Automatic Alias Recognition System for Both English and Arabic

Technology #013-037-diab

With the proliferation of web based data and social media, the number of people using aliases or fake names has expanded enormously. Aliases for famous people are often commonly known (for instance: “George Walker Bush” and “Bush Jr.”), but the set of alternate names that an ordinary person might use in online communication is generally not known or easily discovered.

Present systems exist which can retrieve all the information about a person referred to by various aliases, but those systems rely on an input list of names and aliases which must be already known. The newly invented system is capable of examining a dataset (such as a set of twitter or blog posts) and automatically recognizing when two different names represent the same person. Implementing this system can improve information retrieval, for instance, a query about Mahmoud Abbas should retrieve all the information mentioning ‘Mahmoud Abbas’, ‘Abou Mazen’, or ‘President of PA’. The system can also improve information extraction: it can discover aliases previously unknown to the user, for instance, identifying alternate names for the same entity, medical drug, or person name.

The system is language independent; in addition to working in English, it is the first alias recognition system available for the Arabic language.

Applications:
- Intelligence, Security, Law-enforcement
- Web Search
- Social Media

Advantages:
- Can automatically discover previously unknown aliases
- Language independent; has been tested in English and Arabic
Inventors

Mona Diab

Ayah Zirikly