

Significs Taming the BIG DATA

Named Entities Recognition Technologies

(Human names recognition – quick samples)

Introduction...

The science of Named Entities Recognition [NER] is around the ability to locate and extract predefined categories from free textual inputs and classify them into their appropriate categories –e.g. person names, companies/organization names, commercial brands ...etc.

اشترى أحمد عبد السلام عطية قطعة أرض في الفيوم.
أنتجت شركة أبل جهاز جديد.

With the enormous amounts of textual data existing in information technology databases, internet web portals, newspaper articles ...etc. It became a mandatory requirement to let the computers be able to identify such entities automatically.

As part of Info Arab's **Significs**® technologies, and based on its world-wide recognized Natural Language Processing [NLP] components, Info Arab is proudly introducing its NER capabilities as follows...

Human Names Recognition...

Due to the Arabic language special nature when handling the human names, and as part of **Significs**® continuous research and development to understand human written text chunks, a special version of Named Entities Recognition have been developed for human names.

Significs® NER component is capable to scan any stream of text and identify the mentioning of human names within the text. Through the linguistic semantic analytics features, and the world-class Application Programming Interface [API], the NER technology user will be able to...

1) Extract human names of 1 to N length...

قام أحمد ومحمود بزيارة علي حسن ومحمد حسام البتانوني.
قام أحمد ومحمود بزيارة علي حسن ومحمد حسام البتانوني.
قام أحمد ومحمود بزيارة علي حسن ومحمد حسام البتانوني.

In the above example, the first line was to identify entities of single names. The second line was only compound human names of two or more. While the third was for minimum of three human names.

2) Identify person Gender...

وتضم القائمة سارة وحسين وعصمت.
قامت عصمت بزيارة علياء.

Info Arab®, Arabic Information Systems, Alaa S. Al-Agamawi & Co, 7 Mosadek Street, Dokki, Cairo 12312, Egypt.
Phone: +2 02 37 61 38 33 - Fax: +2 02 37 61 38 39 x115. email: info@info-arab.com
www.info-arab.com, www.info-arab-technologies.com, www.semantic-significs.com

In the previous example, clear Male and Female names are recognized directly. While neutral names require the existence of some grammar rule to be able to identify them.

3) Handle common morphing rules...

خالد كهشام.
أعطي، ماخذ لتامر الكتاب.

In Arabic language, it is quite common that prefixing characters may be connected to human names.

4) Identify relations...

مشيرة وحازم عباس الصاوي.
حازم عباس الصاوي وشقيقته مشيرة.

Significs® NER is capable to identify the full name of both objects through the existence of any logical relation identifier.

5) Common spelling mistakes...

اشترى أحمد أبو الليل من عبد العزيز العقاد.
ثم قام احمد ببيعها إلى عبدالعزيز مرة أخرى.

In the above example, **Significs**® NER is capable to identify that both Ahmed & Abdel Azeez are the same person regardless of the changes in their writing style.

6) Various writing styles...

ذهبت مروءة مع شقيقتها سناء إلى الحديقة، حيث لعبت مروى مع صديقاتها وجلست ثناء تقرأ.

The same concept is applied when dealing with various writing styles for names due to phonological interpretation.

7) Foreign names recognition...

مايكل دوغلاس ممثل شهير يتبارى مع كريستين دانست في الفيلم، إلا إن دانست تفوقت على دوغلاس في مشهد

Significs® NER has a special module to handle the different writing styles of foreign names in Arabic.

8) Transliteration (i.e. Franco Arab)...

Ana ro7t ma3 mahmoud we Laila el nadi we kalemna Nevien
3alashan teegi.

As with all **Significs**® modules, **Significs**® NER can handle names written in Latin characters as well as handling Egyptian slang text streams.

Other modules...

In continuation to its none-stop efforts to realize its mission and aim, Info Arab is sparing no effort nor money in the research and development of other specified NER components.