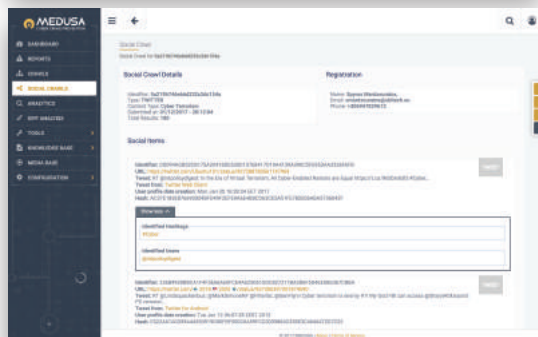
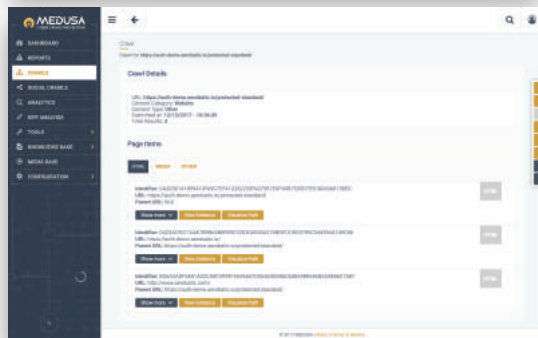
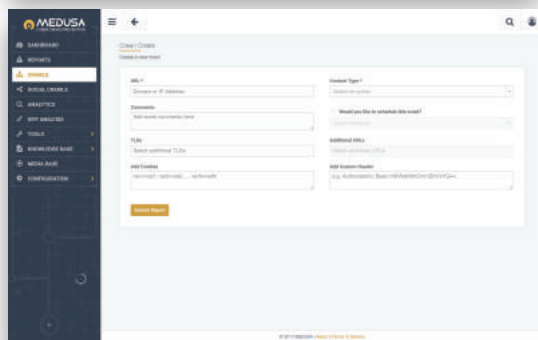
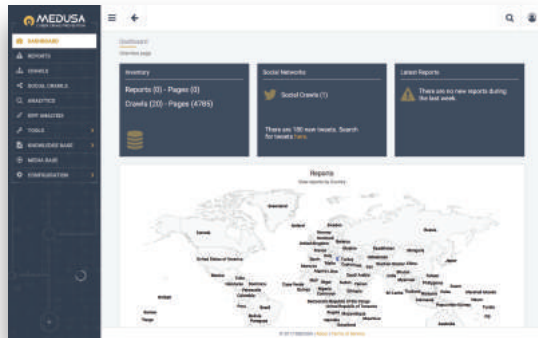


The MEDUSA Cyber Intelligence suite constitutes a sophisticated, modular, highly-configurable and -scalable Web mining and intelligence platform that benefits from Artificial Intelligence and Big Data technologies so as to provide intelligence and real-time insights to non-IT domain experts, satisfying the multi-disciplinary needs of end-user organizations that require advanced Web crawling, processing and analytics services.

Scalable Web Crawling

- highly configurable crawling engine that facilitates full crawl control, configuring one or more seed URL(s), the overall crawl size and depth, the location of servers, the URLs patterns, as well as, the targeted content type and language
- automated expansion of crawling in all variations of the seed URLs using other available “Top Level Domains” (TLDs)
- automated expansion of crawling in all transformations of the seed URLs that derive using hacking alphabets, such as Leet (1337)
- configuration of the politeness (aggressiveness) policy of crawling, so as to remain stealth and avoid detection
- configuration of the revisit policy for each target website so as to capture its dynamic nature (including any creations, updates or deletions), enabling the continuous monitoring of the target
- utilization of custom headers and/or cookies during crawling to impersonate real users or agents
- anonymous crawling of Dark Web via the Tor network, with the transparent usage of a totally integrated Tor proxy
- capturing and fetching all objects and requests of the crawled website, including HTML, XML, CSS, JavaScript, binaries, images, videos
- supporting logged-in, authenticated mode for crawling websites, marketplaces and forums, with the use of the integrated scripting engine for authenticating in specific free and open source forums, such as myBB, phpBB, Simple Machine Forums, etc.
- support for crawling social networks that have APIs (e.g. Twitter), capturing rich data including users’ profiles, posts and multimedia content extracting entities such as posts’ hashtags and URLs
- support for parallel crawling of thousands of target websites



Multi-level Processing Pipeline

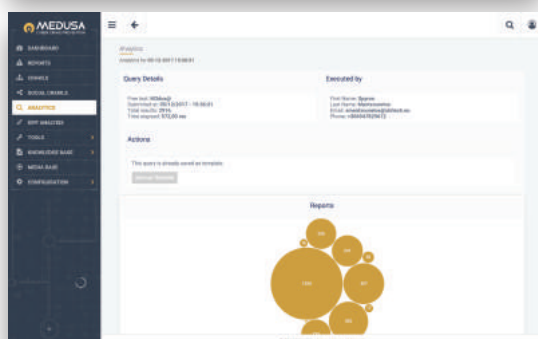
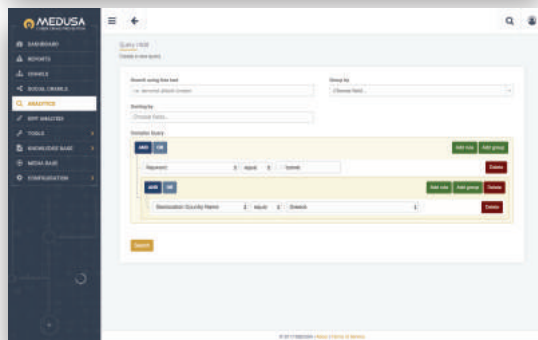
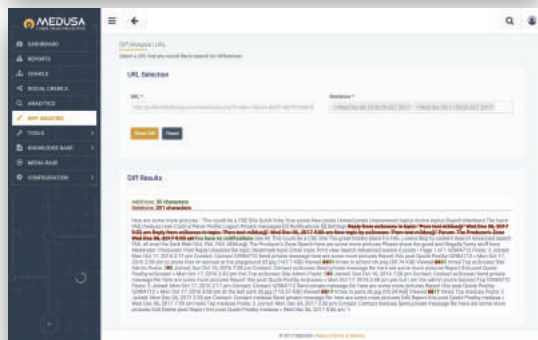
- complete HTML code and text extraction from any webpage
- metadata extraction (like EXIF tag structure) from a range of binary resources and file formats like PDF documents, image files, sound files, office documents, and many others
- face detection for identifying human faces in digital images (and videos) utilizing pre-trained models
- face recognition against a predefined set of “known” human faces, using a totally integrated deep neural network for enabling clustering, similarity detection and classification tasks
- real-time multiple objects detection and recognition in digital images and videos, adopting neural networks technologies
- nudity classification and detection of offensive / adult images
- keyword and regular expression (REGEX) spotting, for extracting email addresses, telephone numbers, IP addresses, Bitcoin addresses, named geo locations and customizable dictionaries of risk terms
- automated auto-generation of knowledge graphs, representing relations and interactions among users, for specific forums (e.g. myBB, phpBB and simple machines forums)

Evidence Collection

- automatic capturing of Document Object Model (DOM) browser during crawling
- electronic sealing and timestamping of the captured artefacts from the target website
- offline browsing of the already crawled websites

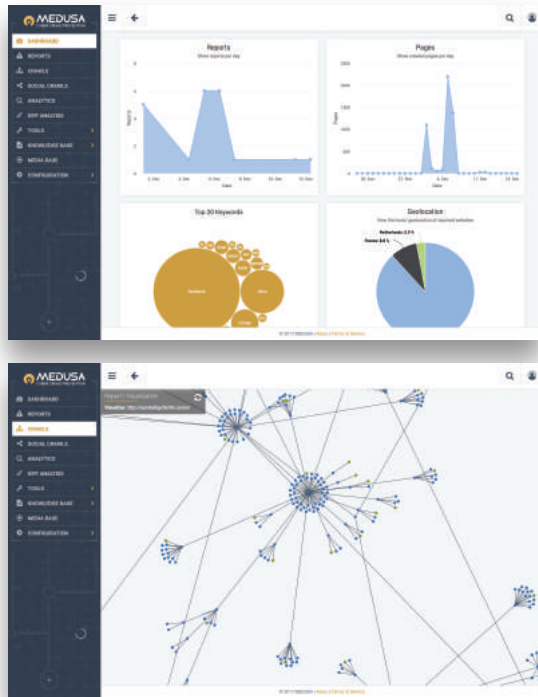
Intuitive Analytics

- semantically-enriched indexing, faceting and categorization of all data fetched from the crawled websites, allowing free-text search, keyword search, entity classification/correlation -based search, phrase search, complex search, geospatial search, term boosting, spell correction, auto-completion, etc.
- automated query expansion, using an unsupervised neural network model that identifies words that occur in similar contexts and/or are also similar in meaning, enabling the natural representation of analogies with "human-like" semantic awareness
- graphical query designer that allows the creation of complex queries in an easy, user-friendly way
- supporting the creation, reuse and extension of query templates, for improving the efficiency and effectiveness of complex and/or repetitive operations
- automated real-time diff analysis, spotting additions, modifications and deletions, among to consecutive visits (crawls) of the same target website



Built-In AI Support and Trained Models

- incorporating a neural network for concept extraction, allowing multiple objects detection and recognition, in real-time, in digital images and videos,
- incorporating a neural network for face detection, in digital images and videos, that supports facial landmark detection, head pose estimation, facial action unit recognition, facial features extraction and eye-gaze estimation
- incorporating a convolutional neural network for nudity assessment, automatically identifying that an image is not suitable/safe for work (NSFW) – including offensive and adult images
- ability to train the aforementioned models with a custom media base, fostering the face and object recognition capabilities of the suite
- ability to define custom keywords and regex expressions for enriching the knowledge and information extraction capabilities from the crawled text



Integrated Network Tools

- ping, checking host connectivity and reporting packet loss and latency
- whois, finding out who owns the domain, when that domain expires, to view the configured logs, contact details, etc.
- dig, querying Domain Name System (DNS) servers
- traceroute, displaying the route (path) and measuring transit delays of packets across an IP network
- mmap, scanning networks for determining which hosts are alive in a network
- nslookup, querying a DNS server for DNS data
- reverse lookup, providing the domain name associated with a particular IP address (reverse DNS lookup)

Multi-modal Alerting

- ability to configure triggering events, e.g. start/stop of crawling task, detection of new object/face/person and new keyword detection
- push notifications through email and SMS service

Reporting

- filtering crawling results based on content type, media classification, geolocation, related cases, etc.
- visual representation of crawling results (spider diagrams)
- reconstructing social graphs and user activity for specific forums (e.g. myBB, phpBB and simple machines forums)


Interoperability with Third-party Tools


- exposure of a sound application programming interface (API) to submit crawling requests
- integration with third-party legacy systems adopting the publish/subscribe (pub/sub) pattern
- exportable results in multiple, structured format (XML, JSON, CSV, binary)


CONTACT

 Thessalias 8 & Etolias 10
15231 Chalandri, Athens, Greece

 26 N.Pattich Str
3071 Limassol, Cyprus

 Paraná 754, Piso 11B
C1017 Buenos Aires, Argentina

 Mr. Dimitris Alexandrou
P: +30 6949193035
E: dalexandrou@ubitech.eu

 Mr. Thanassis Bouras
P: +30 6949193034
E: bouras@ubitech.eu