Automatic translation tool for search terms in Standard Arabic and its dialects

Technology #cu14013

The Arabic language is a collection of spoken dialects along with a standard written language. Spoken dialects can have significant linguistic differences from the written Standard Arabic, complicating translation efforts. This technology describes Dialectal Information Retrieval Assistant (DIRA), a software program which, when given search terms specified by a user in English or Standard Arabic, automatically generates lists of corresponding search terms in different Arabic dialects. Users may configure preferences for dialect variant and inflectional features such as number, aspect, and gender.

A Java implementation and term weighting scheme respectively enable cross-platform integration and tuning of system output for user-specific content.

The DIRA technology can be integrated with various software platforms by virtue of having been developed using Java, a widely used cross-platform programming language. Furthermore, DIRA allows user to customize their content by utilizing a term weighting scheme that tunes the output for particular types of content.

The technology has been demonstrated on a publicly accessible website.

Lead Inventor:
Owen Rambow, Ph.D.

Applications:
- Translation from English or Standard Arabic into various spoken Arabic dialects.
- Automated Arabic translation services.

Advantages:
- Suitable for integration on various operating systems by virtue of having been implemented in Java.
- Result contents can be tailored using a term-weighting scheme to improve translation performance.
• Uses lexemic features to reduce translation errors

**Patent information:**
Patent Pending

**Licensing Status:**
Available for licensing and sponsored research support
Tech Ventures Reference: IR CU14013
Further Information: Columbia | Technology Ventures Email: TechTransfer@columbia.edu

**Inventors**

**Owen Rambow**