EMM OSINT Suite Release I

Description

The EMM OSINT Suite is a desktop software package which consists of various tools based on the JRC's research in open source text analysis and mining.

The software consists of the following core modules:

Data Acquisition

- Search a component to extract search results from online search engines
- Crawler a HTTP crawler module to harvest data from targeted web sites ("crawling")
- Grabber a HTTP client module to download text based or binary documents from web sites for further processing

Data Processing

- Text Extraction extracts texts from different text based and binary formats (XML, TXT, PDF, MS Word, MS Excel, MS PowerPoint, Open Office)
- Entity Extraction a set of modules to extract named entities from raw text. Entity types are people, organisations, locations, address information, VAT numbers and user defined custom types
- Category Matching categorises text according to key word based category definitions
- Translation option to integrate on-demand translation system

Data Analysis

- Reporting a component to create reports for end users of for further external processing of extraction results
- Local Search a local search index to provide full text search of downloaded artefacts
- Entity Browser an analysis component to aggregate found entity data and allows browsing through the results.

User Interface

- Graphical User Interface based on open source
- Online and offline help system

Release Notes

The 2016 release contains the following improvements and bug fixes:

Improvements

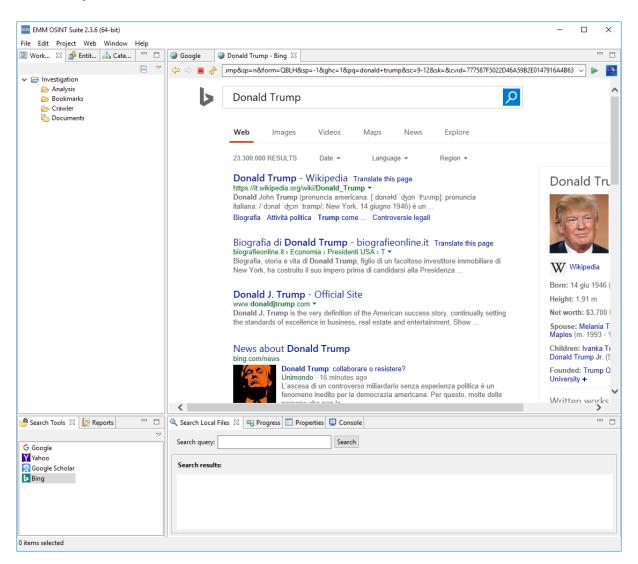
- Data Acquisition
 - New search engine adapters
 - Improved web crawler
- Data Processing
 - Custom entities design view
 - Streamlined handling of custom entities
 - Allow integration of on-demand translation system
- Data Analysis
 - Filtering of entity types
 - Performance improvements and optimised memory utilisation
 - Entity graph new layout types
- System Basis
 - Faster start-up time
 - Latest Java Run Time

Bug Fixes:

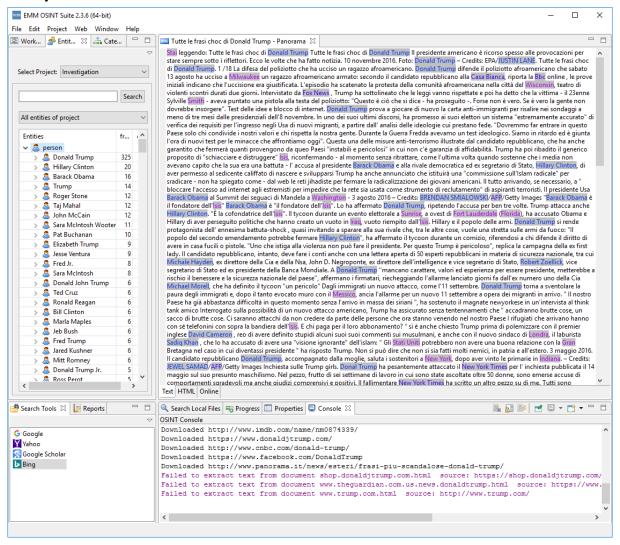
- #OSINT-118 Export Bookmarks to TSV creation date wrong or missing
- #OSINT-97 Show active workspace path and active name variant DB in title bar
- #OSINT-103 Allow custom entity patterns to match against page HTML
- #OSINT-92 EntityBrowser view is not correctly refreshed if selected project is deleted

Screenshots

Data Acquisition



Analysis



User Report

