MEDAR

Mediterranean Arabic Language and Speech Technology

Deliverable 4.3

Education Collaboration

Authors: Bente Maegaard, University of Copenhagen, Mustafa Yaseen, Amman University, Mohamed Attia, RDI, Chafic Mokbel, University of Balamand, Abdelhak Mouradi, University Mohammed V Souissi.

August 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 4
2. INTRODUCTION, MOTIVATION AND RATIONALE FOR COOPERATION IN UNIVERSITY EDUCATION 4
3. FOUNDATIONS SET BY NEMLAR-MEDAR 5
   3.1. PROPOSED COOPERATION ROADMAP FOR THE EDUCATION 5
4. DESCRIPTION OF FOUR CURRICULA, ANALYSIS OF SIMILARITY AND DIFFERENCES 7
   4.1 INTRODUCTORY REMARKS 7
   4.2. THE FACULTY OF IT AT AMMAN UNIVERSITY CURRICULA AND JORDAN PERSPECTIVE 8
   4.3. THE ARAB ACADEMY FOR SCIENCE & TECHNOLOGY AND MARITIME TRANSPORT CURRICULA AND EGYPT PERSPECTIVE 10
   4.4. THE UNIVERSITY OF BALAMAND CURRICULA AND A VIEW ON THE LEBANESE SITUATION 14
   4.5. THE ENSIAS HIGH SCHOOL CURRICULA AND MOROCCO SITUATION 16
   4.6. ANALYSIS OF SIMILARITIES AND DIFFERENCES 18
   4.7. ACTIVITIES AT THE UNIVERSITY OF COPENHAGEN 19
5. COMMON CURRICULA FOR FUTURE COLLABORATION 19
6. CURRICULUM TO BE IMPLEMENTED IN ROADMAP PHASE 1 (2010-2012) 20
7. COOPERATION WITH EUROPEAN PARTNERS: 22
8. IMPLEMENTATION PLAN AND STEPS TO BE TAKEN: 23
9. EXCHANGE OF POSTGRADUATE STUDENTS AND THESIS SUPERVISION (ROADMAP PHASE 2: 2014-2020) 24
10. FINAL REMARKS 25

1. Executive Summary

In the MEDAR Cooperation Roadmap, several elements for cooperation and for advancing language technology for Arabic were mentioned. One of the most important elements is the human resources. Advances in this field can only be made if a sufficient amount of well trained staff is available. Consequently, university education is a key element, and the one MEDAR has chosen as an example in order to see how cooperation can be implemented. This report concludes that there is certainly a will to collaborate, and makes a proposal for how it can be done. The current work includes four universities only, but it is foreseen that it can be extended to many more universities in the same countries and to other countries, once the success has been proven. The proposal also takes collaboration with European universities into account with University of Copenhagen as the example.

2. Introduction, motivation and rationale for cooperation in university education

In the MEDAR Cooperation Roadmap, published in 2009, several elements for cooperation and for advancing language technology for Arabic were mentioned. HLT is a multidisciplinary technology bridging several basic disciplines. There is a strong belief that a good learning process would form young researchers to better understand the different aspects of a complex engine, for example a statistical machine translation algorithm, in order to make significant advances with the hope of some significant breakthrough at different levels.

Key elements for the promotion of Arabic language technology are of course the *players*, i.e. the organizations involved and the human resources they have at their disposal. Human resources have to be highly skilled, and therefore education is a crucial factor in the roadmap. The emphasis on education came from its significance as a driving force towards human resources development. We are looking at universities in the Arab world to take the lead in this direction, through providing the theoretical sound background, proper training and good focused education and proper infrastructures (soft and hard) that are capable of developing proper sustainable industry.

A side-effect of choosing education and training as one of the most important priority areas of collaboration between Arabic speaking countries is that it involves:

- Academic: to incorporate new degrees at the Bachelor, Master and PhD level relevant to the new domains of interests related to HLT and more specifically Arabic HLT. This would include curriculum development, creating new postgraduate programs and related R&D activities, and consequently would mean further development of the academic teams.
- Research: As this is one of the major activities in universities, and would be directed to relevant areas in Arabic HLT.

- Bridging the gap with software industry especially with those interested in the domain, and/or to create and promote interests to attract companies to get involved and develop products for the market.

MEDAR has chosen to investigate how a subset of the universities represented in the consortium might cooperate and establish a joint curriculum. As a first implementation, we propose to establish common courses to teach in:

- The Faculty of IT at Amman University,
- The Arab Academy for Science & Technology and Maritime Transport (AAST-MT) - The College of Computing and Information Technology, Cairo,
- The University of Balamand,
- The ENSIAS high school at Mohammed V Souissi University.

3. Foundations set by NEMLAR-MEDAR

As mentioned above, the point of departure for this work has been the analysis and the recommendations of the roadmap section 5.1, repeated here:

3.1. Proposed Cooperation Roadmap for the Education

The following is extracted from the original Cooperation Roadmap. The activities in Phase 1, have already begun, maybe not as fully complete programs, but rather on developing course material, research projects and graduate studies based on interests of some of the faculty members as in the examples provided by the various institutes covered in this report (Jordan, Egypt, Lebanon and Morocco).

Phase 1 (2010-2012): Laying the foundations

Political/Policy:

- Work towards accrediting and recognition of the newly developed joint degree programs between the participating universities (those mentioned in this report) in the participating countries (Jordan, Egypt, Lebanon and Morocco).

Training and Research:

- Develop initial HLT-enabling curricula, training material, faculty member exchange programs, Arab student placements programs in close collaboration among Arab universities and between universities in EU and Arab countries.
- Develop essential BLARK components, i.e. tools and LRs. This would start at universities as pilot project, which could be seeds for either companies to be established to market and develop such products; or

extension and enhancements of already existing products. Indeed, this activity is exemplified in the cooperation between universities and industry in Jordan[1] and Egypt[2].

[1] Faculty of Information Technology, Amman University and Arabtext company
[2] RDI and AAST-MT - The College of Computing and Information Technology

Phase 2 (2012-2014): Moving forward

Political/Policy:

- Expand the accreditation and recognition of the newly developed joint degree programs in all other Arab countries.

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.
- Research should support the following areas of application[3]: 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.

Phase 3 (2013-2015): First consolidation

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.

In order to achieve success in the strategies related to education, and consequently, implementation of the components and directions in the different stages in the cooperation roadmap related to education:

- Universities and research centres should provide the basic and applied research in cooperation with industry to produce solid products,
- 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)

These elements were already emphasized in the developed cooperation roadmap, thus we believed that by creating networks among the Arab universities in the region, and connecting these networks to European counterparts, mainly from members in the

[3] Some work have been already started in those areas, this will be used to build future directions and will make a good base for further research and cooperation between the various parties, this could be seen in the various examples covered by the institutions who participated in this report

consortium, and from others who have interests in general in the HLT domain, not necessarily Arabic HLT.

Recommendations, Cooperation Roadmap
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 project 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.

Some of these recommendations are taken up in the present report, others will be taken up later.

4. Description of four curricula, analysis of similarity and differences

4.1 Introductory remarks

The four universities reported in this report realized (in accordance with the Roadmap report) that the time is not ready yet to launch a complete specialized programs in Arabic HLT, but rather to do some directed research and/or teaching on both the undergraduate and graduate levels, which will result in trained graduates who are ready to participate in such industry, and who will form the nucleus that could be built around and maybe in the upcoming phases, new stand alone programs could be launched.

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 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 is 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 which are adapted to the pre-existing knowledge the students have and (at least equally importantly) show the students clear links between what they are doing and what HLT would need or offer.

In the following sections, we give a short introduction for each country followed by an overview of related activities and plans in the participating university:

4.2. The Faculty of IT at Amman University curricula and Jordan perspective

At Amman University, there are Information Technology departments that award degrees in Computer Science, Computer Information Systems and Software Engineering. Typical IT courses are offered including Artificial Intelligence, Data Mining, Business Intelligence, etc. No courses are related or even close to the area of HLT or NLP. In the Faculty of Arts, there are departments who teach languages and linguistics but from angles which concentrate on literature and structure of languages. There are neither degrees nor even courses that introduce computational linguistics, HLT or any related field. This applies to all programs in almost all other Jordanian Universities. Same applies to some universities in Jordan (Petra, Zayttona, Zarqa, Yarmook, Jordan University for Science and Technology).

We found some scattered courses, especially on the postgraduate level that deal with HLT in one or the other way. To be more specific, at Jordan University in the faculty of Arts, some faculty members are interested and specialized in the computational linguistics, and thus offering some specific courses in the domain. In Arab Academy for Banking and Finance Sciences, there are postgraduate degrees, M.Sc. & Ph.D. in the Faculty of IT (Computer Science and Computer Information Systems) where dissertations are in the HLT and Arabic NLP, and many degrees were awarded.

The faculty members in this institution are interested in the areas of: Information Retrieval; Phrase-based Evaluation; Text Alignments; Word Alignment Systems; Bootstrapping Parallel Corpora; Multilingual Resources for Entity Extraction; Text Summaries, Extraction & Abstraction; Question Answering Systems; Transliteration of Proper Names; Text Analysis, including Syntactic Analysis (grammars and parsing), Semantic Analysis (word and sentence meaning), and Discourse Analysis; Morphology; Semantics; Corpus Analysis; Speech Recognition; Pattern Recognition; Automatic Translation; Spell Checking; Lexicons; …etc.

Some specific research courses are given in those programs reflecting the areas of interest of the faculty members. On the level of conferences, there are no specialized ones in Jordan yet whose main theme and concentration is in the areas of HLT; but we found some sessions, and some partial areas of interest in the Arabic NLP, where some papers and projects are presented.

We at Amman University are cooperating with those institutions where some of our faculty members in the faculty of IT, supervise graduate students, give seminars and courses, participate in the examination committees and evaluate theses and dissertations proposals. Still much work is needed to enhance the quality and to have more focus in such programs and courses. Collaboration and cooperation are needed from experts in the domain, especially in education, maybe from Europe and the US to transfer the skills and knowledge into our institutions.

We also maintain good relationships and cooperation with companies (e.g. Arab Textware) who are specialized in the Arabic Language Technology. We participate in the activities in the company, provide consultancy services, and have the technical team from the company to work closely with our students and staff. They also give us training sessions, seminars and participate in supervising some of the projects.

Activities at Amman University related to Curriculum

As part of the curriculum for the degrees offered at the faculty of IT, we developed a course *Special Topics in Information Technology*, with the purpose of launching a course which introduces the basic concepts of HLT and has some concentration on Arabic NLP; so students will get acquainted with this emerging field that we are trying to promote.

In this course we attempted to introduce our students, with strong background in computer science and information technology in general, with a minimal exposure to A.I., to the new areas of HLT or NLP and more specifically to using Arabic as the domain of application. We assumed that each student who will enroll in this course must have: good programming skills, knowledge of data structures, some knowledge of formal grammars, familiarity with the basics of propositional and predicate logic, and with elementary probability.

Our students practically have neither experience nor knowledge in the areas of HLT and NLP; they have very basic knowledge in Arabic language and linguistics (as Arabic speakers). The objectives of the course are:

- Expose students to a new area with concentration on Computational Linguistics.
- Equip students with basic knowledge of the domain.
- Expose students to technologies, utilities and resources in the domain.
- Build the interest in the students in the HLT in general and in Arabic NLP in particular.

Some learning outcomes are defined for the course as follows:
Students will be able at the end of the course to:

- Demonstrate knowledge in the domain of HLT and Arabic NLP.
- Develop a prototype application in one of the areas based on their interest
- Extend the prototype application to a graduation project which will enhance and magnify the knowledge they gain and will be potentially demonstrate capabilities and skills in tools and/or resources related to the Arabic NLP
- Pursue postgraduate studies in this domain.

The course included the following topics:

1) Nature of language; including definition of the language; components of language; features and characteristics of language; historical perspective; introduction to linguistics and looking at the language as a system.
2) Introduction to Phonetics and Phonology, with brief description of the International Phonetic Alphabet; applying phonetics to Arabic.
3) Introduction to Morphology, Part of Speech, Parsing & Tagging, Decomposition and Synthesis; defining Arabic as Inflected language; basic concepts in Arabic Morphology, Role of Morphology in disambiguation; Analysis of some Morphological Analyzers.
4) Introduction to Syntax; Structural Ambiguity; Lexical Ambiguity; Grammatical Description Approaches: The Traditional Approach, The Perspective Approach and The Descriptive Approach; Simple and Compound Sentences; Transformational Grammar; Examples from and analysis of Arabic Grammar and Syntax.
5) Introduction to Semantics; Semantic features and Semantic fields; Role of morphology in building meaning; Role of syntax in building meaning; Introduction to Lexicons.
6) Computer applications/systems and language with emphasis on Arabic Language: Discussions and analysis of Samples of language resources and systems; and developments of prototypes.

This course was taught jointly from the Arabic Language Dept. and from the Faculty of Information Technology, Computer Science Dept. It was offered this term (Spring of 2009/2010), results were encouraging but not good enough. The course still needs enhancements and development; it is intended to offer this course in the Fall of academic year 2010-2011; students will be encouraged to enroll in this course. The current enrolments were 10 students from the final year, three of them chose to develop their Graduation Project based on the prototype they developed, the three worked on a small subset of the language Morphological Analyzer.

4.3. The Arab Academy for Science & Technology and Maritime Transport curricula and Egypt perspective

Situation in Egypt

The purpose of this section is to give a brief idea on the courses taught in the academic institutions of Egypt.

The writer has scanned the internet, contacted a handful of such institutions, and consulted the two main “non-for-profit” NLP/HLT Egyptian societies which are in turn mostly academic in nature, and then briefly reported his findings in this document with specific examples.

NLP/HLT Societies in Egypt:

1. The Egyptian Society of Language Engineering (ESoLE)[4] and is academic in nature since its establishment in 1996. Its main activity is the organization of its annual conference[5], the publication of a journal on Language Engineering[6], and several workshops, seminars, and tutorials over the year calendar. The ESoLE has been first hosted by the Faculty of Engineering of Ain Shams University in Cairo.

2. The Arabic Language TEchnology Center (ALTEC)[7] (*the website is yet under construction*) is a recently established (since Feb. 2010) society that is trying to combine leading academic with industrial figures and visions. It comes with a plan that can be said to be an Egyptian version of MEDAR's roadmap for Arabic NLP/HLT; see http://infos2010.fci.cu.edu.eg/workshop2.php.

***Other regular NLP/HLT events in Egypt**:*
The Natural Language Processing (NLP) and Knowledge Mining (KM) track in Infos conference (now under IEEE)[8] held annually by the Faculty of Computers and Information/Cairo University.

***Egyptian Academic Institutions with considerable NLP/HLT activities**:*
Egypt has about 40 public (government-funded) universities; e.g. Cairo University, Ain-Shams University, Alexandria University …, etc., and 20 other (self-funded) Universities; e.g. American University in Cairo AUC), Arab Academy for Science and Technology (AAST), German University in Cairo (GUC), …, etc.

Almost all these universities have Faculties of Engineering, and Faculties of Computers and Information.

In the departments of Electronics and Electrical Communications of the Faculties of Engineering in these universities, one can find two relevant activities:

a. In the final year of undergraduate study (esp. in the second term), usually students have at least one elective course on some advanced topic. Two of the options in these elective courses relevant to HLT - that are appearing regularly in the last decade – are Fundamentals of Digital Speech Processing, and Fundamentals of Pattern Recognition.
b. In some tracks of the M.Sc. prequalification year of these departments, these courses are compulsory.
c. Many graduation projects (no statistics to assess a percentage) of the final-year students in such departments are devoted to implement basic systems of Digital Speech Processing, Pattern Recognition, and to a lesser degree NLP. Arabic and/or English speech and text are typically the subject material of these projects.

[4] http://eng.asu.edu.eg/esole/
[5] http://eng.asu.edu.eg/esole/Files_Both/Conf_Full.html
[6] http://eng.asu.edu.eg/esole/Files_En/Bulletin_En.html
[7] www.ALTEC-Center.org
[8] http://infos2010.fci.cu.edu.eg/index.php

d. Many theses of postgraduate students of these departments are tackling Digital Speech Processing, Pattern Recognition, and NLP problems. The research done in these departments typically focus on the issues of mathematical modeling, numerical processing, system architecture, and efficient implementation.

In the departments of Computer Science of about one third of the Faculties of Computers and Information and Faculties of Engineering, one can find the following relevant activities:

a. In the final and/or pre-final year of undergraduate study, students typically are subject to one or more courses on: recognition/classification methods, and another one or more courses on formal and natural language processing. See for example the curricula of the Computer Science dept. in the Faculty of Computers and Information of Cairo University http://www.fci-cu.edu.eg/?pg=academic/coursecatalog.php&query=EXPAND esp. http://www.fci-cu.edu.eg/academic/coursedetails.php?cid=CS443, http://www.fci-cu.edu.eg/academic/coursedetails.php?cid=CS462, and http://www.fci-cu.edu.eg/academic/coursedetails.php?cid=IT342.
b. In M.Sc. prequalification year of these departments, at least some of these courses are compulsory, with the option of one or more elective HLT courses on HLT; e.g. the course http://www.rdi-eg.com/AAST_NLP_2008.htm titled "An Introduction to NLP" taught at the Arab Academy for Technology & Science and Marine Transport (AAST-MT)/College of Computing & Information Technology/M.Sc. program, Computer Science; 2008/2009.
c. Many graduation projects (no statistics to assess a percentage) of the final-year students in such departments are devoted to implement basic systems of NLP and to a lesser degree recognition. Arabic and/or English text is typically the subject material of these projects.
d. Many theses of postgraduate students of these departments are tackling written and spoken language processing problems. The research done in these departments typically focus on the issues of formal modeling, inference, search methods, and complexity reduction of such problems.

In the departments of Information Technology of the Faculties of Computers and Information and Faculties of Engineering, one can report similar findings (like the last article of the report) but with focus on topics of Information Retrieval, and Data Mining (applied on Text Mining problems).

In recent years, we have started to see the seeds of Computational Linguistics in a limited no. of faculties of Arts and Languages esp. in Alexandria Univ., Ain-Shams Univ., and the American University of Cairo. One can even spot theses at these universities (add Cairo Univ. in this regard) on HLT problems done in these faculties with joint supervision with faculties of Engineering as well as Computers and Information; see for example

http://www.rdi-eg.com/Downloads/Scientific%20Papers/Quran%20Phonology;Quran%20reciting%20rules%20based%20on%20modern%20acoustics_MScThesis%20By%20AhmadRaghib_Oct2004.pdf.

Concluding remarks on HLT in Egypt in general:
The faculties of Engineering & Computers and Information and to a lesser degree Languages and Arts esp. in the Cairo University, Ain-Shams University, Alexandria University, The American University in Cairo (AUC), and the Arab Academy for Science & Technology and Maritime Transport (AAST-MT) are the ones who are most active in educating and conducting research on HLT/NLP.

These activities provide a seed for taking off with this field in Egypt, but are still not enough for a large-scale industry.

Lack of standardization among the surveyed curricula, and the integration with other studies material are apparent weak points. Standardisation and collaboration with others will therefore in itself raise the quality.

Activities at the Arab Academy for Science & Technology and Maritime Transport

The department of Computer Science in the College of Computing and Information Technology of the Arab Academy of Science & Technology and Maritime Transport www.AAST.edu (located in both Alexandria and Cairo - Egypt) is concerned with elevating the human language technologies (HLT) education at the different levels of undergraduate, postgraduate, and research.

At the undergraduate level in the CS department, optional introductory natural language processing (NLP) or digital speech processing (DSP) courses are provided to the students in the final year. At more advanced levels, similar courses are also provided in the M.Sc. prequalification year. Currently, there are 10 theses working on HLT topics (e.g. IR, MT, OCR … etc.) - mostly Arabic enabled ones - are being pursued in the department.

The department has also set up a Computational Linguistics diploma to qualify students with a 1st degree in IT, CS, or Electronics & Electrical Communications to build a career in the field of HLT. The description of this diploma is to be launched in the academic year 2010/2011.

Moreover, the faculty (via its dean) is a founder of the Arabic Language TEchnologies Center (ALTEC) mentioned above. ALTEC adopts an Egyptian version of MEDAR's roadmap. The CS department is encouraging their CS postgraduate students working on some kind of HLT point to be involved into ALTEC activities.

Needs and Horizons of Cooperation with External Partners:

1- For the above mentioned HLT courses for undergraduate and postgraduate students in the department as well as the diploma. The material needs be enriched with lab sessions in order to give the students hands-on experience. Experience of some reputed EU academic partner with a long history in HLT education, may be valuable in this regard (UK universities are normally preferred because of the language).

2- ALTEC in cooperation with the CS dept. of the AAST-MT's college of Computing & IT of CS (and maybe also with the faculty of Computers and Information in the Cairo University) are preparing for a diploma or some other kind of degree to qualify classic Arabic linguists to be computational linguists in order to build enough capacity for building the LR's needed to boost the Arabic HLT applications industry.
The basic aspects of cooperation in this regard are the production process management, validation and evaluation of LR's (where there is still a shortage in our institutions). The valuable expertise of experienced institutions in these aspects like ELDA www.ELDA.org and CST, University of Copenhagen www.cst.ku.dk, make them excellent candidates for partnership to cover these aspects.

3- AAST-MT – in general - has a history in awarding certificates branded with academic institutions in the North America and Europe which proved successful in attracting students to those certificates. Such a branding may also be beneficial for the aforementioned two diplomas.

4- With postgraduate students from the CS dept. in the AAST's faculty of Computing & IT (as well as elsewhere) being involved in ALTEC activities, ALTEC then has a considerable workforce that may allow it to share into joint projects (perhaps with EU partners or consortia) with the advantage it may take high workload at a relatively cheap cost.

5- The typical exchanges, internships, or/and training of postgraduate students as well as young teaching staff (teaching assistants and assistant professors) researchers working on HLT topics with capable EU partners are all welcome.

4.4. The University of Balamand curricula and a view on the Lebanese situation

Situation in Lebanon
Lebanon has a long and rich history with languages in general. Arabic is the official language in the country. However, the Lebanese who do not speak another language are rare. A lot of the old and prestigious press houses of the region have started and operate in Lebanon. There is also a sustained activity in language, especially machine translation. Software development companies have been established and work for the country and abroad projects. This proves that the country is well prepared to a significant activity in HLT. In addition, the higher education in Lebanon is well developed and diverse. About 40 private higher education institutions operate in the country in parallel to the public Lebanese University. Research in HLT is being conducted at different places and not only at the University of Balamand. All the major institutes have significant research activities on languages in general. Several degrees

are also offered which integrate HLT to some degree, for example translation degrees often include some courses on machine translation or on computer linguistics. Regarding HLT one can cite several activities taking place at the Lebanese University and namely at the Institute of Legal Informatics, the Saint Joseph University and, the American University of Beirut.

HLT Activities at the University of Balamand

Within the University of Balamand education for HLT is being conducted at two levels:

- Graduate level in several MSc programs and projects.
- Postgraduate level in several PhDs.

At the MSc level a new degree is being established at the Faculty of Arts and Social Sciences in the Translation department. This degree includes at least two courses related to HLT; a computational linguistic course and a machine translation course. In the computer engineering department a course on speech processing has been developed. Moreover, about ten MSc projects have been conducted in the past few years and cover:

- Speech and language database collection
- Speech recognition
- Speaker recognition
- Language modeling
- Morphological analysis
- Handwriting recognition
- Machine translation

At the postgraduate level several PhDs have been launched in cooperation with European Universities. The PhDs finalized in the past three years:

- Ramy El Hajj, “Reconnaissance hors ligne de texts manuscripts cursifs par l’utilisation de systèmes hybrides et de techniques d’apprentissage automatique” (“Recognition of handwritten texts using hybrid systems and machine learning techniques”), 2007
- Rania El Bayeh, “Reconnaissance de la parole multilingue: Adaptation de modèles acoustiques multilingue vers la langue cible” (“Multingual speech recognition : Adapting multilingual speech models to the target language”), 2009
- Walid Karam, “Imposture audiovisuelle et robustesse de la vérification de l’identité” (“Audiovisual imposture and robustness of speaker verification systems”), 2010

All these PhDs were jointly conducted with ENST-Telecom-ParisTech.

Besides the educational aspects, a speech and image processing group has been constituted at the University. This group has been involved in several research projects and mainly European projects that helped the group to develop know-how and systems through the past years.
(http://www.balamand.edu.lb/english/Research.asp?id=1367&fid=264)

The systems for which expertise exists at University of Balamand:

- BECARS: GMM for speaker recognition (developed at Balamand-ENST)
- HCM: HMM Toolkit (developed at Balamand)
- SRILM: N-Grams (developed at SRI)
- MOSES: SMT

In addition to the above mentioned tools and systems, a large set of smaller tools was developed at the university within different projects.The group also participates at different international competitions in speaker recognition, video indexing, handwriting recognition, speech recognition etc. The group has a particular interest in the Arabic language HLT.

4.5. The ENSIAS high school curricula and Morocco situation

Situation in Morocco

In Morocco, teaching and research in the field of linguistics is traditionally done in the Faculties of Arts and Humanities and also in the King Fahd Higher School of Translation.

Furthermore, the Institute for Studies and Research on Arabisation at Mohammed V Souissi University, conducts research to make Arabic a modern language of work, teaching and science and technology.

The ENSIAS high school curricula

ENSIAS at Mohammed V Souissi University is a high school in software engineering, networks and decision support. There are no specific obligatory courses in HLT but we try to sensitize students to this field through seminars and projects.

For example, we gave in the past a series of seminars covering the following aspects:

Speech synthesis technologies

- Speech coding
- Speech synthesis
- Applications for the Arabic language

Speech recognition technologies

- Speech signal representation
- Hidden Markov Models & speech recognition
- Speaker recognition

On the other hand, students do some projects during the first, the second and the third year of their studies.

Among the projects selected this year we can mention:

Grade	Subject
3A	Etude et configuration du logiciel MOSES : un système de traduction automatique statistique (Study and configuration of the open source statistical translation system MOSES)
2A_1	Etude de quelques logiciels d'alignement de corpus parallèles. Cas du logiciel GIZA++. (Study of some parallel corpora software. The Giza++ case)
2A_2	Outils de traitement automatique du langage. Etude et exploration du logiciel NooJ (Tools for Language Processing. Exploration and Study of the freeware NooJ software)
1A_1	Réalisation d'un mini éditeur de texte pour la langue arabe (Realization of a mini text editor for the Arabic language)
1A_2	Constitution de corpus textuels pour l'apprentissage de modèles statistiques du langage. Application à la langue arabe (Constitution of textual corpora for learning statistical models of language. Application to Arabic)
1A_3	Réalisateur d'un éditeur de signal en utilisant Matlab (Realization of a signal editor using Matlab)

Also, we participate actively in conferences and workshop related to the Arabic language, cf. some selected examples below.

http://www.ensias.ma/cspa09/

http://www.emi.ac.ma/citala2009/scientificcommittee.html

Optional HLT Course at ENSIAS

At ENSIAS, an optional course was proposed to students of the third year (last year) in the software engineering curricula. It is about 12 hours, and is mainly focused on speech analysis and synthesis.

Speech analysis
Students manipulate speech signal using some tools like Goldwave and perform some labs using Matlab.

Speech synthesis
In introduction to speech synthesis methods and techniques are presented to students (synthesis by rules, synthesis by units concatenation, …) and a speech synthesis system, namely MBROLA, is presented.

4.6. Analysis of similarities and differences

The main similarity aspect of the three educational systems of Jordan University, University of Balamand and ENSIAS is that they all provide some optional module of HLT, and none of them offer a full HLT curriculum. This is fully in line with the recommendations of the Cooperation Roadmap, namely to concentrate most efforts on "HLT-enabling" courses, to either engineer/computer scientist or humanities students.

There is obviously a difference in offering courses to engineers/computer scientists and to arts/humanities students. The same course can hardly be offered to students with such different backgrounds. However, there is hope that bringing together students from different background will lead to better creativity in this domain. Therefore, some efforts will be spent during the coming months to investigate the possibility of having a learning process where both areas of knowledge could be combined.

There are also differences related to the length of the course (number of hours). At ENSIAS a 12 hour course is offered, whereas at the other universities the courses are full semester courses of 32 - 50 hours.

At Jordan University the HLT courses are offered at the Faculty of Arts; at the Arab Academy for Banking and Finance Sciences HLT courses are offered by the Faculty of IT. At Amman University a new HLT course is offered at the Faculty of IT. All of these courses in Jordan concentrate on the written language.
There is some interest in Jordan to investigate the International Masters in Natural Language Processing and Human Language Technology offered by four European universities.

At University of Balamand courses are offered by the Faculty of Engineering both within written language and speech, and the ten MSc projects listed cover both the written and spoken language. PhD education is conducted jointly with ENST-Telecom-ParisTech. However, in Jordan this model is not yet implemented due to restrictions

and regulations of Ministry of Higher Education. This is only allowed if there is a joint degree program between a local Jordanian Institute and a foreign institute.

ENSIAS is itself a school of (software) engineering, so their courses are meant for engineering students; here they are similar to the faculties of IT and Engineering of the other universities. A course of 12 hours is offered, as well as a series of seminars within the field of speech. Even if the seminars are related to speech (recognition and synthesis), many of the student projects belong to the field of written language, i.e. the school is able to offer supervision in this field as well (text corpora, MT etc.).

	Engineering/ IT		Arts/ Humanities		Bachelor	Master	PhD
	W	S	W	S			
Jordan University			x			x	x
Arab Academy – AAST-MT	x				x	x	x
Amman University	x					x	x
Balamand University	x	x				x	x
ENSIAS	x	x			x	x	x

4.7. Activities at the University of Copenhagen

As cooperation with European universities is envisaged, we here give a short description of HLT-enabling and HLT activities at the University of Copenhagen (UCPH) as an example of what exists.
At the University of Copenhagen there are courses at BA, MA and PhD level.

- At the BA level, students (from humanities, sciences or other faculties) may take optional courses in Corpus Linguistics, Formal grammar & MT or HLT in Information Retrieval.
- At the MA level UCPH has an MA-degree in *IT and Cognition*, covering computer science elements, information psychology elements and computational linguistics elements. The MA is a 2-year education. Most of it is currently taught in Danish, but it will be taught in English from 2011.
- At the PhD level there are optional courses, also in language technology, and they are taught in English.

UCPH is happy to receive students from the region's countries and also to provide teachers for joint courses.

5. Common curricula for future collaboration

The universities which participated in this report, as seen from their reporting, have focused 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. Those institutions, and may be others to follow, have to increase the number of people with the required skills, and also have to look at the opportunities for the education of a new generation of researchers and developers with adequate skills in HLT.

The education system followed in those universities in this domain, aims 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.

It is also assumed that there is a need to train people to become HLT educators, as this is necessary for a sustainable supply of *HLT-enabled* professionals (as indicated in the Roadmap). Thus the main goal of the education system in the collaborating institutes is to create 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. The participating institutions have not created a completely new HLT discipline with its own professionals, but rather provide graduates capable of carrying out tasks related to supporting HLT industry, and the developed curriculum aims at achieving those objectives, and will provide the basic needed skills.

6. Curriculum to be implemented in Roadmap Phase 1 (2010-2012)

A cooperation plan is proposed between the four institutions to work jointly on a program that will aim at graduating students on all levels (Bachelor, Masters and PhD holders) in the domain of Arabic HLT, this plan will be implemented in the academic year 2011-2012, but initial steps will be taken immediately, as will be shown in section 8. The proposal concentrates on areas as follows:

- Institutes in Jordan (Amman University) and Egypt (The Arab Academy for Science & Technology and Maritime Transport) will concentrate on Written Text.
- Institutes in Lebanon (Balamand) and Morocco (ENSIAS) will concentrate on Speech.

Common background and basic theories will be common for all; any student could transfer credit freely among the institutes, with guaranteed transferred credits. Some difficulties might be encountered if the education systems followed in those universities are different; we assume the four institutes for this program will follow the American Credit Hour System, although some of them have experimented with the European Credit system. The number of credits (American system) will be as followed:

1) Bachelor: 132 Credit Hours
2) Masters: 36 Credit Hours, including 9 Credit hours for Thesis.
3) Ph.D. will be just research based, no courses will be required.

The Curriculum for the Bachelor will be designed following the standard:

1) General Education including courses in communication skills in English.
   Percentage of Curriculum: (25%)

2) Basic IT core courses including programming languages, system analysis and databases, AI, and some aspects of software engineering.
   Percentage of Curriculum: (30%)

3) Basic Arabic Language foundation courses covering: Syntax/Grammar, Morphology, Semantics and Phonology
   Percentage of Curriculum: (30%)

4) Elective courses: Two streams:
   a. Written Text Processing
   b. Speech Processing
   Percentage of Curriculum: (15%)

5) A graduation project will be mandatory for all graduates of this program

On the Masters Level courses will basically be in depth according to the interest of the institutes' faculty members.

Supervision of graduate students (Thesis and Dissertation) will be made jointly as follows: Main supervisor will be from the student's institute, while the co-supervisor will be from any of the other institutes whose interests are in the same domain. Examination and defense committees will be formed from members of all institutes, where they will serve as external examiners, in addition to the local committees.

With the e-learning facilities and video conference technology, it is possible to teach courses remotely and share the content and lecturers among the institutes.

Management:
The program will be managed by a steering committee of five members, one representative from each institute and a representative of a European institute; the committee will serve for a renewable four year term. One of the main objectives for the steering committee in addition to all other academic and management details related to the program is to seek accreditation and recognition in each of the countries of the participating institutions.

After 6 years for the Bachelor program and 4 years for the Masters and PhD programs, the participation of other institutes will be encouraged and this experience will be shared with any interested university in the region.

7. Cooperation with European Partners:

Cooperation and help will be sought from European partners in the MEDAR consortium, which are:

- **University of Copenhagen:** Centre for Language Technology, Denmark
- **University of Utrecht:** Utrecht Institute of Linguistics OTS, the Netherlands
- **Research and Innovation Centre "Athena":** ILSP, Institute for Language and Speech Processing, Greece
- **Université Lumière Lyon2:** Groupe SILAT, France

The cooperation will be in areas of faculty members exchange and sharing of expertise in the HLT domain in general. This will include: teaching, research and supervision of graduate studies, especially at the first stages of the programs. The establishment of long-distance courses on existing platforms has also to be investigated to facilitate the cooperation between the different institutions.

Affiliation and cooperation will also be looked at with similar programs in Europe such as the following:

International Masters in Natural Language Processing and Human Language Technology

This is a program that sounds very interesting, unique and in line with what we are looking for[9]. Along with Université de Franche-Comté (France), Universidade do Algarve (Portugal) and Universitat Autònoma de Barcelona (Spain), the University of Wolverhampton delivers the two year Erasmus Mundus International Masters course during which students attend two or three of the four universities. Students who successfully complete the course are awarded a multiple Masters degree by the universities they study at; i.e. a degree from each university they attend, which together constitute the International Masters in Natural Language Processing and Human Language Technology degree.

Another good example in this area to take into account and to develop relation and cooperation with, is the ISTA (L'Institut Supérieur Arabe de Traduction) founded in 2004. 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).

[9] http://mastermundusnlp-hlt.univ-fcomte.fr/ and http://clg.wlv.ac.uk/teaching/masters.php

8. Implementation Plan and Steps to be taken:

Given the curriculum, thoughts and HE Roadmap explained above, the following steps will be taken to implement this cooperation roadmap: Although the funding issues related to launching this plan are very critical especially with the tight budgets facing institutions in the area, it is proposed that each institution will take care of its part as per related to operations within itself, meaning that the faculty members are already doing their regular work within their institution, and launching this program will be part of their regular work. Cooperation with industry will also be looked at as partial source of funding, especially if the work will involve developing some specific requirements for the industry that could be used and utilized.

HE Phase A: Laying the Foundations & Launching the Programme (2010-2012):

1) Institutes from the four countries, Jordan, Egypt, Lebanon and Morocco agreed to adopt the curriculum as outlined in the above section.
2) Steering committee is formed from representatives from the involved institutions; a representative from CST (UCPH) will be the European member in the committee.
3) Steering committee will work on the details of the program, and on course description. This task will finish January 2011.
4) Certification, Licenses and Accreditation will be sought from different Ministries of Higher Education and Accreditation bodies in all four countries, deadline is June 2011.
5) Programs will be launched on academic year: 2012-2013, where first intakes will start in the fall of 2012 (October 2012), for all three degrees (B.Sc., M.Sc., and Ph.D.)
6) Cooperation with European institutes will be sought, where exchange of students, faculty members and adoption of some courses from those institutes will be taken, especially UCHP and other consortium members.

HE Phase B: Moving Forward (2013-2017):

This phase is anticipated to get beyond the original Phase II as in the roadmap, because the proposed academic program needs to get mature and evaluated through testing the acceptance of the market for the graduates of the program and the strength of the content and curricula of the program matched with the market demand. This is typical duration for any program to test those aspects and improve and/or enhance quality.

The following actions will be undertaken:

1) Graduate students will be exchanged, joint supervision will be encouraged.
2) Cooperation with industry will be looked at to enrich and enhance the programs.
3) Institutions that have similar programs in Europe and the Arab world (such as the MS in HLT mentioned above, and the L'Institut Supérieur Arabe de Traduction) will be approached to find ways of cooperation, and investigate opportunities for adding the Arabic language to the set of languages students can have their Masters in, in addition to the other languages already available.

4) Steering committee will apply for accreditation and recognition of the program in other Arab countries.
5) More cooperation with industry, to create job opportunities for graduates and to enhance the market for Arabic HLT.

HE Phase C: First consolidation (2016-2020)

1) Expand the base and allow more institutions from the Arab world to join the new program.
2) Expand the areas of research and applications in Arabic HLT.
3) Strengthen the ties with industry to create job opportunities for the graduates and to support the newly emerged industry.
4) Strengthen the exchange programs between Arab countries (institutes participating in the program) and other European counterpart.

9. Exchange of postgraduate students and thesis supervision (Roadmap Phase 2: 2014-2020)

During the work with the education cooperation roadmap we have seen the amount of time necessary to implement such cooperation; consequently, we have slightly changed the schedule suggested in the Cooperation Roadmap, as we think this is more realistic.

Mobility at the postgraduate level may be seen as easier to be implemented due to the reduced timing constraint. While at the other educational levels the presence in classes constrain the students to be in one or the other country for a period, if the common course is only to be offered in one country, this constraint does not exist at the postgraduate level. In addition, this mobility allows a good flow of knowledge between the different institutions. One way of sharing resources will be through utilizing video conferences and e-learning techniques in sharing classes and lectures remotely. This would facilitate having specialized faculty members to share their knowledge and expertise without the need to travel and move around, and we would solve the students' mobility associated difficulties.

We suggest that faculty members from the above institutions agree on a set of HLT topics for PhD students based on special interest groups from those who have common interests. We can then consider PhD students mobility and possible joint degrees as explained in the above actions.

Already at this stage, the idea of launching a few PhD projects between the different institutions will be investigated. The support of the European laboratories in this direction is crucial and it is supposed that the PhD students will have to spend some time in the European laboratories, for example UCPH.

In order to establish such cooperation several issues need to be tackled. The process of selection of a PhD project that would be of common interest to two or more parties must be well identified. The process of the selection of the candidate is also crucial as

well as the period of stay in the different laboratories. Finally and as for the other educational levels, the recognition of the degree by the partner institutions is a key to the success of this effort.

10. Final remarks

The work performed by the four universities has shown that there is a clear will to cooperate in the field of HLT education, in order to promote higher education, research, industry collaboration etc. There is a plan for the first implementation of the cooperation, and for overcoming the barriers. Moreover, there will be an evaluation plan to enhance the quality and to correct any problems that might occur through the implementation process.