MEDAR

Mediterranean Arabic Language and Speech Technology

Deliverable 4.1

Cooperation Roadmap

Authors: Bente Maegaard, University of Copenhagen, Mustafa Yaseen, Amman University, Steven Krauwer, University of Utrecht, Khalid Choukri, ELDA
Contributors: All MEDAR partners
July 2010

MEDAR partners

- **University of Copenhagen:** Centre for Language Technology, Denmark (coordinator)
- **ELDA,** Evaluations and Language resources Distribution Agency , France
- **University of Balamand:** Research Council - Speech and Image Research Group (SIR), Lebanon
- **Amman University:** Faculty of Information Technology, Jordan
- **University of Utrecht:** Utrecht Institute of Linguistics OTS, the Netherlands
- **Research and Innovation Centre "Athena":** ILSP, Institute for Language and Speech Processing, Greece
- **RDI,** The Engineering Company for the Development of Computer Systems, Egypt
- **Birzeit University:** Center for Continuing Education, West Bank and Gaza Strip
- **University Mohammed V Souissi:** Ecole Nationale Supérieure d'Informatique Analyse des Systèmes, Morocco
- **CEA,** Commissariat à l'Energie Atomique: CEA-LIST/LIC2M, Laboratoire d'Ingénierie de la Connaissance Multimédia Multilingue, France
- **CNRS,** Centre National de la Recherche Scientifique, Laboratoire LLACAN - UMR 8135 du CNRS, Langage, langues et cultures d'Afrique Noire, France
- **The Open University:** Computing Department, Maths & Computing Faculty, The United Kingdom
- **Université Lumière Lyon2:** Groupe SILAT, France
- **IBM** International Business Machines WTC - Egypt Branch, Egypt
- **Sakhr** Software Company, Egypt

The MEDAR project is supported by the ICT programme

CONTENT

1. Executive summary

This document contains a proposal for a Cooperation Roadmap with the purpose of building a sustainable Human Language Technologies for the Arabic language within and outside the Arabic world.

An introduction to the concept of roadmapping is firstly introduced and the particular purpose of the MEDAR roadmap is elaborated upon. The roadmap (section 4) consists of milestones for three (partially overlapping) phases: *Laying the foundations* 2010-1012, *Moving forward* 2012-2014 and *First consolidation* 2013-2015. The roadmap aspires to address a new perspective on collaboration between the Arabic region and the European Union. In order to do so, the MEDAR consortium adopted a multi-dimension roadmap that combines various impacting factors helping to derive a coherent view. Such factors are related to the state of players, human resources and education curricula, technology development and R&D, evolution of the e-infrastructures (in particular mobile and Internet penetration, attractiveness of ICT environments, growth of e-Content in Arabic). Another dimension which is considered is the Market: both the domestic and international ones are reviewed (focussing on local and international players and the market profile is analysed (products versus services)).

Last but not least, a set of instruments are elaborated upon with the aim to boost cooperation between universities and industries (both within the region but also with the EU and the West in general), to improve the technology transfer (from local R&D players to local/international industries).

All these aspects are analysed and recommendations are drawn by the consortium whenever feasible.

In order to support our analysis and recommendations, annexes are appended to this document with some facts and figures.

Note: This report was written in spring 2009. The present 2010 version is slightly updated in order to take into account the comments we have received, and a few developments, but the overall basic statistics on which the report is building has not been touched, and the perspective is still taking spring 2009 as point of departure.

2. Purpose of MEDAR and of this report

The development of language resources and tools for the Arabic language is important for the economy in the Arab countries; but at the same time it is important for the culture. By focussing on Arabic language technology and making both the technology and the "digital" content available in Arabic, the use of Arabic will grow. At the same time language technology can help access information in foreign languages, even without a very good knowledge of these languages. And finally, it can help spread Arabic ideas and culture to non-Arabic languages.

The goal of the **MEDAR** project, supported by the European Commission ICT programme, is to establish a network of partner centres of best practice in Arabic dedicated to promoting Arabic HLT (Human Language Technologies). The tasks of the project include surveying present language resource needs, organizing a conference, disseminating information on Arabic language technology, establishing development priorities and creating a Cooperation Roadmap for the region. Although the project has a special focus on machine translation and other multilingual tools, including cross-lingual information retrieval, the roadmap will be directed towards the Arabic HLT in general, where it will address several areas of interest in the domain; introduce relevance, importance, impact and potential for applications and developments in the various areas of ICT in general, and the upcoming of Global Information Infrastructure (GII) and the Information Superhighways (digital communication systems). The basic building blocks of the GII include: communication facilities, computing technologies, software interfaces/applications and standards tying together facilities, terminals and applications; services (i.e. information, electronic commerce, applications and content) available on these networks; where one of them which provides a base function in the environment is the human-machine interface where the HLT has a central role. If we consider the 90's term Information Superhighway, it was based on providing and dealing with digital electronic content, and HLT is very critical and a "must have" in such an environment.

MEDAR partners have collected knowledge about existing language resources, players, products, most recently initiated projects related to Arabic e-Content by governmental, local, business institutions and/or international organizations, etc., and based on this knowledge and additional research, the present report has been compiled.

It is the purpose of the roadmap to outline areas and priorities for collaboration.

This report is divided into two parts: One for the proposed Roadmap Report, while the other is for Background (sections 5-7), and Appendices (section 8) that contains statistics and other relevant reports.

3. Introduction to the MEDAR Roadmap

Referring to the work done in Europe in particular under the ELSNET project, we can define a roadmap as "a document that indicates directions for a planned journey, that shows how and in what order goals can be reached and that indicates distances".

It is the purpose of the roadmap to outline areas and priorities for collaboration, in terms of collaboration between EU countries and Arabic speaking countries, as well as cooperation in general: between countries, between universities, and last but not least between universities and industry. This cooperation should lead to a stronger Arabic HLT community, more technology for Arabic and more products on the market. Our primary focus is on multilingual tools, in particular on machine translation and multilingual information retrieval but other areas are mentioned as well.

Usually one focuses either on a roadmap as reflecting expected "technology developments and trends" (technology roadmap) or as "time to market" for a new product (market dimension). In our case we will add a new and essential dimension, which is the roadmap for cooperation between Arabic and European Union countries (cooperation roadmap). So, the Cooperation Roadmap will in a sense consist of three interconnected roadmaps although we will not develop each of them independently, but rather take aspects from all of them into consideration.

At present time we see that much ICT cooperation between EU and Arabic partners is based on third party incentives (e.g. the EC ICT programmes). Some other initiatives are conducted in cooperation with the US. It is good that such incentives exist, and we would need more of this type of support in the future, but at the same time we also face a challenge: to turn these partnerships into strategic partnerships, i.e. long term partnerships based on mutual benefit. It is also noticed, that local interest and governmental support is still very limited; the field needs to see more governmental involvement, more interest and support from local ICT main public and private players specially in the areas of telecommunication and mobile industries, more awareness in the software companies to direct some of their efforts into this area, where there are huge potentials for success and market penetration, and more interest in the R&D arena specially at universities and research centres.

Taking into account the three dimensions listed above, we provide below in sections 5 and 6 an analysis and report on the present situation in the participants' countries, we describe the conditions that need to be fulfilled in order to arrive at particular key achievements and at some strategic partnerships and we describe the steps that need to be taken to get there from where we stand. It is very likely that some of these actions (or similar ones) are being implemented by national ICT efforts but we had no evidence at the time of writing. We do hope to consolidate this report with valuable input from other official sources within the Arabic countries.

The proposed roadmap which is a result of the analysis, is given in section 4. For each item in the roadmap, a reference is given in square brackets, e.g. [5.2]. This indicates that the background for this item is given in 5.2.

We invite comments and proposals from all prospective stakeholders for all aspects of this roadmap document (to the email address at page 2 of this document, or to any of the consortium partners).

4. The Proposed Cooperation Roadmap

In this section we give the proposed cooperation roadmap which is the result of the analyses which are given in the rest of this report. Among other actions, the roadmap contains actions on aspects on which the consortium partners and similar actors in the countries have no power, but which are a prerequisite for the other actions to succeed. These actions are for governments, funding agencies etc. to decide on. They have been retained in the roadmap because they are part of the overall plan which lies behind the roadmap.

The scenario is applicable to all Arab states represented in the consortium, with partially overlapping phases; it can be extended to include all other countries that are not part of the consortium. The roadmap until year 2015 is divided into three phases, with a small overlap between phases 2 and 3.

Phase 1 (2010-2012): Laying the foundations

Political/Policy:

- Organise annual round-table meetings between EU and Arab state players and funders, participants in HLT related actions and in funded projects and grants, and initiate collaboration platform to organise knowledge transfer and take-up.
- Develop e-Government projects; at least a few countries should be investigating the possibilities. 50-60% reduction on Internet rates to encourage its use, collaboration scenarios for dealing with reading and illiteracy through linguistic support functions (e.g. reading aloud) and eLearning on the mobile phone and other mobile devices. More enforcements to copyright and IP laws. Tougher legislations to protect software producers from illegal copying.
- Associating and linking the organizations that initiated the projects and initiatives for enrichment of Arabic e-Content in Arab countries with major regional and international players identified by the surveys conducted by the consortium. Consultancy services by consortium expertise to HLT related activities in the region could be offered.

Training and Research:

- Develop initial HLT-enabling curricula, training material, faculty member exchange programs, Arab student placements programs in close collaboration between universities in EU and Arab countries.
- Identify high priority essential *BLARK*[1] components needed for training. Strengthening of participation of players from the Arab states in Networks of Excellence in fields related to multilingual HLT or HLT in general.
- Research helping to cross language barriers: MT, multilingual search etc.
- Develop methodologies to facilitate reading and fight illiteracy through linguistic support functions.

Industry Development:

- Development of MT systems in Egypt, text analysis tools in Jordan, LRs in Egypt and Morocco, and other prototypes or mature products from partners

[1] BLARK is defined as: The minimal resource kit needed to develop basic R&D activities in language technologies

and others identified by the consortium, also other countries have started efforts in Arabic HLT and this will continue to grow. At least one product should be on the market in this phase.

Phase 2 (2012-2014): Moving forward

Political/Policy:

- Development of schemes of cooperation between the Arab HLT players to be more competitive globally, and continuation of adaptation of copyrights and IP laws.

Training and Research:

- Implement first HLT-enabling curricula in participating Arab states. Teaching staff exchanges, grants for Arab students, Arab student traineeships, and continuation of participation in special grants in EU projects.

- Develop essential BLARK components, i.e. tools and LRs.

- Research should support the following areas of application: Automatic speech recognition technologies for dictation, Language learning, Text to speech synthesis in local colloquial, MT for tourism industry, Better Arabic search engines, Bilingual editing software with grammar checkers, Spell Checkers, Dictation machines (domain specific), Translation memories, etc.

Industry Development:

- Focus on: E-Governance and content creation for the general public, and applications via mobile phones for the illiterate and poor readers, Infotainment services and access to entertainment services. Access to educational services in particular for language learning. At least one application per area and per year should be created in collaboration with universities.
- Development of MT systems, text analysis tools, LRs, search engines. At least one product from each country on the market
- Bilingual tools: need more enhancements and focus should be on niches, i.e. be domain specific.

Phase 3 (2013-2015): First consolidation

Political/Policy:

- Local players can penetrate export markets cooperating with their counterparts internationally.
- The dimension of using and developing Arabic HLT supporting utilities will be a serious growth factor if national agencies consider the use of HLT products and services in the literacy programs. Such a huge market will then attract major players. Efforts to obtain this should be made. Tools for readers and poor readers may help provide the opening of this market.
- Cooperation between local software industries and international players can be seen here in terms of outsourcing models where local industries could participate with the international players in implementing products in their local markets.

- Cooperation could be seen also between software developers (companies) with initiatives and projects related to areas where Arabic HLT is core.
- Successful e-Government implementation will be undertaken by some Arab countries.
- Create facilities for HLT entrepreneurship (universities, SMEs).

Training and Research:

- Implementation of improved curricula on the basis of experience gained and new technological developments
- Regular student and staff exchanges between Arab states and EU, and between academia and industry; and some joint projects and training activities across Arab states.
- Further development of the BLARK and creation of application or domain specific resources and tools for priority areas.
- Joint RTD projects between EU and Arab players to build new applications and services, especially related to multi-linguality.

Industry Development:

- Many small companies will emerge, we expect at least 2 new businesses related to Arabic HLT per university per country per year within this phase.
- By 2015 at least one project should have resulted in an educational product for illiterates and/or poor readers on the Arab market.
- By 2015 we expect improvement on the overall books production by utilizing utilities and tools with the aid of MT and other language technology products, raising the average Arabic books production to 5000 books/year.

Cooperation across sectors:

- Arabic HLT will be utilized more and will be incorporated more efficiently in many areas of applications. The cooperation will be seen in the following areas:
  - Institutional consumers (e-government, e-health, e-… agencies) could adopt the HLT products and provide web-services for citizens.
  - More R&D activities will take place to enhance the products and tools started in Phase I and II, and to come up with e.g. advanced search engines, speech enabled services, multilingual tools to support cultural and tourism sectors, etc.
  - Start adding languages to enrich the multilingual dimension on both the applications and tools/utilities levels.
  - MT prototypes should develop to more mature products especially considering closed domains for specific applications.
  - More enhanced applications and utilities: Text to Speech applications could be very useful in this area, since the public needs to access such services, and given the high degree of illiteracy among citizens in some countries.

Summary of Roadmap tendencies

Development and growth will continue throughout the duration of the period specified for the roadmap (till 2015), in all areas addressed in this report, namely those that have effect and will impact the Arabic HLT in general and therefore will affect the

roadmap proposed in this report. As a result, **cooperation** should be sought binding and linking all players together; this will result in cooperation between all the players to build an Arabic HLT industry; that is encourage the major international players to keep and maintain their interest as well as to support and build the local industry and workforce needed to sustain it.

The following summarises components and directions that will lead into success of the strategy:

- Specialized companies should play a significant role in this area and should build and enhance tools, utilities and applications for Arabic HLT,
- Universities and research centres should provide the basic and applied research in cooperation with industry to produce solid products,
- Governments and funding agencies should facilitate, support and help companies and universities to initiate and sustain their products,
- Governments should launch services/applications for citizens (e.g. e-government sub-projects, initiatives) that will be accessed and navigated in Arabic language
- Universities and other educational institutions should collaborate to create the proper training and re-training (rehabilitation program for personnel from other disciplines who could be re-trained to fit the new requirements)
- International companies specialized and interested in the HLT (and Arabic HLT) in particular should be encouraged:
  - to maintain the interest in Arabic,
  - to providing services to the region and should be given the facilities to make this attractive for them,
  - to maintain relationships with local companies and task forces, and
  - to utilize what is available locally
- Local mobile companies, internet service providers and telephone companies should provide the support and encourage the local companies and universities to direct their efforts towards producing tools and utilities that could be integrated and added to the provided services.

5. Elements of the roadmap

There are a number of elements that all contribute to the current status and the future development of Arabic language technology. In the following we will describe the status for each element and the target in 5-7 years.

Key elements are of course the players, i.e. the organisations involved and the human resources they have at their disposal. Human resources have to be highly skilled, and therefore education is another crucial factor in the roadmap. The next elements that are described are the technology, and the environment (or infrastructure) in which the technology operates and that can be used to disseminate it. We then discuss the market for Arabic HLT (products and services) within and outside in the region, as it is important that the market has a certain size for players to find it attractive. Many of these elements are interrelated, a fact that may give some redundancy, but we hope that the main trends will emerge clearly.

5.1 Players and human resources

In the universe of HLT many different players fulfil their tasks in the long chain from research idea to end user product or service. In this section we will highlight a few of these players, which can be both organisations and humans. In MEDAR Report 3.1 [*Survey of actors, projects, products*] we have brought together a large number of players (or potential players) connected to the field of Arabic HLT, with a special focus on players active in areas that are concerned with multilinguality, such as translation support, machine translation and cross-lingual information retrieval. In the Appendix 2 to this report we have tried to classify them on the basis of their role in the chain in order to get a better understanding of the Arabic HLT landscape, but borderlines between roles are not always easy to draw. The table contains both organisations in the Arab countries and in the rest of the world. It does not claim to be complete although we have tried to be as complete as possible for the countries represented in the MEDAR consortium. It will continue to be updated throughout the duration of the project and, if possible, beyond.

In the table appears in the appendix, we have been able to identify 42 organisations in the Arab states active in HLT education (20), research (29) or development (11). Some are involved in two or three of these activities, which explains why the sum is larger than 42. For a language ranking number 5 in the world, with as many as 280 million native speakers and 250 million non-native speakers (according to http://en.wikipedia.org/wiki/Arabic_language) these numbers are modest.

If we look at the human resources side we can observe that many of the skills required for the production of HLT products and services coincide with those required for ICT products in general (e.g. software engineering, testing, interfaces). We have not systematically investigated the availability of these generic skills in the Arab states since this would go beyond the scope of this report and reliable figures are extremely hard to get (if at all) from existing sources.

The MEDAR Survey (http://www.medar.info/MEDAR_Survey_I.pdf) shows that the number of Arabic HLT professionals is very low, and by no means sufficient to maintain (or even start building) a strong HLT industry in the Arab states. This was confirmed by a recent survey we made through our project partners in the countries

represented in the MEDAR consortium, where all partners without exception reported a shortage of people with Arabic HLT skills.

In this report we focus on the special, additional skills required to build HLT for Arabic, which include both knowledge about Arabic language and linguistics, knowledge about language and speech processing, machine learning techniques, signal processing, statistics, cognitive sciences, for linguists the capability to communicate and collaborate with software engineers, and for software engineers the capability to communicate with linguists. If we want to increase the number of people with the required skills we have to look at the opportunities for the education of a new generation of researchers and developers with adequate skills in HLT.

The main players in the education system are universities and other institutes for higher education. In our survey (conducted in the MEDAR partner countries) we have tried to gather information about the total number of universities (excluding specialized institutes such as medical, agricultural or veterinary institutes) and about the number of those that have HLT oriented courses in their curriculum on a structural basis (i.e. not as side-effects of the presence of individual researchers with a special interest in HLT).

From the figures quoted above from the survey we have seen that the number of institutions offering HLT education is minimal, and our recent survey confirmed this. Below we will discuss what needs to be done to improve the situation.

In our view the education system should aim at providing HLT training both to students who want to graduate from university and to professionals who are already working in the ICT field but who lack specific knowledge about HLT and language in general. In addition to this there is also a need to train people to become HLT educators, as this is necessary for a sustainable supply of *HLT-enabled* professionals. We use the term *HLT-enabled* since we do not believe that the main goal of the education system should be to create a completely new HLT discipline with its own professionals, but rather to provide people who have a firm basis in one of the fields relevant for the advancement of HLT with an additional component of knowledge and skills that allow them to use their specific skills to contribute to the development of HLT related products or services. Typical examples would be courses in e.g. linguistics, phonetics, language or speech processing offered to software engineers, or (vice versa) courses in e.g. computing, machine learning, language or speech processing offered to linguists or to students in these fields.

The underlying assumptions are (i) the experience in European universities has taught that there is little point in trying to convert a linguist to become a software engineer (or the other way around) as it would lead to under-exploitation of a person's specific talents, and (ii) the complexity of language and speech processing is such that only teams where all disciplines are represented (and where team members are sufficiently familiar with each others' fields to be able to communicate) can be successful.

Apart from the complexity of the problems and the view that only mixed teams can tackle HLT problems properly there is another, practical reason why it is not recommended to try to design special HLT curricula leading to a (necessarily deep but narrow) university degree in HLT. If one looks at the employment perspectives for

highly specialized HLT engineers it is obvious that the currently relatively small market for HLT products and services in the Arab countries would not encourage students to go this way.

The best strategy to attract students who aim at becoming HLT-enabled will probably be to make HLT more visible in existing curricula in relevant fields, such as computer science, information science, AI, and linguistics and to offer interesting courses that are adapted to the pre-existing knowledge the students have and that (at least equally importantly) show the students clear links between what they are doing and what HLT would need or offer.

In order to accommodate both university students and professionals in the field it is important that e-learning facilities and part-time courses be made available in addition to standard university programmes.

A good example in this area to take into account and to develop is the foundation in 2004 of the ISTA (L'Institut Supérieur Arabe de Traduction). This post-graduate institute was set up as a scientific and educational institution under the sponsorship/patronage of the Permanent Secretariat of the League of Arab States in Algiers (Algeria). Its activities cover teaching, at the Master level, of translation and interpretations curricula, research projects related to translation and language technologies, and regular translation activities. The remarkable aspect is the cooperation with a large number of instructors/teachers from several universities (Algiers, Casablanca, Beirut, Sharjah, Paris, Lyon, Ottawa, etc …).

Any course activity related to HLT requires at least written and digital material. As for written material (such as text books, journal articles, conference proceedings) we can observe that most of the material used in HLT courses are in English (or in French), and that very little material dedicated to Arabic HLT is available in any language, and almost nothing in Arabic.

As for digital material we have already described in the *BLARK for Arabic* document [MEDAR Report 3.2] and the *Analysis of gaps* [MEDAR Internal Report 3.3] what the basic set of language resources required for education should contain. In the same reports we also gave an account of what is already available (and accessible) and what is still missing.

If such material is made widely available, it is very likely that EU universities will feel encouraged to offer, as part of their HLT curricula, applications and highlights that include Arabic resources and tools which could pave the way to future collaborations.

Recommendations

On the basis of the observations above it appears that there is a need for initiatives which could lead to an improvement of the Arabic HLT education situation. As one of the main objectives of this study is the creation of a collaboration roadmap we will list here a number of possible bilateral or multilateral cooperation actions between EU and Arab states (with support from e.g. the EU framework programmes), or

cooperation between Arab states (supported by agencies in the Arab states) in the form of what in EC programmes is described as Coordination or Support actions:

- Joint training of teachers
- Common curriculum or course development
- Joint schemes for industrial placements or traineeships for students
- Joint development of teaching material (new material or adaptation of existing material)
- Joint development of e-learning courses

or in the form of RTD projects:

- Joint creation of (generic parts of) the BLARK, where many tools might be ported from other languages in collaborative actions

5.2 Technology and R&D activities

Successful R&D leading to the development of HLT-related products and services does not only presuppose the availability of organisations and human resources as described above, but also a wide range of other key ingredients. We list a number of them here and give a brief overview of the present situation and of where we see possibilities for joint actions.

The first aspect is a strong and clear commitment from the Arabic governments to support R&D activities in general and those devoted to ICT and HLT in particular. This is not the case all over the region (we see some useful signs within the Gulf region). As for the real statistics, reliable sources are missing and intricacies of methodologies make them hard to compare. It is clear from our survey that the budgets involved are far behind the amounts devoted by the developed and emerging countries to this area (see for more details:
http://www.unu.edu/unupress/unupbooks/uu09ue/uu09ue00.htm#Contents).

From the cooperation point of view priority should be given to tools and resources (i) that are re-usable, (ii) that can be shared with other players, and (iii) that adhere to formal or de facto representation and interoperability standards, so they can have a maximal impact. The survey shows that at present time external standards (formal or de facto) do not play a significant role in Arabic HLT. This is a serious obstacle for any form of collaboration.

From the specific MEDAR perspective, where the focus is on cooperation, a number of observations can be made. First of all the universe of HLT players in the Arab world is small and fragmented with a majority of small players (as far as their HLT efforts is concerned) and a very few larger ones [see *Survey of actors, projects, products*, MEDAR Report 3.1]. Many of the available HLT resources and applications originate from outside the Arab world, some from large global players such as Microsoft, IBM and very recently Google got interested in this area.

The global players (some of which already have local branches in the Arab world, especially in Egypt, such as Microsoft, IBM and Orange Lab) are in an excellent position to establish relationships with local players or even to acquire them if the global players feel they need their technologies and/or tools/applications.

In this fragmented landscape it will be extremely hard for other local players to enter into competition with the big players, so they need to join forces. The European experience has shown that in spite of the linguistic fragmentation in Europe many small players have entered into lasting cooperation relationships leading to a strengthening of HLT in Europe. In the Arab world, politically and economically less united than Europe, this may not be as easy, but at the same time it should be noted that for Arabic HLT an enormous cooperation potential follows from the fact that in MSA they share a common written language, and that even in the case of spoken language, where many local dialect variations exist, many technologies can still be shared across dialects.

The BLARK [see *BLARK for Arabic*, MEDAR Report 3.2] has already been mentioned in the previous section as an important facility for the education of researchers and developers. Although its generic nature will most probably not make it suitable for the development of specific end user products and services it will be a useful instrument for researchers at the pre-development stage as well as for pilots and demonstrators. The present state of the BLARK for Arabic, including gaps and priorities is described in the above document and in the *Analysis of gaps* [MEDAR Internal Report 3.2], and it is recommended to take initiatives to complete it.

In addition to the BLARK a variety of language resources and tools will be needed for the development of specific applications and services. In the internal MEDAR Report 3.3 [*Analysis of gaps*] an overview is given of the tools and resources that are missing. We recommend undertaking actions to fill them. Much attention should be paid to tools for the creation, acquisition and annotation of new resources, and the training of staff to use them. This will help to fill the identified gaps and it will also support the creation of new resources for future applications and services.

Cooperation, especially at the international level, is very unlikely to happen spontaneously and cannot be enforced either, but is an absolute necessity if local players want to continue to coexist and collaborate with the global players. This collaboration could be based on national or trans-national R&D programmes in the Arab states, but in addition to that cooperation in the context of the EU programmes could be a key factor here in that they offer both the organisational framework and the funding possibilities for cooperation between EU and Arab players. In view of the MEDAR focus on multilingualism the members of the consortium strongly recommend actions especially in this field as it offers opportunities for cooperation between EU and Arab country partners to form consortia with complementary skills, and where partners from different Arab states would work together. This will not only create lasting partnerships between Arab and EU players, but also between Arab players.

- Focus on crossing language barriers: MT development, multilingual search etc.

Cooperative projects could take many different shapes, such as medium size and large RTD projects (STREPs or IPs in EC jargon), participation in Networks of Excellence, Coordination and Support actions, but also staff exchanges involving both academic and commercial organisations' leading to better transfer of knowledge between academia and industry end between EU and Arab partners.

As already stated above representation and interoperability standards are an important instrument to ensure re-usability and sharability of resources and to create opportunities for cooperation aimed at integration of tools and services with a view to offering more advanced products and services. It is therefore recommended that players from the Arab states be offered ways to participate in EU-wide actions aimed at the creation of infrastructures to share resources and technologies, and at providing recommendations for policies regarding language resources and technologies (e.g. CLARIN, FlaReNet, ELRA), where standards play a crucial role. Without that they would isolate themselves from what is going on in mainstream HLT in Europe.

Recommendations

To summarize and conclude we list a number of typical examples of possible actions aimed at strengthening the advancement of Arabic HLT and at creating lasting partnerships, and that should be included in the HLT Cooperation Roadmap for Arabic HLT.
This includes Coordination and Support actions aimed at:

- Strengthening of participation of players from the Arab states in Networks of Excellence in fields related to multilingual HLT or HLT in general
- Participation from the Arab HLT community in HLT related actions concerned with language resources infrastructures, standards and policies
- Developing schemes for staff exchange
- Developing schemes to bring players from the Arab HLT community together in order to make them more competitive vis-à-vis the global players on the Arab HLT market (incentives could be tax deductions and fiscal advantages).
- Offering attractive conditions to R&D labs which are relocating to off-shore countries

And it would include joint RTD actions involving EU and Arab partners aimed at

- Creation of basic resources (development of the BLARK)
- Creation of application or domain specific resources
- Creation of HLT products, applications and services, especially in the field of multilingual technologies

5.3 E-infrastructure

An important factor for the further development and use of Arabic language technology is the availability and penetration of enabling technologies, in general ICT infrastructures and in particular internet access, available within the Arab regions, as these technologies are a prerequisite for the development of the market. The global information infrastructure (GII) is the developing communications framework intended to eventually connect all telecommunications and computer networks worldwide. The GII would eventually make all electronically stored or transmitted content accessible from anywhere on the planet, and HLT will definitely play a critical functional role in accessing and manipulating the digital content. Translation utilities are needed to make the multilingual content available in the local languages; in our case MT will definitely play a critical role in providing some kind of Arabic equivalent of the content (even if initially it will be weak from the language point of view). Tools and applications for Arabic IR will play a better role and may achieve better success rates. A special attention should be directed towards search engines, they will play a significant role in documents retrieval taking into account the special

linguistic needs and requirements of the Arabic language. It is noticed that this technology still needs a lot of work and concentration since what is applicable to Latin languages, have fallen short from satisfying the Arabic language documents and terminology searching techniques. The future will witness a lot of advancements and quality of the outcome will definitely be improved.

Internet

Less than ten percent of people in the Arab world are internet users; according to a study (March 2008) conducted by Internet World Stats (www.internetworldstats.com) the internet penetration percentage in the Arab region is only equal to 9.4%. Internet penetration in Arab Region is still very low, when it is compared to the percentages in the developed countries which reached 88% in Norway, 85% in Iceland, 65% in Germany, 66% in the UK, 69% in Japan and 71% in North America (See Appendix 3).

The highest percentages for PC penetration are in the Gulf Cooperation Council (GCC) countries where the average percentage is equal to 25.37%; the chart in appendix 3 shows the variance in the PC percentage between all Arab countries.

As mentioned above, it is evident that the Internet penetration in the Arab countries is much less than in e.g. Europe in average. But it is also important to consider the growth rate, and here several Arab countries show a very good growth rate for the period 2000-2008. It is true that growth rates of several thousand percent appear only when the penetration has been very low, and such growth rates cannot be maintained when a higher penetration has been reached, but we believe this shows a positive development for most countries. Several Arab countries will reach a reasonable level compared with other regions if they can keep up the speed to a reasonable extent. The growth rates for the Arab countries are higher than for Europe in average, and considerably higher than for countries with a large penetration, such as e.g. the Nordic countries (Norway has a growth rate of 82% and Denmark of 126% for the same period).

Note: The June 30, 2010, Internet World Stats shows that the Arab speaking part of the Middle East region had reached an Internet penetration of 28 % and still had a growth rate of more than 1000 % in average, so it looks like this development is very positive, even if this is only slightly above the world average of 26 %, and even if some countries are still at less than 2 %.

In a report from the Economist Intelligence Unit of 2008, "*How technology sectors grow: Benchmarking IT industry competitiveness*", showing the IT industry competitiveness index, only three Arab countries could make it in the 40th (Saudi Arabia), 53rd (United Arab Emirates) and 65th (Egypt) ranks, amongst the overall scores and ranks of 66 countries. Figure 4, in Appendix 3 shows details of the ranking including other top 10 countries for the years 2007 and 2008.

A lot of efforts are needed to improve the ranking of the Arab countries, and to be competitive in this sector enhancements to human resources and IT methodologies and infrastructures are needed in almost all of the Arab countries.

There are several ways to support the growth in Internet penetration. One way is to reduce the rates that users have to pay; e.g. the Internet rates in Jordan have recently been reduced by 30%, and the same was done in Saudi Arabia and other countries. Egypt is a pioneer in this regard as very low cost dial-up Internet (at the cost of a local landline phone call) is available for everybody since 2001. Another way of boosting the Internet penetration is for the governments to increase their services on the Internet, i.e. developing e-Government. This is a well-known method that has proven its effectiveness in other parts of the world. However, e-Government can only take off when a reasonable amount of users are online, so maybe only some of the Arab countries are ready for this development. We hope to see more leading roles for workers in the Arabic HLT to help in enriching the Arabic e-Content on the internet. As MEDAR is concentrating on MT, this will contribute to populating the Internet with Arabic content. Moreover, the interest of the project is also in the IR domain and such domain will strengthen the usage and access for Arabic speaking people to the internet. Search engines of all types will be potentially important to be available to Arabic users.

- Try to focus on eGovernment and content creation in order to reach the general public (eGovernment requires focus on security and on secure payment).

Arabic e-Content on the Internet:

Cooperation on Arabic HLT between research centres, international organizations and industry (both on the local and international levels) is taking place in many countries in the region. Appendix 1 gives examples of some initiatives in some Arab countries as examples of the important initiatives in this line, and we will briefly highlight some of them in this section.

As one good example we can mention the "King Abdullah (of Saudi Arabia) Initiative for Arabic content", where efforts are directed towards enrichment of the Arabic content in the Internet, and launching projects to achieve the target and the objectives of the project. The project is aiming at open source based solutions, and looking to avail all results for public use. We believe that the directions MEDAR is taking in the two domains that were selected, namely: MT and IR will definitely be of use and benefit to the project, and will go in line with helping in this direction. We can see there is room to expand on this initiative, and propose several directions to contribute to enriching the Arabic content. In a very recent press release at Riyadh Saudi Arabia, on June 10th 2009, "*King Abdullah Initiative for Arabic content has recently adopted a number of important projects for the advancement of Arabic content on the internet (e-Content) and for the preservation of identity and heritage”*.

In Syria, eSyria {http://www.esyria.sy} is a national initiative launched by Syrian Computer Society for Informatics in Syria, this initiative aimed to enrich the Syrian national content on the internet and raise the level of electronic services, also, provided to citizens in Syria with the changing needs of information and useful for all events and activities in all provinces of the country. This project aims to increase the Arabic content on the Internet and the Syrian special services.

A national conference on Digital Arabic Content Industry with a view to searching for the DAC's national strategy and to setting up an incubator and a share structure for the

industry was held. It was a result of cooperation between Communications and Technology Ministry and the Syrian Computer Society (SCS) in cooperation with UNESCO's Representatives in Syria. This is a cooperation effort between a local government, professional society and International organization.

Arabeyes is a Meta project that is aimed at fully supporting the Arabic language in the Unix/Linux environment. Arab Science and Technology Foundation (ASTF) is an independent, nongovernmental, non-profit regional and international organization. The Foundation aims to identify and support outstanding scientific research activities being conducted by women and men of science and technology from the Arab world. It also attempts to act as a mediator between those who produce, develop, finance or benefit from scientific research, while seeking to become a centre for assessing the performance of scientific programs and also a powerful international Arab entity that defends the region's interests in scientific and technological progress. Dahsha (www.dahsha.com) is an Arabic online encyclopedia and one of the serious efforts to enrich Arab content on the internet. A lot like Wikipedia, the encyclopedia relies on volunteers to write articles and publish content.

Talal Abu-Ghazaleh Organization (TAGorganization www.tagorg.com/) recently announced the launch of the first Arabic electronic encyclopedia (TAGIPEDIA), the first initiative of its kind that targets Arab academics, scholars and readers, and provides them with general knowledge and sciences from an Arabic cultural, historical, and educational perspective. TAGIPEDIA aims at founding a free-access database that delivers a broad variety of knowledge in all humanities sciences, with the highest standards of quality, integrity, and accuracy. It also gives the chance for readers/users to augment the content through contributing to the Encyclopedia without prejudicing the rights of its original author.

Collaborative efforts resulted in the creation and initiation of the Fekr Rama (www.fekr-rama.com) project in Egypt. It is a new web portal that provides rich Arabic content in the fields of culture, education, heritage, religion, sport and the arts for Arab and other audiences. A collaborative action between the Ministry of Communications and Information Technology (MCIT), Al Azhar, the Ministry of Culture, the Ministry of Information, the Egyptian Publishers Union and the e-Learning and Business Applications Union, the portal was launched in March 2008.

On the 21st April, 2009 - UNESCO and 32 partner institutions, including Bibliotheca Alexandrina launched the World Digital Library, a Web site that features unique cultural materials from libraries and archives from around the world. The site www.wdl.org includes manuscripts, maps, rare books, films, sound recordings, prints and photographs. It provides unrestricted public access, free of charge, to this material.

A report by The United Nations Economic and Social Commission for Western Asia (UN-ESCWA), released in 2009, for evaluating the progress and ranking of Arab countries in relation to ICT, Information Society and Internet utilization, with emphasis on content and promotion of cultural activities. There are some projects identified in the report, we think that there are areas of intersection between those projects and our interests, we can utilize our resources and capitalize on the available opportunities to build many cooperation agreements. We think UN-ESCWA represent

a strong and good potential to be in alliance with and to extend our network for cooperation to include such an organization.

The greatest number of such initiatives however seem to belong to the United Arab Emirates and Qatar, which are both aiming to become an active cultural pivot as far as preserving and promoting Arabic culture and heritage are concerned.

Dubai-based research and consultancy company, Madar Research Group, has recently conducted a study on Arabic content on the Internet (both in Arab countries and abroad) using a methodology based on world search engines. The results showed that the number of Arabic web pages is still considerably low compared to the overall number of web pages published in other languages.

Summary on Arabic e-Content initiatives

Above we have seen examples of initiatives and leads in several Arab countries in the areas of e-Content which show the potential as well as the need for Arabic HLT. Such an effort needs a lot of cooperation between industry and government, and it needs tools and utilities specially in searching mechanisms and Information Retrieval systems. This is another effort which the roadmap is trying to address, and this could be identified as one major project for cooperation. Those efforts are in the same direction that we are trying to identify, it can clearly demonstrate two main points:

- It emphasizes the importance of the areas the roadmap is trying to identify and build activities directed towards achieving the main goals.
- It is a proof that the efforts are scattered and that the need to cooperate is very much needed.

It is excellent to see these Arab teams investing in knowledge and working to create more Arab content online. Those projects are very critical, they have a significant role, and they all need a lot of cooperation and collaboration between many groups from both the academic and business sides, in addition to many tools and utilities to simplify and facilitate the job, such as MT and IR, which are the main concentration of the Roadmap. They could play a major role, and we could find many cooperation opportunities in those projects and initiatives. We believe that these goals of more Arabic e-Content should still be pursued with focus and strength, and more countries and groups should support the directions so that the goals could be achieved faster, with many steady steps in the same direction.

Those initiatives, especially the ones initiated by ESCWA and UNESCO, in their meaning and mechanisms are manifesting the coincidence and similarity of their actions and what we are proposing to achieve. They clearly demonstrate that such achievement cannot be realized without the cooperation and unification of efforts from governments, international bodies, countries from all over the world, and room is open for such groups like ours and many others from the industry and private sectors. It is the hope that our roadmap could participate in contributing to such efforts and provide significant tools and utilities to enhance and support such impressive gigantic projects. Technologies and cooperation could be extended to initiate projects in this area, utilities and tools such as: MT, with address to multilinguality, IR, search engines, Text to Speech systems and utilities, involvement

of partners from the local and international sides could also be of interest to all parties.

Suggestions and recommendations on Arabic e-Content

The various initiatives conclude with a set of recommendations, and from our perspective, we can see a significant role and contribution in terms of cooperation and utilization of the technologies and trends discussed in the roadmap, our part is within the recommendations; by "our contribution" we mean collaborative efforts from the local as well as European partners and major players in the domain; it is worth mentioning that the funding agencies should be approached and encouraged to fund and support such projects and initiatives:

1. Accelerate the implementation of e-government projects through the allocation of additional funds, since these projects specifically increase online Arabic content;
2. Set up effective governmental programmes that aim to develop Arabic content, establish non-profit Arabic digital information banks, and crystallize a plan for regional cooperation in this respect;
3. Support individual projects related to cultural and linguistic diversity through free hosting of these website projects and furnishing them with financial assistance to sustain their activities;
4. Increase governmental infrastructure investment to promote use of the Internet. Focus on the Internet as a distinct advertising environment which can convey the product message to the consumer in a more effective way than the traditional methods, which will help develop commercial content;
5. Accelerate the legislation of laws and statutes related to the rights of digital publishing;
6. Alleviate controls and minimize strict filtering of websites in some ESCWA member countries as much as possible;
7. Provide an enabling environment that motivates the private sector to participate in establishing the Arabic content industry;
8. Focus on e-learning content within the framework of a public strategy for lifelong education;
9. Focus on the media of the local groups and consider the cultural diversity within Arab countries as a support element in developing the Arabic content industry (in particular set up an incentive policy to help them move their content to the Internet and digitalize their archives).

Our contribution with those initiatives can be summarized by the following:

- Consultancy services and providing expertise knowledge for the domain of Arabic HLT
- Associating and linking those organizations with major regional and international players
- Identifying prototypes or mature products from partners and others, supported by the consortium, like: Search engines, IR tools, MT, …etc
- Making available MT prototypes that would develop to more mature products especially considering closed domains for specific applications; multilinguality could also be considered here.

- Promoting the use of Text to Speech applications which could be extremely useful in this area, since the public needs to access such services, and suffer the high degree of illiteracy among citizens in some countries.
- Local education material for development of e-infrastructures for companies and universities.

Information and Communication Technologies (ICT):

As for the ICT in general in the Arab countries, although ranking is acceptable among the developing countries, but again much efforts by governments are needed. The ICT Development Index (IDI) 2002–2007 by ITU (International Telecommunication Union) *Measuring The Information Society Report*, 2009, given in Figure 5 in Appendix 3, shows the ranking of 154 countries from all over the world, including 17 of the Arab countries.

Arab countries ranking according to this index came as a sort of reflection on the economic strength and financial status (to a certain extent). The GCC countries scored the highest among the Arab countries: UAE 32, Bahrain 42, Qatar 44, Saudi Arabia 55, Kuwait 57, except Oman came at rank 77; then came Lebanon 64, Jordan 76, amazingly Palestine scored 79, while Libya 81, Tunisia 83, Syria 89, Egypt 94, Algeria 97, Morocco 101, Sudan and Yemen ranked the lowest at 120 and 124 respectively.

Mobile phones

Mobile phones (and PDAs) are considered today as an integral part of the ICT boosters. The mobile phone industry in the Arab countries shows aggressive growth in almost all of the countries. Many providers who operate their services in the region are local and are available in many countries, providing facilities and cheaper rates to their subscribers. In Appendix 3 we quote reports and statistics related to the Mobile phones business in some Arab countries, to demonstrate the penetration rates, extra ordinary expansion and large spread in the related countries. The reports and statistics along with the dates for each report were taken from:
http://www.itu.int/ITU-D/ict/newslog/CategoryView,category,Arab%2BStates.aspx

A new report from Arab Advisors Group released on 2/12/2008, analyzes and ranks 30 fixed services operators and 50 cellular operators in nineteen Arab countries. STC's Al Jawwal, Egypt's Mobinil and Vodafone Egypt are the largest Arab cellular operators in terms of subscribers. Cellular subscribers in 19 examined Arab countries reached 194.533 million. UAE recorded the highest cellular penetration rate by H1 2008, which stood at 198.6% followed by Saudi Arabia (123.3%). UAE also had the highest fixed line penetration rate by H1 2008, which stood at 29.4%.

We do consider mobile telephony as a key instrument and a crucial media that could be exploited to enhance Arab countries access/supply of information and Multimedia Content. Given the rates of illiteracy the use of mobile devices and speech input/output (Automatic Speech Recognition and Text to Speech Synthesis) could play a very efficient role.

- Linguistic support for the illiterate and poor readers via mobile phones.

Recommendations

The IT competitiveness and ICT readiness still needs a lot of efforts and collaboration among the public and private sectors; education should also play a significant role in terms of IT-readiness and IT-competitiveness, by enhancing the educational curricula and by bridging the gaps with the industry. Moreover, governments need to improve their incentives towards the IT industry to allow them to compete locally (in the region) and internationally on the global level.

From the reports and statistics for some of the Arab countries listed above, it is noticeable that the mobile phones usages are growing at a very high speed, and many companies are looking at more penetration through enhancing services and providing clients with an edge and added value for subscribing with their networks. One good reason will be to utilize language technology into mobile added services, and by approaching the mobile companies to support R&D in this direction.

In order to ensure that governments, when deciding about their national ICT policies, are made aware of the opportunities that might be offered by collaboration with EU partners through present or future EU programmes we recommend to investigate the possibility of organising annual or biannual round table meetings where the following groups should be represented:

- Arab HLT and ICT players
- EU HLT and ICT players
- Government agencies in the Arab states
- Other funding bodies in the Arab states

These meetings should result in joint recommendations to governments and the EU which would open up possibilities for coordinated collaborative actions. They could be organised on a national or regional basis, depending on what seems most effective.

- Round table meetings

Many of these desired developments do not lie immediately in the field where MEDAR can act. But as we believe they are necessary we have above suggested collaboration in sectors dependent on this development.

5.4 Market

Arabic Market(s) and HLT

This section will briefly elaborate on the existing market and its different components (suppliers, consumers, impacting factors, business models). The crucial question is how HLT would help develop the Arabic region from various facets: technological, societal, economic, education and at the end its GDP. It is also important to address the HLT market per se and envisage various scenarios of its own development.

Let us summarize herein our views on these aspects.

It is essential to consider that LRs and LTs are the key components of HLT and are mandatory for any deployment of applications. Furthermore, HLT are becoming key components of a large number of IT and ICT based applications. So all assumptions about ICT will impact and be impacted by HLT. For instance the mobile industry may

grow even more (see statistic in this report) if users can easily send SMS messages (voice dictation, predictive typing, etc.) which are typical HLT applications.

Economic growth may be boosted by the use of HLT to improve competitiveness but also to supply HLT components. The Arabic regional markets (see below for more descriptive details) can be characterized through a small number of activities. The major development trend these days, in almost all Arabic countries, is the off-shore activities. Such activity requires essentially reduced labour costs, high skills and competencies of human-resources and both geographical and cultural proximity. The off-shore market may contribute to the development of the GDP, through its services in very profitable niches (e.g. HLT as a component of Call centres, semi-automatic conversion of voice-messages to SMS and human validation, etc.), but also through some product developments for the targeted markets (development of MT systems in Egypt, text analysis tools in Jordan, LRs in Morocco, etc.[2]).

This is a very competitive arena and according to Gartner, when reviewing the software outsourcing industry, Morocco and Egypt are the first Arabic countries to enter the top 30 attractive countries. Arabic countries are competing against countries like Czech Republic, Poland, Romania, Hungary (which benefit from their European Union status) but also Israel, Mexico, Vietnam, Malaysia, Thailand, etc.

Among the criteria used by Gartner we can list spoken languages and language proficiency, political stability, geographical proximity, time zones, culture, educational system, the official support (grants and/or fiscal deductions, venture/public capital), available skilled human resources, etc.

Within the economic demand of the offshore markets, some local players may realize that they can also benefit from similar services and products and develop an extra demand to be fulfilled by the local players. The last dimension is that local players may also feel that they can penetrate export markets independently from the off-shore policy.

Domestic versus International markets

The market foreseen here comprises the Arabic domestic market which could be addressed or penetrated by commercial offers of products and services but also foreign market that could attract exports from players located within the Arab countries. We will also consider the different suppliers and users: those can be located in Arab countries and supply HLT products/services to the markets mentioned above, but could also be located outside the Arabic region and supply products/services to Arabic markets as well as international ones. The idea of the project is to primarily identify scenarios to boost the offers of Arabic players (for domestic and export markets) but also the international offers of Arabic HLT for the Arabic markets.

[2] Activities of Orange Labs, Arabtext, MLTC, etc.

This is illustrated here:

Players	From Arab countries	From outside Arab countries
Markets		
Arab countries	Yes (e.g. Sakhr, Arabtext)	Yes (Systran selling to the Arab League)
Outside Arab countries	Yes (Sakhr selling in France)	Yes (Systran selling to USA agency an Arab=> English MT)

Fig. 1. Players and markets

It is clear from some of the countries' policies that the favourite scenarios are to encourage local productions mostly for the international market; As a side effect, such players may also serve the domestic consumers; Let us list some examples of such services. It is very likely that a service offering the recitation of the holy Quran, with speech input and speech output, could be profitable. A similar service using web-based tools with IR and/or CLIR would be also very much used. We can also imagine similar services for music, movies, etc.

Services versus product distribution markets

In our analysis we also distinguish the market of product distribution (buying an off the shelf MT by a user) and the market of services (access to a web-based MT service). In some cases it is hard to "locate" the service supplier or the user.

Such players (and others involved in HLT development for other languages) will be encouraged to address the needs of Arabic language and the corresponding markets only if the demand is strong enough to offer a good return on Investment. Such demand is interlinked to the ICT infrastructures offered within the country and the capacity of the users to purchase and use them.

Therefore we have to take into consideration the profiles of our consumers/users of HLT products and services. We distinguish at least three categories:

- individual consumers,
- professional/business users, and
- institutional users.

Such categories have to be located within Arabic countries and outside. It is clear products for individual users (e.g. spell/grammar checkers, MT[3], TTS, ASR, etc.) will be impacted by the software business in general (acquisition of software packages). Such profiles are listed below considering products versus services adopters.

[3] MT: Machine Translation, TTS: Text-To-Speech (also Speech synthesis), ASR (Automatic Speech Recognition)

For product acquisitions: a crucial issue is the "buying" of intellectual products in general and software packages in particular in the Arabic world. This is an issue that could be interpreted along the usual dimensions (usefulness, usability, etc.) in addition to a large number of factors: poverty, illiteracy, trading customs.

This is highlighted by the annual survey of the Business Software Alliance (BSA) conducted with IDC about the software piracy (the 2007 BSA and IDC Global Software Piracy Study). This survey covers piracy of all packaged software that runs on personal computers (PC). Our HLT products when packaged for PCs e.g. MT, Dictation, etc. are part of this.

The worldwide PC software piracy rate increased three percentage points to 38% from 2006 to 2007 (meaning that 38% of the packages are not "paid"). Yemen is the first Arabic country (ranked 7 with 89%), Libya (ranked 8 with 88%), Algeria ranked 14 with 84%, UAE is within the low rates (with 35%). When Egypt's government made a deal with right holders to provide software packages for government and educational use, piracy rate dropped to 60%. Similar efforts were taken in Jordan (60%), Saudi Arabia (piracy dropped to 51%). Tunisia piracy rate is about 76% while Morocco is 67%.

This illustration should also be interpreted as an indicator of the whole market size. With 60% of piracy and losses of 2,446M$ we estimate the whole market at 4,000M$ which is certainly not big enough for a mass-market (but may be attractive for niche players).

International players may be discouraged to invest in such markets while local software industries can be crippled by competition from pirating actors. IDC finds that "*most of the benefits that result from lowering software piracy accrue to locally based resellers, software services, and channel firms–meaning the greatest proportion of economic benefits from lowering software piracy remain within a country*".

- These analyses are in agreement with the trend to offer most of these applications as web based services. This is also the big trend in software industry.

The other dimension mentioned above is the illiteracy. This has an impact on the global Arabic economy and the PC/Internet penetration (see the E-infrastructure section of this report for more details). This dimension can be also taken as a serious growth factor if national agencies consider the use of HLT products and services in the literacy programs. Such a huge market could then attract major players.

Some consumers from "privileged" classes could better adopt such technologies as we see from the internet and mobile penetration. Here we assume a "customer" behaviour comparable to what we see in regions like the EU and USA. Today's business behaviour is almost identical to individuals except a small number of players.

Most business enterprises (in most significant sectors e.g. financial, touristic, industrial … and even educational) are basically bilingual, and sometimes (like touristic) even multilingual; in most Arab countries they deal mainly (beside Arabic) in English and in some countries in French. Although on government level Arabic is

the official language, sometimes they also deal in the other two languages. This aspect supports the need for multilingual HLT tools and utilities, and will support the directions that the project is taking in this regard.

Given our statement about individual users and the countries' efforts to fight piracy we can assume that business consumers (SME and large companies) would adopt such technologies if they are useful economic growth instruments. In that scenario, some well designed technologies could be very useful: Automatic speech recognition technologies for dictation, language learning, text to speech synthesis in local colloquial varieties of Arabic coupled with news and media content suppliers, MT both desktop and on mobile for tourism industry, better Arabic search engines, etc. In addition we can also imagine supplying HLT tools for the content production players (media, news, publishers, etc.): tools such as bilingual editing software with grammar checkers, translation memories, etc.

Institutional consumers (e-government, e-health, e-… agencies) could adopt the HLT products and provide web-services for citizens. This requires more R&D to come up with e.g. advanced search engines, speech enabled services, etc.

Saudi Arabia, for example, planned to spend $20 billion on Information and Communications Technology (ICT) this year, making the Kingdom the top investor in ICT, the Arab News reported July 27, 2008. According to a recent report by international research company Global Insight, Europe, the Middle East and Africa (EMEA) region plans to spend $73 billion on ICT, exceeding $95 billion over the next three years. In 2008, the region is projected to overtake the United States in the purchase of ICT products and services.

Businesses in the Middle East have experienced a very high growth rate, compared to Europe and the US. As one of the fastest growing regions in the world, the Middle East is strategically positioned as a global technology hub, and ICT companies are seeking to capitalize on this attractive market with increased investment in the area. Moreover, Saudi Arabia is currently enjoying a massive boom in its personal computer industry since the deregulation of 2002. PC per capita has exploded to nearly 43% of the population in 2005 from just 13% in 2002 leapfrogging over the rest of West Asia. The chart given in Appendix 3, demonstrates the steady growth in expenditure over a period of 6 years for three countries: Egypt, Saudi Arabia and UAE as examples for Arab countries expenditure to improve the ICT infrastructure.

Recommendations

One can let the market evolve by itself, and given the low degree of ICT friendliness of most of the Arabic countries (and the huge requirements to enhance this area), no development can be achieved by 2015. Of course this is not desirable and there is a lot to be done if we want a successful Arabic HLT to emerge.

We can identify the following scenarios, directions and factors affecting the Arabic HLT and having an impact on the development of the industry:

1) The only optimistic aspect that can be foreseen is the relocation of R&D activities and software development to some of the countries with very low labour rates and highly skilled personnel e.g. Egypt, Jordan, Lebanon. The

[4] tools, Speech recognition and Synthesis as interfaces to content, search engines, etc.

2) Given the new initiatives, the ICT friendliness plans in some key Arab countries and created needs for Arabic HLT, there are huge opportunities for R&D, utilities, tools, applications and system integration for the software industry in the Arab world.
3) The products would follow the web model in particular and would be provided as web services (and not as typical packaged products). Here we assume an economic model more like Google than Microsoft.
4) Some specific domain will attract attention of the industry, and the need will drive developments and R&D directions. Such areas can include religious affairs (*access to advice, fatwas, recitations/pronunciation of the Quran*, etc.) would be also a factor to consider.
5) Another driving force for the coming 5-6 years can be expected to be a number of big media players, e.g. the *AlJazeera Satellite Channel*, to be an important early adopter of such technologies and to contribute to popularize these within the content/media sectors (news agencies, TV channels, etc).
6) Improvement is expected by 2015, on the number and quality of local companies who got their business based on the HLT domain, more Arabic content will be managed and be available over the net, while more systems, applications and service providers of the Arabic HLT will be available. We expect the number to grow between 20-30 companies in most of the Arab countries who are currently having some sort of interest in this domain. This is especially true based on the increasing interest in enrichment of the Arabic e-Content by many Governmental, Business and International bodies (See Appendix 4).
7) The very poor reading "habits" (less than 900 books published/year in Morocco, over 200,000 in UK ranked number 1, Turkey 6,500 (ranked 26), the first Arabic country is Lebanon (ranked 29 with 3,686 books), all Arabic countries together produce less than 12,000 books per year); this is an aspect that can be turned into a positive consumption of web-based content, assuming a serious growth of PC and Internet penetrations (and tools like Text-To-Speech). By 2015 we expect improvement on the overall book production by utilizing utilities and tools with the aid of MT products, raising substantially the average book production per country.
8) We can also imagine that the new globalization, the GII effect, Trading and Commercial effects, and the geopolitical reasons around the world, with special emphasis on relations between EU countries and USA from one side and the Arab Countries from the other side; would boost the use of Arabic and will demand an enhanced Arabic HLT. This could foster the development and deployment of applications for these purposes worldwide.
9) Consortia such as NEMLAR, MEDAR and European joint projects could lead the way in establishing cooperation between the EC groups, companies and organizations working in language technologies with counterparts from the Arab countries. This has a potential of reaching concrete agreements on concrete projects especially in the currently initiated projects.

[4] TM : Translation Memories

10) Encourage producing and distributing a model of Arabic NLP/HLT educational books.

6. Instruments

In order to promote Arabic language industry there are several avenues to follow. One way is to build on collaboration between higher education institutes and industry, allowing cooperation, collaboration, joint projects, technology and knowledge transfer and at the higher level more integration between the two bodies and mechanisms of funding and expertise exchange. In the following we will highlight the venues of cooperation and the channels of working together.

6.1 Universities and Industry need to know each other, collaborate, do technology transfer

Universities are places where Research is a major component in the activities, any faculty member is asked to do three activities as part of his/her duties:

1) Teaching
2) Research and Publication
3) University and Society Public Services

Typically, most of faculty members in Arab universities are very much involved in teaching, less in Research as exemplified by the Shanghai Academic Ranking of World Universities [http://www.arwu.org/, ranking from August 2008] which ranked the first 500 top universities in the world in terms of research activities, none of them were Arab universities. A sign which was taken seriously by most universities and by most governments in the Arab world, raising questions such as “Why?”, “What are the obstacles?” And “What are the mechanisms needed to enforce and enhance such activities and how to motivate faculty members to be involved in research more?”

Moreover, directions and measures were taken by many universities to promote applied research, rather than just doing research for the purpose of publications. Another dimension of the reason and at the same time the solution for such a problem is the gap/divide between the industry and the universities. Industries are interested in research that will enhance their products and production lines while universities are interested in theoretical research. Industry doesn't seek help nor do they communicate with universities to utilize their researchers and scientists in the development processes, and on the other hand, universities do not facilitate communication and cooperation between their faculty members and industry. On the industry side, universities cannot be blamed completely for this lack of interest, which originally developed as a result of non-existence of needs for R&D activities in the industry; most of their work is basically copying and manufacturing of products without much improvements.

These observations apply in general, to most of disciplines; in the case of HLT it is even worse because there is no real noticeable such industry. There aren't many companies who are currently working in development and/or production of tools, resources, technologies or systems. As shown in the surveys conducted in NEMLAR

and MEDAR not much significant work is undertaken in the Arabic HLT and not many applications are really available (compared to other fields).

There are many initiatives to bridge the gap between ministries of higher education and universities from one side and industries from the other side to promote interaction and exchange of expertise. Faculty members can through these initiatives work for industry as part timers (i.e. Doctor per Factory initiative at the MoHESR in Jordan); while industry experts/engineers/executives can work in universities as adjunct faculty members (e.g. MoHESR in Jordan new rules and regulations for Research), so the two parties can come closer together and can have better understanding and cooperation.

Moreover, strengthening the partnership and the cooperation between the private and public sectors to boost the industry growth and development is a major objective that should be looked at as a target. Some of the main initiatives are being planned by the ministry of ICT-Jordan to be implemented in cooperation with the private sector, most notably the 'Internship Program Initiative' that aims at building ICT graduates capacity to fit the ICT Industry needs.

Human language technology (HLT) is becoming increasingly essential to analyze, understand and exploit the huge amounts of information currently available in the form of web, textual or spoken documents. HLT gives people the possibility of using speech and/or natural language to interact to conversational machines and support advanced services. Arabic HLT can be considered as a newly emerged field, this will be of great benefit and will direct the applied research in universities towards becoming products and to find a place in the market through companies and businesses. Service and technology providers are therefore strong potential supporters of this technology; cooperation should be sought between Arab universities and such providers.

6.2 Roles of Funding in Promoting the Arabic HLT Industry

There are new trends in some Arab universities recently towards adopting (even partially) the Entrepreneurship University Model where universities will host and incubate start up companies initiated by faculty members and their students based on some research that could be turned into a product (tool, system, …etc) and be launched in the market. Funding is very critical for such initiatives and would be of great help to initiate, and convert the ideas into products.

- Create instruments to facilitate HLT entrepreneurship.

Universities will mainly play the roles of incubators and entrepreneurs promotion bodies, curricula should be developed to spread the culture and knowledge of entrepreneurship within faculty members and students; procedures and setups should be put together to encourage faculty members, researches and students to convert their innovations, ideas and thoughts into products and systems specially in the areas of Arabic HLT where those ideas and innovations only need implementation utilizing ICT and converting them into systems and products that could be launched in the market as competitive products, where they would survive in the market, given the proper support, both logistically and financially. This could be achieved as results of supporting and funding of research from interested agencies and governmental bodies.

The ultimate goal of ICT professionals is to build systems that are easy to use and user friendly, systems which can interact with simple users by their own language. Significant advancements and complete systems are still very far to be achieved; there are some realistic short term achievements in systems and tools such as: grammar and spell checkers, morphology analyzers, intelligent email sorting and response generation, document categorization, classification, extraction and summarization applications. The work on Arabic HLT can span a wide range of tasks that will ultimately utilize novel techniques and building applications resulting in products that could be market as profitable Arabic language technologies' products.

6.3 International cooperation

The globalisation has and will have a serious impact on the Arabic countries economies. Competition will be stronger and more aggressive. Examples of Silicon Valley experiences like seen in India, China, show the needed ingredients for an IT growth: performing education systems producing highly skilled talents, IT infrastructure, capital (mostly venture capital), incubation policies. These can be offered at country level (with Inter-Arabic competition as is today) or could be set up through a harmonised policy, a pipe dream today.

A large number of projects was supported by European Commission and resulted in cooperation between European and Arabic partners from different countries in the Arab world. The objectives and scope of work concentrated on Arabic HLT in general and so far all were successfully concluded. Among those: NAPLUS, ALMA, NEMLAR, MEDAR and OrienTel. In addition the European Union, in the framework of EuroMed, has launched ANIMA, an Investment Network and a multi-country platform supporting the economic development of the Mediterranean. The network gathers around 40 governmental agencies and international networks. The objective of ANIMA is to contribute to a better investment and business climate and to the growth of capital flows into the Mediterranean region.

NEMLAR: (Network for Euro-Mediterranean Language Resources) The NEMLAR project was started in order to help pave the way for collaborative effort for Arabic language resources in the Mediterranean area. 14 partners from 6 Arab countries and 4 European countries were involved in the project.

MEDAR: (Mediterranean Arabic Language and Speech Technology) MEDAR addresses International Cooperation between the EU and the Mediterranean region on Speech and Language Technologies for Arabic. MEDAR is structured around 3 pillars: (1) producing a knowledge base on HLT players, existing LRs and processing tools, activities and products for Arabic, (2) designing a strong cooperation roadmap between EU and Arabic countries, within the Arabic countries, and between academia and industry, (3) focusing on Machine Translation (MT) and Multilingual Information Retrieval (MLIR) for which required technology components, LRs, benchmarking methodologies will be identified; and 4 main objectives, and a number of instruments. MEDAR has 15 partners from 6 Arab countries and 4 European countries.

NAPLUS: Project aimed at the creation of infrastructure in language technology for the processing of Arabic. The tools developed are for the processing of Arabic in terms of morphology, syntax and semantics. The long-term aim of this project was to integrate these tools in machine translation systems such as translation memories, text retrieval information etc. 6 partners were involved in the project from 1 Arab country and 3 European countries.

ALMA: ALMA project involves the development of a technical platform providing access to information and multilingual and multimedia content in Arabic, English and French via the Internet thanks to high-level search and translation tools adapted to the end-users needs and domains. This project includes innovative issues such as the introduction of a real linguistic expertise in an information system with a real interaction between information provider and NLP systems. It also involves a dynamic online user platform for terminology management that gives access to corpus tools ensuring safety of data and giving access to validation protocol. Coupling cross-lingual search and MT will allow users in Maghreb/Machreq to have a good understanding of documents not in their native language retrieved from the multilingual network. Another innovative aspect of the project will be the creation of operational Arabic to English and French MT Systems. 8 partners were involved in the project from 2 Arab countries and 2 European countries.

OrienTel: Project that focuses on the development of language resources for speech-based telephony applications across the area between Morocco and the Gulf States, including several variants of local German French, English, Arabic, Cypriote Greek, Turkish and Hebrew. OrienTel is conducted by an academic and industrial consortium consisting of 9 partners, none of them from Arab countries.

Many other projects have also supported the positive development, e.g. DIINAR, ALFARABI, etc. which were, at least partly, consolidated within the above mentioned ones.

The NAPLUS, NEMLAR and MEDAR projects have created a steadily increasing partnership between EU and Arab HLT players from industry and academia. We recommend exploiting the network that has been created and aiming at the creation of an open and more sustainable collaboration platform where EU and Arab HLT players from industry and academia can meet in order to

- exchange experience and expertise
- promote technology take-up
- collect and disseminate information about EU and local funding opportunities
- coordinate dissemination and education activities
- maintain a continuous survey of important developments (new markets, new players, new technologies)

- Roadmap: Collaboration platform

6.4 Available and/or Potential Funding Agencies in the Arab World

Many funding organizations could play significant roles in providing the support and funds needed to launch and maintain research activities for the Arabic HLT. Those funding agencies and supporting bodies are normally interested in innovation and development, other bodies of interest could be service providers and industrial bodies who are interested in applications and systems on the Internet and web; and who are dealing with the information society. In appendix 4 we enumerate some of the most famous potential agencies who are currently supporting funding of R&D in the Arab world, and who at current stage are not directing much attention to Arabic HLT, but we believe that with the roadmap and approaching them properly this might result in convincing them of the criticality of this domain and might direct their attention to supporting this important area:

1) **Content Management; Text Mining, Business Intelligence & Data Warehousing; Internet Service Providers, Telecommunication and Mobile Companies:**
   The available media on the internet combine text, graphics, sound and movies, but at the end navigation and retrieval will be based on search using natural language. The content and information which is available on the web need systems and tools for browsing, navigating, filtering and processing. Multilinguality of the content adds more challenges for HLT. Cross lingual language barriers for information and knowledge content management will be facilitated and will be useful in the applications related to e-commerce/trade, education and international cooperation. This is where universities can play a significant role in Arabic HLT and attract the attention of industry to allocate funding for supporting such activities and helping industry to achieving their targets.

   Companies such as: Orange, Zain, Jawal, MTN, Etisalat, Vodafone, Qtel (and other mobile companies), Microsoft, Google, etc., are all potential targets for initiating projects that could be focused on tools that could be used as added values and utilities to help in utilization of the interests of each domain, e.g. search, indexing, content management, etc.

2) **Funding Agencies/Bodies: Public/Private: Funding programmes for HLT (education, technology transfer, R&D … etc):**
   The projects organized and supported by the region officials and resources are still extremely limited. Minimal attention is directed towards supporting Arabic HLT by the main grants providers and research support agencies. Governmental level interest is very low, and interest in general in this domain almost does not exist.
   Awareness of existing funding is also needed: In some countries available funds are not used by HLT players (e.g. in Morocco ANRT collects funding for ICT/Telecom research through 1% tax on all telecom operators revenues, many institutions are not aware of that).

   An education and a proper aggressive strategy should be developed to convince and engage those organizations in allocating part of their funds for the Arabic HLT related projects; this will enhance the field and boost it, the

outcome will be considerable and the return on investment would prove to be of value. Moreover, the Arab researchers and technical staff in this domain will be an important component in the promotion of such industry and be critical part in the cooperation projects in this domain with European and other partners from the world, cf. section 5.1.

Appendix 4 displays a number of funding institutions and organizations who have allocated budgets for supporting R&D and initiatives to help governments, research centres, private sector business, higher education institutions and individuals to do products and services for helping the society in achieving advancement of some sectors of interest specially those related to Arabic heritage, culture, tourism and promoting and enrichment of the Arabic content. Arabic HLT could attract many of those funds specially if associated and supported by consortia like NEMLAR/MEDAR.

7. Final remarks

Basically our recommendations are expressed in the cooperation roadmap of section 3.

From the survey of the education systems we can conclude that the academic institutions in the Arab world do not provide enough academic programmes or training courses to ensure the competency needed for the technical tasks of industry. The recommendation for this shortcoming is to develop programmes and curricula for Arabic HLT as mentioned above in the section related to players and human resources. This can be done in win-win partnership with international institutions from Europe, USA but also cross-Arabic countries.

From the technology and market perspectives, we can conclude that as a result of the spread of ICT in general, more penetration, more trends, and consequently more needs will attract the major international players to this area and start planning for dominating or getting more shares in this emerging, promising market. Major companies such as the mobile and telecommunication companies will have needs to enhance their services by facilitating the utilization of the services in the Arabic language. Major international applications and software providers and manufacturers need to add tools and utilities to their products, such as Microsoft and Google; who basically deal with language aspects and always in need to enhance their services with professional, not superficial tools and utilities. Their focus is directed towards utilities for the enhancement of Arabic processing (spell checkers, morphological analyzers, syntactic analyzers, lexicons, search engines based on Arabic main features, etc).

We can imagine two different scenarios for future development:

1) The major international players will dominate the market, develop technologies and enforce their vision, methodologies, procedures, and as a result monopolize the whole industry.
2) Local efforts will be undertaken, governments and major funding agents will encourage and incubate such activities, giving the local companies opportunities to grow and develop its own; build the capacity in local forces; and build and maintain national industry that, as a result, could be competitive on the international level.

What we recommend is a hybrid solution between the two scenarios; the Arabic HLT should get the benefit out of the huge international interest, huge allocation of investments, and immediate need for the services and applications in the ICT in general, and in related applications to Arabic HLT in particular (search engines, mobile services, e-commerce, e-government, e-learning, etc). The cooperation roadmap is focussing on promoting this view.

8. Appendices

Appendix 1. Ongoing Arab initiatives

Saudi Arabia:

King Abdullah (of Saudi Arabia) Initiative for Arabic content Project:

Cooperation between research centers, international organizations and industry (both on the local and international levels) is taking place in many countries in the region in the Arabic HLT. As one good example we can mention the "*King Abdullah (of Saudi Arabia) Initiative for Arabic content*", where efforts are directed towards enrichment of the Arabic content in the Internet, and launching projects to achieve the target and the objectives of the project. The King Abdullah Initiative for Arabic Content and Google-Knol is a competitive project to enrich the Arabic e-Content where students and faculty members at universities and schools write essays on subjects that fall in their fields of interests and then a group of reviewers assessed the articles periodically for awards nomination of best articles. The aim of this competition is to motivate the community to enrich the Arabic content, encourage reading and writing, writing skills in many areas of science, spread awareness of the importance of participating in writing and publishing of Arabic content articles.

The project approaches many directions: locally, regionally and on the international levels, moreover, cooperation is sought by calling for proposals to supply tools and utilities to help in the building process. The project is aiming at open source based solutions, and looking to avail all results for public use. We believe that the directions MEDAR is taking in the two domains which were selected, namely: MT and IR will definitely be of use and benefit to the project, and will go in line with helping in this direction. We can see that there is a room to expand on this initiative, and propose several directions to help in enriching the Arabic content. In a very recent press release at Riyadh Saudi Arabia, on June 10th 2009, "King Abdullah Initiative for Arabic content has recently adopted a number of important projects for the advancement of Arabic content on the internet (e-Content) and for the preservation of identity and heritage. The Initiative supervised by King Abdulaziz City for Science and Technology (KACST), has implemented the Interactive On-Line Dictionary "*Almuajam*" as a collaborative efforts between KACST and ALECSO (The Arab League Educational, Cultural and Scientific Organization) to be an open source dictionary for all. This project is expected to be completed within one year period.

KSU tops universities in Arabic e-Content

RIYADH – King Saud University has topped universities in the enrichment of the Arabic Internet content. The KSU led the contenders in the competition “the enrichment of Arabic content on the world wide web” which concluded very recently. This has been the result of cooperation between King Abdullah Initiatives for Arabic Content and Google-Knol. KSU advanced the nearest competitor in the number of participation with more than 5,000 knols as it recorded 7,062 knols.

Syria:

eSyria, Enrich the Syrian National Content

eSyria {*http://www.esyria.sy*} is a national initiative launched by Syrian Computer Society for Informatics in Syria, this initiative aimed to enrich the Syrian national content on the internet and raise the level of electronic services, also, provided to citizens in Syria with the changing needs of information and useful for all events and activities in all provinces of the country. This project aims to increase the Arabic content on the Internet and the Syrian special services, and encouraging the citizens to get advantage through added Internet services to all segments of the Syrian people, the purpose from eSyria comprehensive development of knowledge and sharing knowledge of Syrian society. It is intended as a step in the direction of the e-government projects and aimed to be the Syrian website query for e-government services.

The 1st National Conference on Digital Arabic Content Industry

For the first time Syria held a national conference on Digital Arabic Content Industry with a view to searching for the DAC's national strategy and to setting up an incubator and a share structure for the industry. The three-day event that was launched on June 13, 2009 at Omayad Palace in Damascus, and was coordinated by Communications and Technology Ministry and the Syrian Computer Society (SCS) in cooperation with UNESCO's Representatives in Syria. This is a cooperation effort between a local government, professional society and International organization. The conference's goals are summarized by the following:

1. Exchanging experiences and getting benefit from considerable expertise in the field of digital content.
2. Drawing up standards and indicators for digital content.
3. Reaching an integrated system to realize a digital content's national strategy.
4. Defining components of the Digital Syrian Content.
5. Defining mechanism of evaluation and locally following up implementation.
6. Developing role of governmental institutes and research centers in this domain.
7. Reactivating role of private sector and the NGOs in the Digital Syrian Content.

United Arab Emirates:

Arabeyes.org

Arabeyes is a Meta project that is aimed at fully supporting the Arabic language in the Unix/Linux environment, it is designed to be centrally located to standardize the Arabization process {*http://aosaf.org/links/7-arab-open-source-technology-organization-website/1-arabeysorg.html/*}. *Arabeyes* relies on voluntary contributions by computer professionals and enthusiasts all over the world.

Unlike other Arabized products, *Arabeyes* is Open Source, completely FREE and abides by the ideals of the open source communities. *Arabeyes* aims for all active participants and volunteers to partake in the development process and to speak their mind on this professional effort. The idea of Arabizing Unix is not a new one. However, most of the attempts were done by Arab computer science students studying outside of the Arab world. Once those individuals graduated, their projects

were abandoned, resulting in very sparse code and in every subsequent attempt in having to re-invent the wheel. *Arabeyes* aims to eliminate this redundancy and lack of information by not depending on any one person or group - *Arabeyes* is aimed to succeed irrespective of the interest or availability of the parties involved.

Arab Science and Technology Foundation

The Arab Science and Technology Foundation is an independent, nongovernmental, non-profit regional and international organization. Scientists and researchers from inside and outside the Arab world as well as representatives of Arab and international science centers and organizations have contributed to its establishment. It is based in Sharjah, United Arab Emirates, and is seeking to set up branches and links in those Arab and world capitals that have scientific entities willing to participate in its activities {*http://aosaf.org/links/7-arab-open-source-technology-organization-website/2-arab-science-and-technology-foundation.html/}*

ASTF was founded upon a recommendation from the symposium on "Scientific Research Outlook in the Arab world and the New Millennium: Science and Technology" organized by the University of Sharjah during 24-26 April 2000 and attended by 375 scientists from inside and outside the Arab world. Its establishment was declared in the closing session of the Symposium upon the initiative of His Highness Sheikh Dr. Sultan Bin Mohammed Al-Qassimi, Member of the Supreme Council of the UAE, Ruler of Sharjah. Its board of directors and advisory board are composed of an elite cluster of renowned Arab scientists whose outstanding scientific achievements are well known on the national and international levels.

The Foundation aims to identify and support outstanding scientific research activities being conducted by women and men of science and technology from the Arab world. It also attempts to act as a mediator between those who produce, develop, finance or benefit from scientific research, while seeking to become a center for assessing the performance of scientific programs and also a powerful international Arab entity that defends the region's interests in scientific and technological progress.

Dahsha Arabic On-line Encyclopedia Project:

Dahsha (www.dahsha.com) is an Arabic online encyclopedia and one of the serious efforts to enrich Arab content on the internet. A lot like Wikipedia, the encyclopedia relies on volunteers to write articles and publish content.

Volunteers can post books, reports, studies, articles and even video, audio and useful programs to the encyclopedia. This project covers some types of content that Wikipedia doesn't support. The system isn't as open as Wikipedia though, as users can't change or edit articles by other users; they can only reply or comment on them. Which means the site's administrators have to be more involved in reviewing the submitted content, its quality, and whether publishing it violates any copyright laws. The encyclopedia has already published a big volume of content covering different categories and topics from computer science and technology related content to art, literature and religion.

Jordan:

Talal Abu-Ghazaleh Organization (*TAGorganization www.tagorg.com/*) recently announced the launch of the first Arabic electronic encyclopedia (*TAGIPEDIA*), the first initiative of its kind that targets Arab academics, scholars and readers, and

provides them with general knowledge and sciences from an Arabic cultural, historical, and educational perspective. *TAGIPEDIA* aims at founding a free-access database which delivers a broad variety of knowledge in all humanities sciences, with the highest standards of quality, integrity, and accuracy. It also gives the chance for readers/users to augment the content through contributing to the Encyclopedia without prejudicing the rights of its original author.

TAGorganization is a private organization, with an aim to enrich the Arabic digital content, in a media that is known to be very poor in providing Arabic content; Arabic documents in the famous Wikipedea is very small, it is estimated to be 0.25% which is not a good representation if compared to the size of the Arabic speaking users over the internet. *TAGorganization* initiated a cooperation agreement between the Association of Arab Universities, due to the fact that both institutions are highly concerned in promoting knowledge and culture, aiming at enriching the Arabic content over the internet in general and on the Wikipedia in particular. In order to preserve the authenticity and quality of all contributions, in addition to providing academic and scientific support to the Encyclopedia's content, the Organization has drafted this agreement with the Association of Arab Universities.

The agreement aims at enhancing cooperation between the two parties in order to enrich the content, defining the references and sources of the incoming information whether they be institutions, universities, or research centres, and drawing up agreements with these parties in this regard.

Through the expertise and qualifications at *TAGorganization* along with the Association of Arab Universities, both will be working relentlessly on bringing *TAGIPEDIA* to light for its indispensable role in supporting Arab universities to raise generations capable of serving and developing the (Arab) nation and enriching the digital content.

Another cooperation agreement was signed between *TAGorganization* and the Emirates Internet Group to enrich *TAGIPEDIA*'s content with a whole set of innovative information and ensure the achievement of the Encyclopedia's goals. The Emirates Internet Group will be responsible for building bridges of collaboration to accomplish the project's objectives in all of the GCC countries, Yemen, and Iraq, as well as following up with potential sponsors to provide financial support as to the awards and incentives related to the project.

In this respect, the Chairman of both the Emirates Internet Group and the Arab Union of Internet Societies, Mr. Khaled Bin Thani said, "This initiative comes as a result of the lack of significant Arabic contributions on the internet, which emphasizes the importance of creating an Arab academic and scientific online reference that provides credible information." *TAGorganization* will provide the essential technical support; take the responsibility of assuring constant development of quality information, in addition to establishing the project in Middle Eastern and North African countries. Consequently, a website was created to give the opportunity to those who are qualified and willing to contribute to the content of *TAGIPEDIA*. Interested users may access the site http://register.tagipedia.com and register as users.

Jordan HLT related Industry:

As an example of the values of business related to the industry of the Arabic HLT related products, according to Int@j (Information Technology Association - Jordan) the IT Industry of Jordan Statistics 2007 the percentage of business close to areas that might be related to HLT is very small (1.31% for Domestic local sales and 3.14% for Export to the region and the world) of the total ICT industry in Jordan; if it is directly associated with HLT then it is almost negligible.

Type of Service	**(Domestic) Percentage**	**Revenue (USD)**	**(Export) Percentage**	**Revenue (USD)**
Web Design, Development and Administration	0.99%	6,792,024	3.12%	6,143,520
E-Commerce	0.31%	2,126,795	0.00%	0
E-learning	0.01%	68,606	0.02%	39,382
Total	1.31%	**8,987,425**	3.14%	**6,182,902**

Figures for the telecommunications sector have not been included in the industry statistics nor in the survey produced by Int@j, as they are compiled and reported annually by the TRC (Telecommunication Regulatory Commission) in Jordan. According to the TRC website, it is estimated that 2007 telecom figures have crossed the USD 1.1 billion figure bringing the total ICT industry revenue figures to over USD 2 billion.

These are good indicators, we believe that similar indicators could be found in many other countries, and should be considered carefully, to enhance and boost the Arabic HLT directly related services, tools and products; and associate them with Internet technology, e-Learning, e-Commerce and most importantly with Mobile industry.

Intel Capital

Intel Capital, Intel's global investment organization, makes equity investments in innovative technology start-ups and companies worldwide. Intel Capital invests in a broad range of companies offering hardware, software, and services targeting enterprise, home, mobility, health, consumer Internet, semiconductor manufacturing and cleantech. Since 1991, Intel Capital has invested more than US$9 billion in over 1,000 companies in 46 countries. In that timeframe, 174 portfolio companies have gone public on various exchanges around the world and 231 were acquired or participated in a merger. In 2008, Intel Capital invested about US$1.59 billion in 169 investments with approximately 62 percent of funds (excluding Clearwire) invested outside North America. This year, Intel invested in two companies in Jordan who are working in content development.

Cooperation between Amman University and Arabtext Co.

An agreement for cooperation and joint venturing is signed between Amman University and Arabtext Company, to work together to develop the research and resulted prototypes into commercialized products. The cooperation will concentrate on Arabic language technologies, tools and utilities. This is an achievement step towards linking industry with academia and will enrich the field by constructing industrial lines and commercialized products. It will also open other areas of

cooperation and will bridge the gap between the academic institutions and the business and industrial world. It will provide funding to work on more projects that will be developed into products and services.

Egypt:

Fekr Rama Portal Project:

Collaborative efforts resulted in the creation and initiation of the Fekr Rama (www.fekr-rama.com) project in Egypt. It is a new web portal that provides rich Arabic content in the fields of culture, education, heritage, religion, sport and the arts for Arab and other audiences. As a collaboration between the Ministry of Communications and Information Technology (MCIT), Al Azhar, the Ministry of Culture, the Ministry of Information, the Egyptian Publishers Union and the e-Learning and Business Applications Union, the portal was launched in March 2008.

Fekr Rama is the result of two years of preparation, and currently contains more than 6,000 titles, 3,000 of which are free to access. By 2009, some 20,000 titles will be available. The portal contains e-books, magazines, directories, news and audio files, as well as animation for children and programs that promote learning. Fekr Rama, derived from Panorama El-Fikr El-Araby (Arab Intellectual Panorama), focuses on Egypt as a pioneering civilization in the spread of knowledge and as a role model in preserving Arab heritage for future generations.

Project Objectives

- To **preserve** Arabic heritage for coming generations through the enrichment of Arabic digital content in all forms
- To **support** and promote the Arabic digital content industry and related industries
- To **boost** investment in the hosting of Arabic digital content
- To **raise** Egypt's competitiveness as a supplier of Arabic digital content
- To **increase** job opportunities in the national ICT sector, particularly for youth

Portal Components

Books

The portal contains a wide selection of e-books sorted into more than 28 categories for simple searching. Free books are available for download.

Animated Books

Animated versions of 50 children's plays, allowing viewers greater interaction than the original print versions.

Photos

A rich library of historical and artistic photos from different eras in Egyptian history, with a focus on natural and cultural heritage.

Multimedia

Multimedia resources include national, classical and children's songs.

Project-related Protocols

- **MCIT and the National Library and Archives of Egypt** (July 2006): Digitization of more than 50,000 titles with full bibliographic information

- **MCIT and Dar Al Maaref** (July 2006): Digitization of 10,000 titles
- **MCIT and the National Theater** (July 2006): Digitization of 250 plays
- **Agreements** with more than 50 partners – 10 governmental and 40 from the private sector

Launching the World Digital Library Project:
On the 21st April, 2009 - UNESCO and 32 partner institutions, including Bibliotheca Alexandrina launched the World Digital Library, a Web site featuring unique cultural materials from libraries and archives from around the world. The site www.wdl.org includes manuscripts, maps, rare books, films, sound recordings, prints and photographs. It provides unrestricted public access, free of charge, to this material.

The launch took place at UNESCO Headquarters at an event co-hosted by UNESCO Director-General Koïchiro Matsuura and U.S. Librarian of Congress James H. Billington.
Mr Billington first proposed the creation of a World Digital Library (WDL) to UNESCO in 2005, remarking that such a project could "have the salutary effect of bringing people together by celebrating the depth and uniqueness of different cultures in a single global undertaking." Mr Matsuura welcomed the proposal as a "great initiative that will help to bridge the knowledge divide, promote mutual understanding and foster cultural and linguistic diversity."

In 2007, the Library of the Congress asked for Bibliotheca Alexandrina's technical assistance, and an agreement was signed between both parties outlining their collaboration to develop this huge project. Accordingly, they worked together on the design and implementation of the database, search engine and interface for the project. In addition, Bibliotheca Alexandrina contributed its particular expertise in the search and display of Arabic texts.

Moreover, Bibliotheca Alexandrina has enriched the content of the WDL with a digitized copy of the valuable, "Description de l'Egypte", a work of scientific observation carried out by French scholars during Napoleon's military foray into Egypt in 1798. The whole 20 volumes of text and plates were digitized in 2007 at Bibliotheca Alexandrina's premises.
The WDL functions in seven languages – Arabic, Chinese, English, French, Portuguese, Russian, and Spanish – and includes content in more than forty languages. Browse and search features facilitate cross-cultural and cross-temporal exploration on the site. Descriptions of each item and videos, with expert curators speaking about selected items, provide context for users and are intended to spark curiosity and encourage both students and the general public to learn more about the cultural heritage of all countries.

Institutions contributing to the WDL include national libraries and cultural and educational institutions in Brazil, Egypt, China, France, Iraq, Israel, Japan, Mali, Mexico, Morocco, the Netherlands, Qatar, the Russian Federation, Saudi Arabia, Serbia, Slovakia, South Africa, Sweden, Uganda, the United Kingdom, and the United States. "UNESCO welcomes the creation of the World Digital Library which reflects the values and priorities of our Organization," Mr Matsuura declared. "WDL offers an invaluable platform for the free flow of information, for international solidarity, for the celebration of cultural diversity and for the building of inclusive knowledge

societies. With projects like the Digital Library, the cultural and societal potential of digital technologies come into their own."

"We are honored to be working with so many great libraries in this venture," said Mr Billington, "and thankful for the strong support that UNESCO has given to this project. What we launched today is a first step. We look forward to seeing this project realize its ambition to bring people together, deepen our understanding of each other, and help electronically oriented young people enjoy what is best in traditional culture, using the new media".

Partners of this tremendous venture have all confirmed the important role that World Digital Library plays in the appreciation of other cultures and nations and in bringing together different countries and peoples in mutual understanding and enrichment. They believe that the creation of the World Digital Library has emphasized the spirit of equality and open understanding between nations.

The World Digital Library showcases cultural treasures from all over the world that include; Arabic scientific manuscripts from the National Library and Archives of Egypt; early photographs of Latin America from the National Library of Brazil; the Hyakumanto darani, a publication from the year 764 from the National Diet Library of Japan; the famous 13th century "Devil's Bible" from the National Library of Sweden; and works of Arabic, Persian, and Turkish calligraphy from the collections of the U.S. Library of Congress.

This project is considered an open channel of cultural and intellectual understanding between peoples of the world in today's language- the Internet

A Regional Perspective:
Regional Profile of the Information Society 2007

A report by The United Nations Economic and Social Commission for Western Asia (UN-**ESCWA**), released in 2009, for evaluating the progress and ranking of Arab countries in relation to ICT, Information Society and Internet utilization, with emphasis on content and promotion of cultural activities. There are some projects identified in the report, we think that there are areas of intersection between those projects and our interests, as we can utilize our resources and capitalize on the available opportunities to build many cooperation agreements, we think UN-ESCWA represent a strong and good potential to be in alliance with and to extend our network for cooperation to include such organization:

Comparative analysis of cultural diversity and identity, linguistic diversity and local content in the ESCWA region

There is a wide variation in the level of ICT usage to promote and preserve the Arabic culture, heritage and language among ESCWA member countries. The United Arab Emirates undertook the first initiative several years ago by launching a portal created by the Cultural Foundation in Abu Dhabi, which included hundreds of Arabic heritage books. Alwaraq.com is now considered the biggest online Arabic heritage portal so far.

1) ICT in support of cultural and linguistic diversity

Cultural and linguistic diversity is seen as one of the factors encouraging artistic creativity and scientific discoveries, as well as a means of coping with the demands of modernization. Some countries of the ESCWA have launched various initiatives aiming to support cultural and linguistic diversity in the past few years. Thus far not many of these initiatives achieved much progress, although it is still too early to assess their relative success or failure, given the amount of time involved since their conception.
In Egypt, three protocols have been signed in 2005 and 2006 between the Ministry of Communications and Information Technology, the Ministry of Culture and Dar El Maaref in order to digitize a collection of Arabic books and national theatre. In Bahrain, the Ministry of Information has published a web-based cultural magazine. In addition the digital library provided by the Bahrain Center for Studies and Research to its members is regarded as a contribution to support cultural content.

The greatest number of such initiatives however seem to belong to the United Arab Emirates and Qatar, which are both aiming to become an active cultural pivot as far as preserving and promoting Arabic culture and heritage are concerned. Both countries' initiatives aim to establish leadership in preserving the culture, heritage and language on multiple levels: Arabic, Islamic and international levels. Both countries plan to implement major cultural projects, such as establishing museums and libraries, and translating books published in other languages into Arabic (e.g. Abu Dhabi's Kalima Project), which will contribute to the Arab cultural and content diversity through electronic websites accompanying such projects.

As for the rest of the ESCWA member countries, some individuals have, with the use of their personal resources, undertaken initiatives to support cultural and linguistic diversity as well. They have managed to do so through several websites, which include some which carry information about the Syriac language[1] in the Syrian Arab Republic, and heritage-related manuscripts in Egypt[2].

2) Local and foreign digital content development

Dubai-based research and consultancy company, Madar Research Group, has recently conducted a study on Arabic content on the Internet (both in Arab countries and abroad) using a methodology based on world search engines. The results showed that the number of Arabic web pages is still considerably low compared to the overall number of web pages published in other languages, and accounting for just about 0.16 per cent at the end of 2006. The insignificance of such value becomes more apparent when one considers that Arabic speakers constitute about 5% of the world population. The brighter side, however, lies in the prevailing growth rate of Arabic web pages which has been in excess of 400 per cent per year since 2005.
Some ESCWA member countries have undertaken certain initiatives that directly aim at increasing digital Arabic content. These include an initiative launched by Egypt's Ministry of Communications and Information Technology (MoCIT) in 2005 to create digital Arabic content portals. The initiative aimed to provide about 300,000 URLs on the Internet, and was planned to be operational by mid-2007. Progress, however, has been slow and there seem to be no indications so far whether the project will be completed.

Additionally some governmental and non-governmental organizations have designed an incentive-based approach to encourage individuals and institutions to contribute to

the development of Arabic content on the Internet. This includes organizing competitions, which provided monetary rewards and prizes. Examples of such competition are the National e-Content Competition in Egypt, and the Digital Excellence Prize initiated by the Ministry of Communications and Information Technology in Saudi Arabia. Both projects aim to enrich Arabic digital content at the same time enhance Arab talents in the development and design of websites.

Remarkably Egypt was the only country, which developed in 2007 a strategy for developing software tools for creating Arabic digital content.

3) The distribution of Arabic versus English language web content in the ESCWA region

Madar Research Group created a methodology using the search engines available on the Internet (Google, in particular) to determine language distribution in domains (or URLs) of websites deployed under the Country Code Top Level Domains (ccTLDs) of ESCWA member countries. These domains do not include sites registered under generic TLDs, such as dot-com and dot-org.

Results gathered using this methodology were not very accurate, but they do provide a rough estimate about Arabic content on the Internet. Based on these results, the United Arab Emirates has the largest amount of web content (in English and Arabic languages) among ESCWA member countries, accounting for 18.8 per cent, as of mid-2007, followed by Saudi Arabia with 18.5 per cent.

Saudi Arabia, however, led in terms of the overall Arabic web content in the ESCWA region, accounting for 20 per cent, followed by the United Arab Emirates (17 per cent), the Syrian Arab Republic (14 per cent), Kuwait (13 per cent) and Egypt (10 per cent). The Syrian Arab Republic ranked first in the amount of content growth (Arabic and English), which increased about 160 times over a two-year period (mid-2005 to mid-2007). This means that about 160 Syrian web pages were added for each web page available back in 2005. Kuwait came second in terms of content growth, which increased about 58 times over the same period .In terms of the proportion of the number of Arabic pages to the total Arabic and English pages, Iraq ranked first (98 per cent), followed by the Syrian Arab Republic (96 per cent), and then Kuwait (82 per cent). Iraq's leadership, however, has to be viewed from the vantage point that it has the least number of Arabic and English web pages compared to the rest of the ESCWA member countries, accounting only for a dismal .0001 per cent.

Except for Lebanon, the amount of Arabic content has grown more than the amount of English content in all ESCWA member countries, most likely due to the increase of e-government related projects in the region.

Suggestions and recommendations

The report concluded with a set of recommendations, from our side, we can see significant role and contribution in terms of cooperation and utilization of the technologies and trends discussed in the roadmap:

1. Accelerate the implementation of e-government projects through the allocation of additional funds, since these projects specifically increase online Arabic content;
2. Set up effective governmental programmes that aim to develop Arabic content, establish non-profit Arabic digital information banks, and crystallize a plan for regional cooperation in this respect;

3. Support individual projects related to cultural and linguistic diversity through free hosting of these website projects and furnishing them with financial assistance to sustain their activities;
4. Increase governmental infrastructure investment to promote use of the Internet. Focus on the Internet as a distinct advertising environment which can convey the product message to the consumer in a more effective way than the traditional methods, which will help develop commercial content;
5. Accelerate the legislation of laws and statutes related to the rights of digital publishing;
6. Alleviate controls and minimize strict filtering of websites in some ESCWA member countries as much as possible;
7. Provide an enabling environment which motivates the private sector to participate in establishing the Arabic content industry;
8. Focus on e-learning content within the framework of a public strategy for lifelong education;
9. Focus on the media of the local groups and consider the cultural diversity within Arab countries as a support element in developing the Arabic content industry.

Appendix 2. Players

In this section we provide a list of players, sorted by country. This list will be extended and improved continuously.

First we list the Arabic-speaking countries (in alphabetical order), this includes both partner countries and other countries.
Then we list other countries, in Europe, America and Asia.

We used the following codes to indicate the type of player

A = Association, Project, Network etc
D = Developer
E = Educator
I = Integrator
O = Other (e.g. government agency)
R = Research
S = Services
V = Vendor

Please help us improve the overview! Kindly provide information additional payers, and help us get the codes right, email nemlar@hum.ku.dk.

Arab states	
Algeria	**Type**
CRSTDLA (Scientific & technical Research Center for Arabic Language Development)	ER
Egypt	
AlKhawarizmy	DV
ALTEC (Arabic Language Technologies Elevation Center)	R
Arabize	S
Cairo Microsoft Innovation Center (CMIC)	R
Coltec, Egypt	D
Data Mining & Computer Modeling Centre of Excellence (DMCM-CoE)	R
DEEC-FECU – Department of Electronics and Electrical Communications, Faculty of Engineering, Cairo University, Egypt	ER
ESLE – The Egyptian Society of Language Engineering, Egypt	A
Faculty of Computers and Information, Cairo University, Egypt	ER
FCIS – Faculty of Computer & Information Sciences, Egypt	ER
IBM WTC, Egypt	DR
MobiDev, Egypt	DS
ITIDA (IT Industry Development Agency) "Funding and Support Agency emanated from the Egyptian Ministry of Communications and IT"	A
Orange Labs-France Telecom-Egypt Group	RS
RDI – – The Engineering company for the development of computer systems, Egypt	DR

Sakhr-Egypt	D
Tayait	DIR
The Arab academy for Sciences and Technology, Egypt	ER
The Egyptian Society for the Arabisation of Science, Egypt	A
Jordan	
Al-Ahlya Amman University – Faculty of Information Technology, Jordan	ER
Arabic Textware, Jordan	DSV
Arab Banking Institute for Financial Studies	ER
Arab Open University – Amman	ER
Petra University	ER
Al-Balqa Applied University	ER
Princess Soumaya University for Science	ER
Kuwait	
Sakhr, Kuwait (and Egypt)	D
Lebanon	
Hariri Canadian Academy of Sciences and Technologies, Lebanon	ER
Jinny Paging company, Lebanon	V
UOB – University of Balamand – Department of Computer Engineering, Lebanon	ER
Morocco	
MLTC, Morocco	D
ENSIAS – University of Mohammed V Soussi - Ecole Nationale Supérieur d´informatique et d´analyse des Systèmes, Morocco	ER
Institute for the Study and Research on Arabisation, Morocco	R
Laboratoire de Recherche en Informatique et Telecommunications Faculte des Sciences, Morocco	R
University Cadi Ayyad - Faculty of Sciences, Morocco	ER
Saudi Arabia	
ArabNet Technology	D
King Abdulaziz City for Science and Technology, Saudi Arabia	ER
Syria	
Higher Institute for Applied Sciences and Technology (HIAST), Syria	ER
West Bank and Gaza Strip	
Birzeit University – Birzeit Information technology UNIT (BIT) & Arabic Department, West Bank and Gaza Strip	ER
IT College / Birzeit University, West Bank and Gaza Strip	ER
Unit for Learning Innovation- Birzeit University, West Bank and Gaza Strip	ER

EUROPE	
Belgium	
Catholic University Leuven (KUL), Belgium	ER
Denmark	
Copenhagen Business School CBS	ER
SDU – University of Southern Denmark, Denmark	ER
UCPH – University of Copenhagen, Denmark	ER
France	
CEA, Commissariat à l'Energie Atomique: CEA-LIST/LIC2M, Laboratoire d'Ingénierie de la Connaissance Multimédia Multilingue, France	R
Cimos, France	D
CNRS – Centre National de la Recherche Scientifique - Laboratoire LLACAN - UMR 8135 du CNRS, France	R
ELDA, France	SV
France Telecom R&D ORANGE Labs, France	RD
Lyon2 – Université Lumière Lyon 2 Faculté des Langues, France	ER
Systran, France	D
Université du Maine	ER
Xerox Research Centre Europe	R
Germany	
RWTH University of Aachen	ER
Greece	
ILSP –Institute for Language and Speech Processing, Greece	ER
Hungary	
Morphologic Ltd.	D
Italy	
Istituto di Linguistica Computazionale – CNR – Italy	ER
South Africa	
Interpret	D
Spain	
TALP Research Center - Universitat Politecnica de Catalunya, Spain	ER
Universidad Politecnica de Valencia, Spain	ER
Switzerland	
European Association for Machine Translation (EAMT)	A
UK	

Applied Language Solution	D
ATA Software Technology Ltd.	V
Exeter University	ER
Queen Mary University of London	ER
The Open University	ER
Translution	S
University of Cambridge	ER
University of Edinburgh	ER
University of Lancaster	ER
University of Leeds	ER
University of London School of African and Oriental Studies	ER
University of Manchester	ER
University of Sheffield	ER
AMERICAS	
USA	
AppTek Inc	D
BBN Technologies	RD
Carnegie Mellon University	ER
Columbia University Center for Computational Learning Systems, USA	ER
Google	D
IBM Watson Research Center	R
Indiana University, USA	ER
Johns Hopkins University	ER
Language Engineering Co.	D
Linguistic Data Consortium (LDC)	DV
MaximumEdge.com Translation	S
Microsoft Corporation	S
MITRE Corporation	D
MTM Linguasoft	DS
Smart Link Corporation	S
Taragana (USA and India)	D
Transclick Inc.	S
University of Maryland, College Park, USA	ER
University of Southern California, Information Science Institute	ER
U.S. Army Research Laboratory	R
Worldlingo Inc.	DS
ASIA	
Japan	
CJK Dictionary Institute, Japan	D

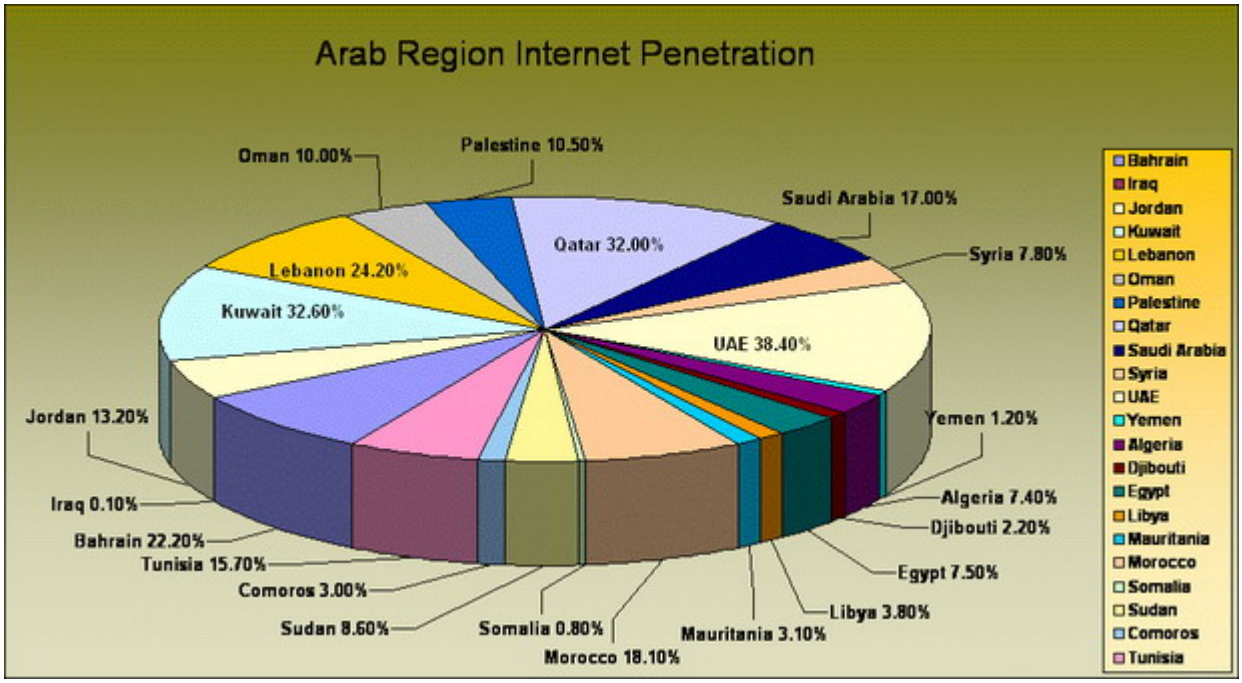
Appendix 3: E-infrastructure and Market, details

The following table and figures below illustrate the status of Internet usage and access in all Arab countries:

Arab Countries Internet Usage and Population Statistics						
Arab Countries	Population	Usage Dec 2000	Usage Mar 2008	% Population (Penetration)	(%) of Arab Region	Use Growth (2000-2008)
Bahrain	708,573	40,000	157,300	**22.20%**	0.49%	293.30%
Iraq	27,499,638	12,500	36,000	**0.10%**	0.11%	188.00%
Jordan	6,053,193	127,300	796,900	**13.20%**	2.49%	526.00%
Kuwait	2,505,559	150,000	816,700	**32.60%**	2.56%	444.50%
Lebanon	3,925,502	300,000	950,000	**24.20%**	2.97%	216.70%
Oman	3,204,897	90,000	319,200	**10.00%**	1.00%	254.70%
Palestine	2,535,927	35,000	266,000	**10.50%**	0.83%	660.00%
Qatar	907,229	30,000	289,900	**32.00%**	0.91%	866.30%
Saudi Arabia	27,601,038	200,000	4,700,000	**17.00%**	14.71%	2250.00%
Syria	19,314,747	30,000	1,500,000	**7.80%**	4.70%	4900.00%
UAE	4,444,011	735,000	1,708,500	**38.40%**	5.35%	132.40%
Yemen	22,230,531	15,000	270,000	**1.20%**	0.85%	1700.00%
Algeria	33,333,216	50,000	2,460,000	**7.40%**	7.70%	4820.00%
Djibouti	496,374	1,400	11,000	**2.20%**	0.03%	685.70%
Egypt	80,335,036	450,000	6,000,000	**7.50%**	18.78%	1233.30%
Libya	6,036,914	10,000	232,000	**3.80%**	0.73%	2220.00%
Mauritania	3,270,065	5,000	100,000	**3.10%**	0.31%	1900.00%
Morocco	33,757,175	100,000	6,100,000	**18.10%**	19.09%	6000.00%
Somalia	12,448,179	200	94,000	**0.80%**	0.29%	46900.00%
Sudan	39,379,358	30,000	3,500,000	**8.60%**	10.96%	11566.70%
Comoros	711,417	1,500	21,000	**3.00%**	0.07%	1300.00%
Tunisia	10,276,158	100,000	1,618,440	**15.70%**	5.07%	1518.40%
TOTAL	**340,974,737**	**2,512,900**	**31,946,940**	**9.37%**	**100.00%**	**920.20%**

Arab Countries Internet Usage and Access Statistics, all data conducted by (www.internetworldstats.com)

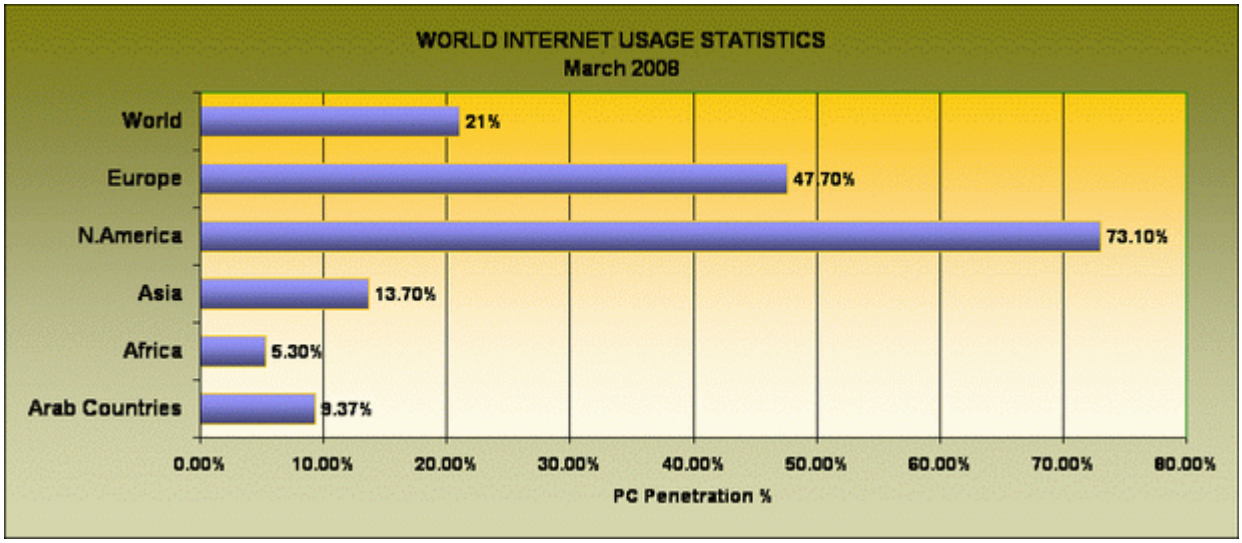
Figure 1: Internet usage in Arab countries



PC Penetration Percentage In all Arab Countries, percentages conducted by (www.internetworldstats.com)

Figure 2: PC penetration in the Arab world

The figure below illustrates the Arab countries Internet usage in comparison with the world regions:



PC Penetration Percentage in all world regions; conducted by (www.internetworldstats.com).

Figure 3: PC penetration in the world

Economist Intelligence Unit 2008 Report, *How technology sectors grow: Benchmarking IT industry competitiveness*

Country	**Score**	**2008 Rank**	**2007 Rank**
United States	74.6	1	1
Taiwan	69.2	2	6
United Kingdom	67.2	3	4
Sweden	66.0	4	7
Denmark	65.2	5	8
Canada	64.4	6	9
Australia	64.1	7	5
South Korea	64.1	8	3
Singapore	63.4	9	11
Netherlands	62.7	10	12
Arab Countries			
Saudi Arabia	32.3	40	38
Egypt	25.3	53	55
Algeria	18.5	65	59

Figure 4: IT industry competitiveness index

Economy	Rank 2007	IDI 2007	Rank 2002	IDI 2002
Sweden	1	7.50	1	6.05
Korea (Rep.)	2	7.26	3	5.83
Denmark	3	7.22	4	5.78
Netherlands	4	7.14	6	5.43
Iceland	5	7.14	2	5.88
Norway	6	7.09	5	5.64
Luxembourg	7	7.03	21	4.62
Switzerland	8	6.94	7	5.42
Finland	9	6.79	8	5.38
United Kingdom	10	6.78	10	5.27
Hong Kong, China	11	6.70	12	5.10
Japan	12	6.64	18	4.82
Germany	13	6.61	14	5.02
Australia	14	6.58	13	5.02
Singapore	15	6.57	16	4.83
New Zealand	16	6.44	19	4.79
United States	17	6.44	11	5.25
Ireland	18	6.37	26	4.36
Canada	19	6.34	9	5.33
Austria	20	6.32	20	4.64
Macao, China	21	6.25	23	4.41
Italy	22	6.18	24	4.38
France	23	6.16	25	4.37
Belgium	24	6.14	15	4.91
Taiwan, China	25	6.04	17	4.82
Estonia	26	5.97	31	3.93
Spain	27	5.91	28	4.10
Slovenia	28	5.88	22	4.47
Israel	29	5.60	27	4.24
Malta	30	5.54	29	4.04
Portugal	31	5.47	32	3.87
United Arab Emirates	32	5.29	40	3.27
Lithuania	33	5.29	43	3.17
Greece	34	5.25	30	3.94
Hungary	35	5.19	36	3.49
Latvia	36	5.01	39	3.30
Cyprus	37	4.97	33	3.78
Slovak Republic	38	4.95	35	3.51
Poland	39	4.95	37	3.34
Czech Republic	40	4.88	34	3.74
Brunei Darussalam	41	4.80	41	3.27
Bahrain	42	4.69	38	3.30
Croatia	43	4.68	42	3.19
Qatar	44	4.44	47	2.84
Bulgaria	45	4.37	51	2.74
Romania	46	4.16	60	2.48
Argentina	47	4.12	44	3.06
Chile	48	4.00	45	2.97
Uruguay	49	3.88	46	2.90
Russia	50	3.83	52	2.71
Ukraine	51	3.80	59	2.50
Malaysia	52	3.79	50	2.74
Jamaica	53	3.78	48	2.79
Belarus	54	3.76	57	2.53
Saudi Arabia	55	3.62	73	2.13
Trinidad & Tobago	56	3.61	58	2.50
Kuwait	57	3.57	49	2.77
Bosnia	58	3.54	66	2.33
Turkey	59	3.49	63	2.41
Brazil	60	3.48	54	2.55
Panama	61	3.46	62	2.42
Mauritius	62	3.45	61	2.45
Thailand	63	3.44	70	2.17
Lebanon	64	3.43	56	2.53
TFYR Macedonia	65	3.42	53	2.65
Costa Rica	66	3.41	55	2.54
Venezuela	67	3.34	69	2.18
Moldova	68	3.31	74	2.13
Kazakhstan	69	3.25	68	2.18
Colombia	70	3.25	72	2.13
Maldives	71	3.16	88	1.96
Armenia	72	3.12	81	2.03
China	73	3.11	90	1.95
Peru	74	3.11	71	2.15
Mexico	75	3.09	64	2.38
Jordan	76	3.06	65	2.36
Oman	77	3.00	76	2.12

Economy	Rank 2007	IDI 2007	Rank 2002	IDI 2002
Iran (I.R.)	78	2.94	92	1.93
Palestine	79	2.92	67	2.20
Georgia	80	2.91	75	2.13
Libya	81	2.84	78	2.08
Ecuador	82	2.75	85	1.97
Tunisia	83	2.73	94	1.86
Fiji	84	2.73	83	2.00
Albania	85	2.73	93	1.92
Azerbaijan	86	2.71	100	1.71
South Africa	87	2.70	77	2.11
Mongolia	88	2.67	84	1.97
Syria	89	2.66	102	1.69
Dominican Rep.	90	2.65	87	1.97
Philippines	91	2.63	79	2.07
Viet Nam	92	2.61	107	1.59
Kyrgyzstan	93	2.61	86	1.97
Egypt	94	2.54	95	1.81
Cuba	95	2.53	91	1.94
Paraguay	96	2.52	82	2.02
Algeria	97	2.51	105	1.61
Bolivia	98	2.45	80	2.03
El Salvador	99	2.43	99	1.74
Sri Lanka	100	2.38	97	1.75
Morocco	101	2.34	111	1.37
Honduras	102	2.28	114	1.31
Guatemala	103	2.28	106	1.60
Turkmenistan	104	2.23	89	1.96
Cape Verde	105	2.18	103	1.67
Tajikistan	106	2.14	96	1.76
Gabon	107	2.14	110	1.48
Indonesia	108	2.13	109	1.54
Botswana	109	2.10	101	1.70
Uzbekistan	110	2.05	98	1.75
Nicaragua	111	2.03	112	1.37
Namibia	112	1.92	108	1.58
Swaziland	113	1.73	113	1.32
Ghana	114	1.63	122	1.10
Bhutan	115	1.63	118	1.17
Kenya	116	1.62	116	1.21
Lao P.D.R.	117	1.60	125	1.08
India	118	1.59	117	1.19
Myanmar	119	1.57	104	1.64
Sudan	120	1.56	131	1.03
Cambodia	121	1.53	126	1.07
Gambia	122	1.49	139	0.96
Lesotho	123	1.48	119	1.15
Yemen	124	1.47	129	1.04
Cameroon	125	1.46	120	1.12
Zimbabwe	126	1.46	115	1.29
Pakistan	127	1.46	146	0.89
Côte d'Ivoire	128	1.41	134	1.01
Zambia	129	1.39	124	1.08
Nigeria	130	1.39	123	1.09
Senegal	131	1.38	142	0.95
Congo	132	1.37	121	1.10
Madagascar	133	1.36	140	0.96
Mauritania	134	1.36	135	1.00
Benin	135	1.28	149	0.76
Haiti	136	1.27	127	1.05
Togo	137	1.26	130	1.03
Bangladesh	138	1.26	132	1.02
Nepal	139	1.23	133	1.01
Uganda	140	1.21	143	0.92
Malawi	141	1.17	141	0.95
Comoros	142	1.17	145	0.91
Rwanda	143	1.17	136	0.99
Papua New Guinea	144	1.14	128	1.05
Tanzania	145	1.13	138	0.96
Mali	146	1.12	150	0.75
Ethiopia	147	1.03	147	0.78
Mozambique	148	1.02	148	0.77
Eritrea	149	1.00	137	0.96
Burkina Faso	150	0.97	151	0.68
D.R. Congo	151	0.95	144	0.92
Guinea-Bissau	152	0.90	153	0.56
Chad	153	0.83	152	0.65
Niger	154	0.82	154	0.51

Source: ITU.

Figure 5: ICT Development Index (IDI) 2002 – 2007

Mobile phones statistics

United Arab Emirates-based telco Etisalat has announced it has reached 500,000 residential broadband subscribers, bringing the country's broadband penetration rate to 11.4%[5]. Gulf Daily News reports[6] that Saudi Telecom Company (STC) has cut internet tariffs by 70%. Currently, Asiacell is at the forefront of **Iraq**'s mobile telecoms sector, at the end of 2008 it had a wireless subscriber base of six million, up from 4.2 million a year earlier[7]. **Qatar** Telecom has posted a high record of more than 1.8 million mobile customers in Qatar[8]. The subscribers have now exceeded the total population of Qatar. **Lebanon** will see a downfall in the mobile call rates by April'09[9], the country's subscriber base is likely to rise to 2 million. In **Oman**, the subscriber base has crossed the mark of three-million for the first time[10].

The new Arab Advisors' survey[11] of **Jordan**'s Internet users also revealed that WiMAX operators had an 8.3% share of residential broadband Internet accounts in the country. Retail e-commerce in Jordan reached an estimated US$ 181 million the 12 months between November 2007 and November 2008. **Moroccan** regulator Agence Nationale de Reglementation de Telecom (ANRT) has released its statistical report for the second quarter of 2008[12]. Total mobile telephony subscribers increased by 3.86%, or 796,000 in the three months to the end of June to 21.412 million subscribers.
Israel dumped from third to fifth position in the Middle-East & Africa (MEA) penetration rankings in Q1 2008 as both the Seychelles and Qatar overtook it to line up behind the United Arab Emirates and Bahrain, which have held the top two places for the last year. If the second quarter even came close to matching the first in the UAE, then that market will have finished June with a penetration rate in excess of 200%, the rate having stood at 192.5% at the end of March.

[5] 16/3/2009
[6] 12/3/2009
[7] 11/3/2009
[8] 10/3/2009
[9] 3 /2/2009
[10] 15/12/2008
[11] 2/12/2008
[12] 17/7/2008

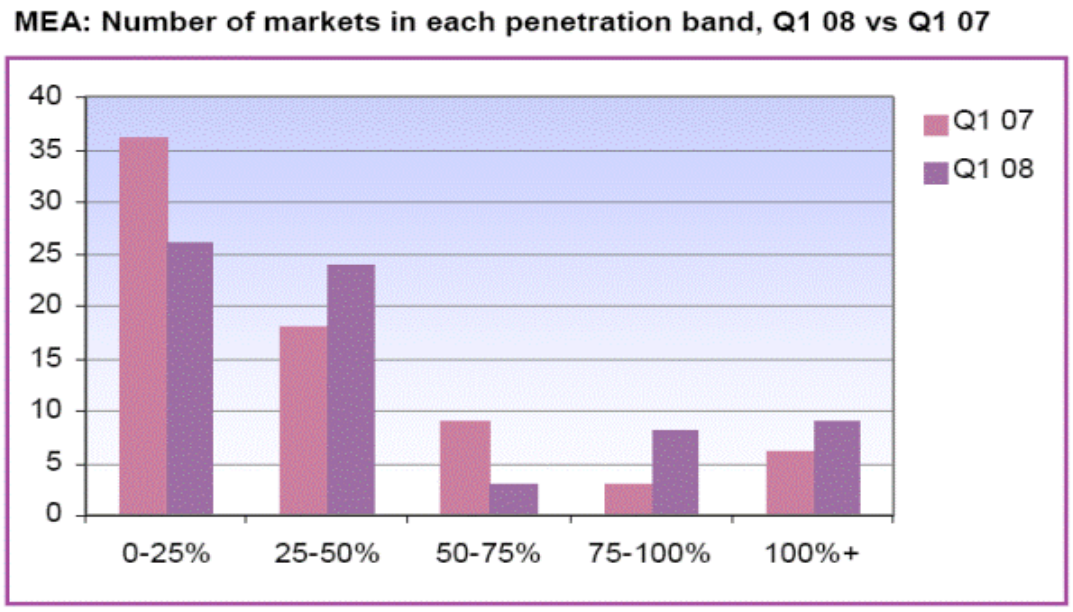


Figure 6: Number of markets in each penetration band, Q1 08 vs Q1 07

The chart shows the fastest growing businesses in the MEA region over a twelve month view[13]. The list includes nine of the names which features in the Q1 08 list, with Zain Iraq coming in at Algerie Telecom Mobile's expense. The top two places are the same in the year as they are in the quarter, with TCI's 9.12m just shading Irancell's 7.95m. The other really strong performances in the region were spread across five main markets - with Mobily in Saudi Arabia taking third place ahead of two Egyptian companies, three from Nigeria and one each from Iraq and Kenya.

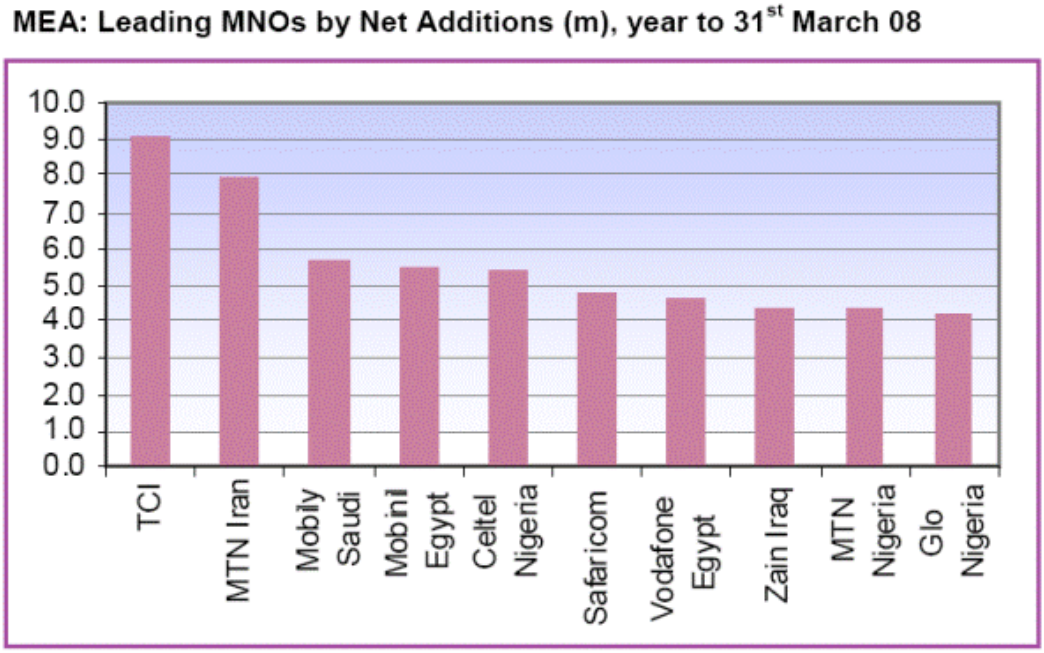


Figure 7: Leading Mobile National Operators by Net Additions (in Millions)
1 January – 31 March 2008

[13] 3/7/2008

Piracy rate
The piracy rate by region is:

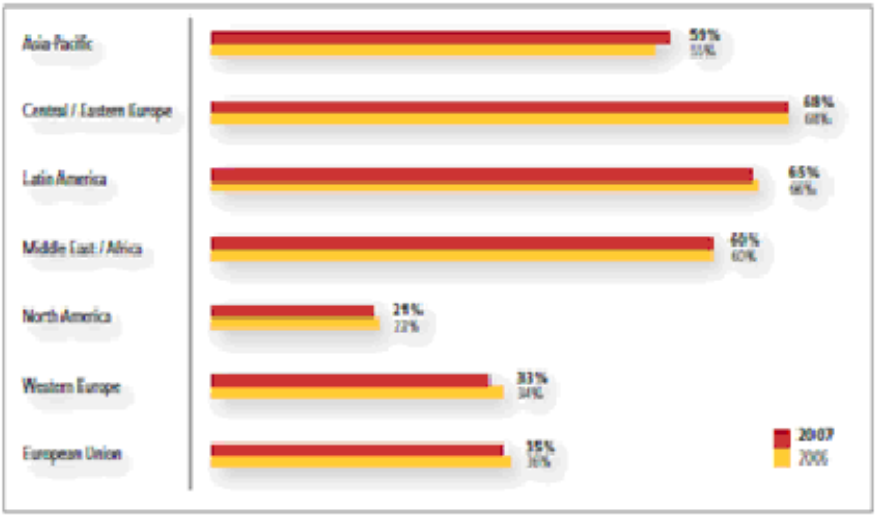


Figure 8: Piracy rate per region

Yemen is the first Arabic country (ranked 7 with 89%), Libya (ranked 8 with 88%), Algeria ranked 14 with 84%, UAE is within the low rates (with 35%). Egypt's government made a deal with right holders to provide software packages for government and educational use, piracy rate dropped to 60%. Similar efforts were taken in Jordan (60%), Saudi Arabia (piracy dropped to 51%). Tunisia piracy rate is about 76% while Morocco is 67%.

It is crucial to relate these rankings to the losses suffered by software industry:

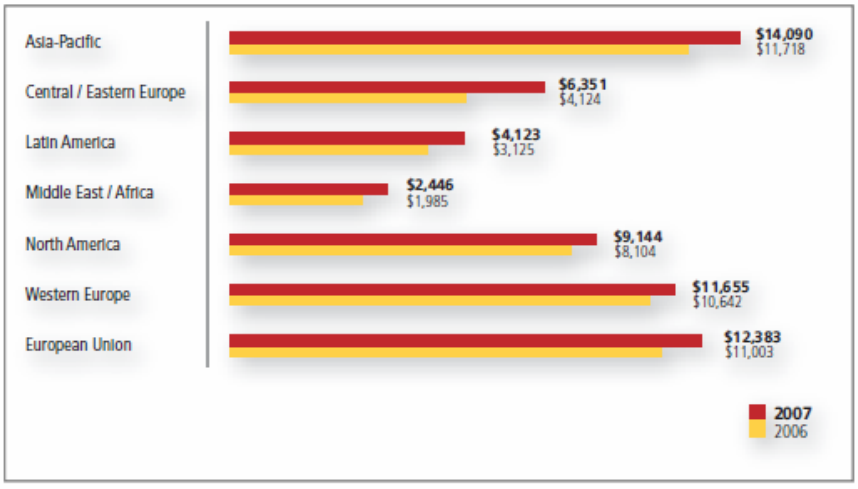


Figure 9: Dollar losses by region ($M) 2007

This illustration should also be interpreted as an indicator of the whole market size. With 60% of piracy and losses of 2,446M$ we estimate the whole market at 4,000M$ which is certainly not big enough for a mass-market (but may be attractive for niche players).

Sales of hardware, software and support

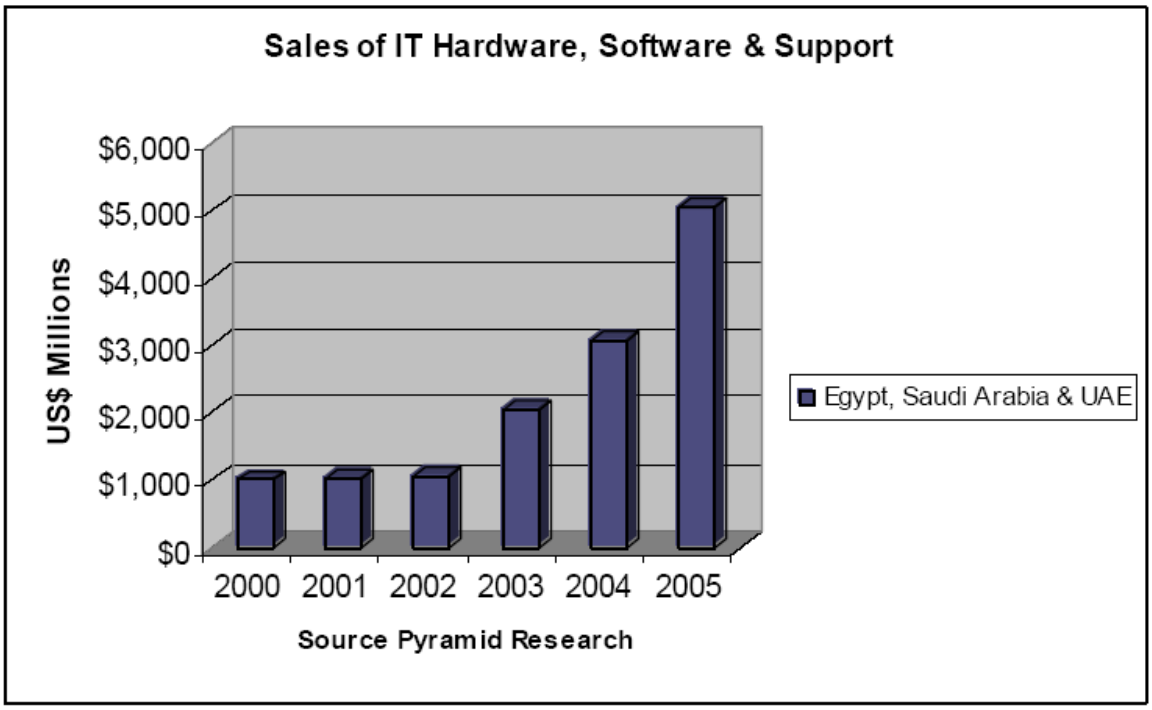


Figure 10: Sales of hardware, software and support

Appendix 4: Potential Funding Agencies in some of the Arab Countries

The following is a list of some Arab countries, under each is a list of potential funding agencies who are involved in providing support for research projects in various areas in the Arab World:

a. Jordan:
   - Royal Scientific Society – Jordan: UNDL project
   - Scientific Research Fund at Ministry of Higher Education & Scientific Research of Jordan
   - Shouman Fund for Young Scientists
   - The Arabic Language Association - Amman

b. Kingdom of Saudi Arabia
   - King Abdul Aziz City for Science & Technology (KACST)
   - King Abdullah Education Initiative
   - King Fahad Library (interests in Supporting Translations Activities)
   - Islamic Development Bank of Jeddah

c. Morocco
   - Agence Nationale de Réglementation des Télécommunications, (ANRT)
   - APEBI (Association des Professionnels des technologies de l'information), http://www.apebi.org.ma/apebi.php?id_rubrique=10

d. Egypt:
   - ITIDA (Industry IT Development Agency) www.ITIDA.gov.eg
   - The Library of Alexandria is responsible for the Arabic part within the UNDL project.

e. Qatar
   - Qatar National Foundation (QNF)
   - Qatar Educational Council

f. United Arab Emirates
   - Mohammad Ibn Rashed Funds for Innovation
   - Abdul Latif Jamil Fund for Innovation

g. Kuwait
   - KFAS: Kuwait Foundation for Advancement of Science:
   - KISR: Kuwait Institute of Scientific Research
   - Al-Babtain Funds

h. Lebanon
   - Arab Thought Foundation

i. Syria
   - The Arabic Language Association - Damascus

- The Syrian Computer Society (which supports high quality translation to Arabic of a set of Computer Science books)

j. International Organizations:
- ALECSO - Arab League Educational, Cultural and Scientific Organization
- ISESCO - Islamic Educational, Scientific and Cultural Organization
- UNESCO and other UN organizations have expressed interest to develop HLT for the Arabic language.

k. European Commission:
- Framework Programmes