠٠.

- Albert Breton (editor), ***"Economic approaches to Language and Bilingualism"*** New Canadian Perspectives, Department of public works and Government Services, ١٩٩٨.

- Francois Grin, ***"Languages and the Economy: European Research on the Economics of Language: Recent Results and Relevance to Canada"***, Department of Political Economy, Universite De Genere, www.pch.gc.ca

- Proceeding of the Colloquium ***"Official Languages and the Economy"*** ٢٧١ pages, www.pch.gc.ca

- Ngugi Wa Thiongo, ***"The role of colonial Language in Creating the image of a savage Continent"*** www.trinicenter.com

- Vikas Nath, "***Heralding ICT enabled Knowledge Societies: Way forward for the developing countries"*** http://members.tripod.com

- ***"Global Competitiveness Report ٢٠٠٠/٢٠٠١"*** Center for International Development, Harvard University, ٢٠٠٠.

- M. Mrayati, "***Knowledge-Based-Economy, and the Need to Arabization of Information Technology in the ESCWA Region***", Expert Panel on IT and Development priorities competing in a knowledge-based global economy, UN-ESCWA, ١٥-١٦ May ٢٠٠٠. And Review of S&T in the ESCWA Member Countries, No. ٣, ٢٠٠٠.

- M. Mrayati, "***Arabization and Localization of E-Business: Status and Requirements"***. **EGM** on Trade Facilitation and e-commerce in the ESCWA region, UN-ESCWA, Beirut, ٨-١٠ November ٢٠٠٠.

- M. Mrayati, "***Technology Transfer and Emerging Markets***", Environment ٢٠٠١ Conference: Challenges and Solutions for Sustainable Development, Abu Dhabi, ٤-٧ February ٢٠٠١.

- محمد مراياتي، *"الإنترنت والوطن العربي"* مجلة العلوم، الألكسو، تونس، كانون الأول/ديسمبر ١٩٩٩

- محمد مراياتي، *"تغيير منظومة العلم والتكنولوجيا إلى نظام وطني للإبداع من ضرورات التنمية في القرن الحادي والعشرين"* مجلة العلوم، الألكسو، تونس، كانون الأول/ديسمبر ١٩٩٩.

- محمد مراياتي، *"التكنولوجية الحديثة والمصطلح العلمي العربي في ظل اقتصاد المعرفة"*، ندوة إقرار منهجية موحدة لوضع المصطلح العلمي العربي وتوحيده وإشاعته، اتحاد مجامع اللغة العربية، دمشق، ٢٥-٢٧ تشرين الأول/أكتوبر ١٩٩٩.

- محمد مراياتي، *"تعريب المعلوماتية في ظل اقتصاد المعرفة ودور التربية والتعليم فيه"*، ندوة أسئلة التعريب ورهاناته في التعليم العالي بالمغرب وسورية، فاس، جامعة سيدي محمد بن عبد الله، ٢٥-٢٦ تشرين الثاني/نوفمبر ١٩٩٩.

- محمد مراياتي، *"صناعات تكنولوجيا المعلومات واللغة العربية: الفرص الكامنة"* ندوة تعريب البرمجيات، الجامعة اللبنانية الأمريكية، بيروت، ٢٠٠٠/٣/٣٠.

- محمد مراياتي، *تطور مهن المعلوماتية ومردودها الاقتصادي"* ندوة الدراسات الإنمائية، المؤتمر الوطني السابع عشر للإنماء، بيروت، ٩-١١/١١/٢٠٠٠.

- محمد مراياتي، *" تعريب العلوم والتكنولوجيا وضرورته مع توجه العالم نحو اقتصاد المعرفة"*، ندوة الترجمة والتقانة العربية، المجلس الوطني للثقافة والفنون والآداب، الكويت ٤-٢٠٠١/١/١١.