



Social Science Automation, Inc. (SSA) offers a wide range of text analysis products and services. Using unstructured data, the Profiler Plus Text Coding Platform performs sophisticated analysis in combination with industry-specific coding schemes. Working together, these systems help you answer your most challenging questions.

## Capabilities for Arabic Natural Language Processing

SSA has developed a number of tools for automated coding of Modern Standard Arabic (MSA). In addition to our coding capabilities, we also have a robust web reporting service, a large corpus of Arabic media documents and the capability to continuously download and analyze new media texts. The following is a brief description of each capability.

### Automated Coding Tools

#### Arabic Tagger Parser

Arabic script does not separate words as unique text strings as in English. In order to analyze text and identify individual terms, SSA has developed a tool which first parses MSA text into individual words, and then tags each term with relevant grammatical information. This tool will:

- Remove attached pronouns, preceding determiners, and preceding prepositions
- Identify parts of speech for all words
- Standardize spellings of verbs and plural forms of nouns and adjectives
- Identify tense, gender, number, and verb form of words
- Identify participles and verbal nouns

#### Arabic Valance

The Arabic Valance tool identifies terms that have positive and negative connotations affiliated with them. The lexicons currently contain 671 terms/variants of positive terms, and 1725 terms/variants of negative terms. Terms are determined based on hand coding by validated hand coders and have been developed primarily against corpuses of media documents that describe America, Israel, and Palestine.

#### Entity Extraction and Identification

The entity extraction and identification tools identify references to specific entities and the people that make up or are affiliated with that entity and label them as part of that entity. These tools identify reference variants, standardize their spellings, and aggregate them at country and/or group levels. In addition to general purpose entity extraction routines, specific routines identify references to all official country names, national adjectives, and capital cities as per US government list of FIPS Codes. Even more comprehensive entity identification routines have been developed for:

- America
- Israel
- Palestine
- Fatah
- Hamas
- Palestinian Islamic Jihad
- Hizbullah
- Lebanon
- Iraq

## Grammatical Connections

The Grammatical Connections tool identifies grammatically determined associations between terms. This tool has served primarily to establish connections between valanced language and entity references. The tool can establish connections between terms and multiple different entities, including establishing connections between single terms associated with multiple entities.

## Data Delivery Methods

Delivery of data is flexible and tailored to a clients needs. An XML feed is available, or SSA has developed an internet based reporting service that allows third party users to access results from the automated coding tools on a dynamic, interactive website. The reports offer the user complete flexibility by allowing the user to choose which indicators are displayed in the charts and graphs, over what time frame, and for which source. These reports include both aggregated and overtime analyses, as well as moving averages and bubble plots. All web reports are dynamic: media texts are downloaded, processed, and results are incorporated on the website within hours of their publication.

## Sources

SSA has collected MSA media documents from several sources. Documents are available for analysis for the sources and dates below. Those with “ongoing” as the end date are available for automated download at client request:

*Al Ahram* (Egypt) 2000 - ongoing  
*Al Manar* (Lebanon Hizbullah) 2006 - ongoing  
*Al Mustaqbal* (Lebanon) 2003 - ongoing  
*Qudsway* (Gaza-PIJ) 1998 - ongoing  
*Al Arabiya* (UAE) 2005 – ongoing

### Available by request:

*Al Ittihad* (UAE)  
*Al Forat* (Iraq)  
*Al Jazeera* (Qatar)  
*Al Watan* (Syria)

*Hayat al Jadid* (West Bank) 2005 - 2009  
*Al Liwa* (Jordan) 2004 - 2009  
*Dar al Hiya* (Lebanon) 2002 - 2009  
*Syria News* (Syria) 2005 - 2009  
*Al Ayum* (Saudi) 2002 - 2009  
*Al Ayyam* (West Bank) 2004 - 2009

*Palestine Information Center* (Gaza-Hamas)  
*Al Thawra* (Syria)  
*Al Watan* (Saudi)  
*Al Zamman* (Iraq)

## About Social Science Automation

Since 1997, Social Science Automation, Inc. has provided state-of-the-art automated text analysis products and services to business, government, and academic clients. Founded by Dr. Michael Young and Dr. Margaret Hermann, the company identified a need within government organizations for automated text analysis services and expert consultation. Since then, SSA has diversified its offerings to include solutions for media analysis, campaign and election media evaluation, athlete achievement, profiling, and forensic psycholinguistics.