Ain Shams University
Faculty of Computer and Information Sciences
Data Strucutres   Project

Title	**Wikipedia Arabizer**
Description	It is not necessary to explain here how great and important the Wikipedia is. I think everyone knows this already. The Arabic Wikipedia (I mean the Wikipedia pages that are available in Arabic) is still, however, far away behind the needs of the Arabic reader. Since the English Wikipedia is currently mature to some extent, one could start translating it into Arabic. Knowing that more than 400 million people speak Arabic, one can imagine the great impact of this. This project aims at providing a tool that makes it *not difficult* to arabize the English Wikipedia. There are three challenges for a person to translate a Wikipedia page: (1) to translate the text, (2) to format the translation using the wiki markup language, and (3) to reproduce the non-text elements such as figure, tables, links, ...etc in the translation. The good news is that these three challenges can be made easy if the suitable SW tools are available. The goal of this project is to develop a kind of an integrated environment that is dedicated for the specific task of helping people translating the Wikipedia pages into Arabic. Following is a typical user scenario that explains what the program should do: 1. The user launches the program. Clear :). 2. Import and preview an English page from the Wikipedia. 3. On the user's request, the program invokes an automatic translation service (e.g. google translate, Yahoo Babel Fish, ...) to produce a rough translation of the page. The Arabic translation is displayed side by side with the English page in the program's GUI. 4. The user edits the Arabic translation. Meanwhile he will be using tools that help him editing and formatting the translation. ***The program is required to fulfill these mandatory requirements:*** 5. Parse the wiki markup language, and generate the preview. 6. Reproduce the elements of the English page in the Arabic translation (e.g., photos, tables, bullets, ...etc). 7. Invoke automatic translation services. 8. Provide toolbars and menus for editing the wiki markup language. 9. Provide a tool for translating the links. The English page will normally contain links to other Wikipedia pages. The Arabic translation needs also to refer to something. It is required to provide a search tool for finding the most relevant Arabic pages, and link to them.
Group size	4 – 6 members.

Ain Shams University
Faculty of Computer and Information Sciences
Data Strucutres   Project

Duration	Until the final data structures exam.
Deliverables	1- A program that implements the requirements above. 2- A short (2 pages) user manual + a technical document describing the used techniques.
Bonus extensions	1. Maintain user profiles, or 2. Manipulate the user's Wikipedia account and publish the translated pages directly from the GUI, or 3. Context sensitive translation and synonyms, or 4. Support languages other than English.
Mentor	Dr. Mahmoud Attia Sakr.
Notes	In order to implement this Plugin you'll need to self study the wiki markup language (easy). The program should have a nice GUI. It can be implemented as a desktop/windows application, or as a web application, or "better" both :). It might cost money to use the automatic translation API. For example the google translation APIs charge 20$ per 1M characters. I am not sure whether there are free translation APIs some where else. Unfortunately, you'll need to pay these costs, if any, from your own pocket. The course instructors didn't negotiate any fund sources. This page might be helpful giving some related ideas: http://en.wikipedia.org/wiki/Computer-assisted_translation This project requires good programming and research skills. Unless you are an A+ team, it might be risky to apply for this idea.