Information Transfer

COLTEC

ANEE
Arabic Named Entity Extractor

ANEE: Arabic Named Entity Extraction

Don't just search. *Find meaningful information.*

Finding vital information in a sea of documents can be a daunting task. With ANEE, important data pops to the surface. Using advanced discovery search methodologies, ANEE extracts critical information from large amounts of structured and unstructured data using human semantic concepts – it analyzes concept and context and delivers results based on meaning.

Integrated into data management applications, ANEE is the ideal tool for processing large amounts of data – it discovers names, places, dates, mentions and things to help analysts uncover pertinent data for further analysis, separating the wheat from the chaff. Say you wanted to uncover information linking US cities with anyone of Arab descent – simply select "US cities" and "Arabic names" and every mention of "New York City, Washington DC, Los Angeles, and Boston" found with "Mohamed, Ahmed, Ussama, and Tarek" would be found Just select the category and ANEE finds every mention, comprehensively and automatically. For persons of interest, ANEE even finds and displays alternate names and aliases – it knows when a name and a person are related, even when they are completely different, such as Abu Ammar and Yasser Arafat, for example.

ANEE enables the most comprehensive data discovery for Arabic entities

What sets us apart from the competition is our extensive understanding of the linguistic sophistication of Arabic that comes from our Middle-Eastern roots and years of R&D. As a leading provider of Arabic language processing technology for almost two decades, we've developed the world's most thorough entity extractor using both linguistic and statistical methods – our NLP techniques, developed by experienced Arabic-speaking linguists, provide meaningful context and human-like judgment while computer algorithms using statistical techniques support to increase speed and accuracy. The result? The most efficient, thorough tool for extracting meaningful Arabic information, taking into account all the complexity and subtlety of the Arabic language without being restricted by simple look-up tables or rigid rules.

ANEE features over 25 pre-set entity categories and 100 subcategories such as presidents' names, international airports, municipalities, corporations and government agencies that instantly enable users to discover all matching entities in their database. Through proprietary techniques we have developed, ANEE extracts named entities even if they have prefixes and suffixes.

ANEE extracts ALL related named entities, including those with prefixes and suffixes:

للعراقيين
To the Iraqis (plural)

العراقيون
The Iraqi

عراقي
Iraqi

Information Transfer

COLTEC

ANEE
Arabic Named Entity Extractor

ANEE Arabic Named Entity Extraction

Sample Entity Extraction Output

Named Entities Found (11)

Government Organizations (1)
وزارة الخارجية الإيرانية

Person Names (1)
محمد علي حسيني

Energy/Fuel Related (2)
Nuclear Energy Related
الوقود النووي
تخصيب اليورانيوم

Geographic Locations (3)
Geographic Region (1)
الغرب
Countries (1)
إيران
Capitals (1)
طهران

Military Related (2)
العسكرية
Weapons: Nuclear (1)
أسلحة نووية

أكد المتحدث باسم وزارة الخارجية الإيرانية محمد علي حسيني مجددا تمسك بلاده بتقنية الوقود النووي.
وأشار في مؤتمر صحفي طهران إلى أن بلاده لن تناقش "حقها الصريح" في إتقان دورة الوقود النووي وتخصيب اليورانيوم.

لكنه أعرب عن استعداد طهران لإجراء محادثات من شأنها طمأنة الغرب بأن خططها لا تهدف لصنع أسلحة نووية.

وأضاف "ينبغي أن يكون هناك هدف للمحادثات ولن يناقش حق إيران الصريح، نريد محادثات بلا شروط مسبقة لإزالة مواطن الغموض وطمأنة الأطراف الأخرى بأنه لن يكون هناك تحول للاستخدامات العسكرية.

Based on years of research and development coupled with internationally recognized ontologies, ANEE's taxonomy can be easily customized to suit your specific needs, including adding taxonomies or modifying entity concepts – for example, one person's idea of "terrorist" can be different from the next. In addition, extracted entities – by themselves or, for context, in the sentence in which they appear– can be exported into lists for additional analysis and transcription. Using a tool such as our WORDCON, extracted entity lists can be easily romanized from the original Arabic, yielding all the potential English variations of a named entity for further search, discovery, and analysis. For example "Mohamed" can also be spelled as "Mohamad" or "Mohamid" or "Mohammad" or in at least 20 other ways – all corresponding to a single Arabic spelling " محمد "

Sample taxonomy subcategories

- Place Name
  - Continents
    - Countries
      - Cities
        - Capitals
    - Governorates
      - Streets
- Person Name
  - Political Leader
    - President
      - US President
    - Minister

ANEE even enables you to set relationships between entities so it only extracts them when it sees them in relationship to each other -- you may not want every mention of "George W. Bush", but you do want everything related to President Bush and the Persian Gulf. It also enables you to extract only data related to one concept for a multiple-concept entity. Keyword search the date "6th of October" in Arabic and you will get a variety of results including anything with that general date as well as the Egyptian Armed Forces Day, an Egyptian industrial area, and a famous bridge in Cairo. With ANEE, select just the "date" category and it will uncover only information about the Egyptian Armed Forces Day. Unlike a conventional search engine, ANEE actually analyzes the context in which "6th of October" appears and correctly distinguishes it from any other use.

ANEE is the ideal tool for mining vast amounts of data for meaningful information – quickly, easily, and efficiently. It is available as a system development kit (SDK) for integration into an existing application, regardless of platform – Windows, Macintosh, or Linux – including use as a browser toolbar application to quickly analyze web pages for entities. ANEE also can be used as a stand-alone application.

COLTEC