

Arabic Natural Language Processing (NLP)

More than 10 years experience in NLP developed a real competency in our NLP portfolio of products

● ArabMorpho®

The main RDI's Arabic NLP core engine which is the basis of morphological analysis, full word diacritization, prosodic phrasing, POS tagging, and statistical disambiguation.

● ArabDiac®

Automatic Arabic text diacritizer that directly maps to Arabic phonetic transcription based on ArabMorpho®. ArabDiac® can be used as basic infrastructure for many Speech Processing technologies; e.g. Arabic TTS.

● ArabTagger®

Automatic Arabic Part-Of-Speech (POS) tagger based on ArabMorpho®.

● Swift®

Arabic derivative search engine (Indexer & Searcher) based on ArabMorpho®.

● ArabDictions®

RDI's automatic online dictionary binding System that can be used to enrich any Arabic text based content.

Arabic Speech Processing

● Speech Verification

Self learning of Languages, first applied on the recitation rules of the Holy Qur'aan; Haffs®.

● ArabTalk ®

Arabic hybrid/concatenative Text-To-Speech(TTS) with a small size < 20 M.Bytes.

Human Language Tools Assets (HLT's)

● ArabDictions®:

RDI's Arabic dictionaries with about 7,500 text entities explaining every morpheme of Morpho4's that are used to build the whole Arabic vocabulary.

● ArabDictions®:

2,800 roots & 30,000 stems. Expanded version: 4,500 roots, 60,000 stems.

● Arabic Morpho Analysed Corpus®

Morphologically analysed (according to RDI's formalism) and manually revised Arabic text corpus of a size around 1,200,000 morphemes (and persistently growing).

● Arabic POS Tagged Corpus

The POS tagging (according to RDI's formalism) with manual revision of the same Arabic text corpus.

● Arabic concatenative Text To Speech male & Female DB



Qur'aan Tools

RDI is one of very few companies who manipulate Qur'aan as text in Ottoman font (not as static scanned images) which is essential for the other features enlisted below:

- Drivers, editors, and viewers of text in **Ottoman font** with different sizes.
- On-line Mus'haf designer and **viewer**.
- Binding Qur'aan text in Ottoman font with speech material; e.g. Qur'aan Recitation, Spoken Tafseer,....., etc.
- Binding Qur'aan text in Ottoman font with other relevant text in Arabic or in Latin; e.g. Translations, written Tafseer, Asbab Annuzoul, syntax analysis.
- Multi modal search features in the text of Qur'aan; Semantic, Syntactic, and Derivative.
- Dynamic coloring of Qur'aan characters according to their Ahkaam of Tajweed for any combination of these Ahkaam.



Main Projects

- **www.nemlar.org** This project is funded by the **EU** to elevate the status of Arabic Human Language Technologies (HLT'). RDI has been a main participant in the project all over that successful project that endured from Feb. 2003 till Aug. 2005.
- **Egyptian Data Mining Center-of-Excellence** has been launched in late 2005 by the **Egyptian MCIT** to boost high advanced R&D in the field of CIT. RDI had been elected to be one of the founding members of this CoE and is sharing with other consortium members an ambitious project to develop a novel set of Arabic Text Mining Tools with a main contribution of RDI 's Arabic Lexical Semantic Analyzer.
- **OrienTel www.OrienTel.org**; which has been an important Euro-Mediterranean speech resources project organized by giant companies like IBM, Lucent Technologies, Phillips,...., etc. It has been coordinated by the European Language Distribution Agency (ELDA) www.elda.fr to cover telephony speech over the whole middle east. RDI has finished the segment of Egypt as a subcontractor of IBM.
- **Qur'an sciences portal of the Kuwaiti Ministry of Endowments & Religious Affairs** which contracted with RDI to build its new comprehensive Qur'an Sciences Encyclopedia web portal where RDI is deploying its Qur'an tools as well as its novel technology for the interactive self-learning of Tajweed (**Hafss**).

