EECE 6XX – Introduction to Computational Arabic (3 credits)

Catalog description:

This course will focus on knowledge necessary to develop software applications and systems that deals with Arabic data and tends to Arabic users. The course will discuss computational challenges specific to the Arabic language including representation, display, rendering, processing, directionality, structure, interface, and recognition. The course will also discuss multilingual texts where Arabic takes part. We will visit several text processing techniques and algorithms such as encoding, matching, tokenization, search, indexing, and pattern matching and introduce the necessary changes to accommodate the Arabic language. The last part of the course will discuss the state of the art in automating Arabic language processing, understanding, and recognition.

Areas: Software engineering, Arabization, localization

Required or Elective:

Elective for CCE / ECE

Level: Third year, senior or graduate standing

Prerequisites:

By course: EECE 330

By topic: Students are expected to have basic knowledge of data structures, and considerable programming experience. Consent of the professor.

Textbooks:

The course will use journal and conference papers and industrial technical reports.

Course objectives:

The objectives of this course are to give students:	***Correlates to Program Educational Objectives***
Knowledge and experience of the different digital representations of the Arabic language.	**1,2,4**
Knowledge and experience with solutions of computational challenges specific to the Arabic language.	**1,2,3,4**
Knowledge of text processing techniques specific to the Arabic language.	**1,2,3,4**

Experience in building, writing and presenting Arabic software systems.	**2,3,4**
Knowledge and experience in user interfaces specific to the Arabic language.	**1,2,4**

Topics

No.	**Subjects covered**	**75 Min. Lectures**
1	Language definition	1
2	Representations of the Arabic language (native character sets, proprietary representations, transliteration, etc …)	2
3	Rendering and displaying Arabic text (fonts, glyphs, context aware rendering, etc …)	2
4	Bilingual Graphical User Interface	2
5	Arabic string manipulation (storage, retrieval, tokenization, matching, etc …)	2
6	Arabic search algorithms (haystack, regular expressions, suffix and prefix matching, etc …)	4
7	Typesetting scientific Arabic documents	2
8	Dictionary based Arabic applications	2
9	Structure based Arabic applications	2
10	Automated understanding of Arabic	2
11	Automatic recognition of Arabic (OCR)	2

12	Projects and presentations	4

Class/laboratory schedule

a) Two 75-minute lectures per week.

b) Use of computer lab or personal computer is needed for working on the project and the assignments.

Course outcomes:

At the end of the course students should be able to:	***Correlates to Program Outcomes***		
	H	M	L
1. Build software that reads and writes Arabic text.	**A** C	**E**	**F**
2. Design software with Arabic GUI.	**A C**	**E**	
3. Design and implement text algorithms that handle Arabic digital text.	**B** C	**E**	
4. Build simple Arabic text processing utilities such as tokenizing and matching utilities.	**C** E		F
5. Write high render quality scientific Arabic papers.	**A,B**	**E,j**	**F,l**
6. Use Arabic friendly development environments.	**A**	**J**	**F**
7. Port current systems to Arabic or any local language.	**A,d, E**		**F**
8. Understand the issues and complexity involved in Arabic optical character recognition.	**A**	J	F
9. Understand and appreciate the impact of technology in the local Arab world.	**H, i**	A,j	

10. Contribute and disseminate the advanced computational technology to local Arabic users.	**H, i**	B,j	
11. Build and use Arabic understanding libraries such as dictionaries and spell checkers.	**CE**		F

Resources for the course:

Books, arctiles, publications, online material

Evaluation:

1. Class participation: 25 %
2. Exams: 25 %
3. Paper: 20 %
4. Project: 30 %

Students will have topics and read papers and come ready to discuss them in class. They will present on many occasions and will work in teams if possible to appreciate and solve some of the presented problems and techniques. There will be one midterm to evaluate the understanding of students of materials presented in class. Students will write one term paper on a selected subject. An acceptable paper will cover state of the art material and will at least offer a critical review and suggest possible enhancements. Students will submit a project that can be an implementation of the suggested enhancements.

Professional component:

Engineering topics: 80%

General education: 10%

Mathematics and basic sciences: 10%

Computer usage:

Students pick their platforms and programming languages.

Person(s) who prepared this description and date of preparation:

Fadi Zaraket, Dec 2009