J. Def. & Sec. Technol., vol. 1, no. 1, p0-pn, 2017
https://doi.org/10.xxxxx/jdst.vvnpn

www.jdst.eu, ISSN XXXX-YYYY, e-ISSN XXXX-YYYY

Research Article

Tools for Discovering and Analyzing Web Resources Containing Homemade Explosive Information

G. Kalpakis, C. Iliou, T. Tsikrika, S. Symeonidis, S. Vrochidis, I. Kompatsiaris

*Information Technology Institute, CERTH, 6th km Charilaou-Thermi Road, GR-57001,
Thermi- Thessaloniki, Greece
http://mklab.iti.gr/*

ABSTRACT

This work presents a novel application for discovering Web resources containing recipes for manufacturing Home Made Explosives (HMEs), analyzing multimedia content for determining its relevance to the HME domain and detecting the HME-related objects it contains. The discovery of HME Web resources both on the Surface and the Dark Web is based on a hybrid infrastructure that combines two different approaches: (i) a Web crawler focused on the HME domain; (ii) the submission of HME domain-specific queries to general-purpose search engines. Both approaches are accompanied by a post-processing classification for reducing the potential noise by re-ranking the discovery results. The multimedia analysis detects HME-related semantic concepts in visual content and determines whether it is relevant to the HME domain. The proposed tools are developed in a user-driven manner based on the needs of law enforcement and security agency personnel, as well as HME domain experts. Overall, the ultimate goal of this application is to provide law enforcement agencies with additional operational means in their fight to keep the citizen safe.

✉ George Kalpakis: Tel.: +30 2311257807 Fax: +30 2310474128; E-mail:kalpakis@iti.gr

ARTICLE INFO

RECEIVED: 09 Oct 20xx
REVISED: 10 Nov 20xx
ACCEPTED: 30 Nov 20xx
ONLINE: 12 DEC 20XX

KEYWORDS

concept detection, Dark Web, domain-specific, focused crawling, homemade explosives.

I. INTRODUCTION

The escalation of terrorist attacks during the past years clearly indicates that law enforcement agencies (LEAs) need to adopt effective mechanisms for preventing, investigating, and suppressing terrorist activity worldwide not only by following conventional methods, but also by considering alternative solutions provided by the rapid technological progress. At the same time, the growth of the Internet and broadband services has facilitated the communication and the diffusion of knowledge among users worldwide and has thus proven useful for terrorists, since it constitutes a means used for supporting their acts either by communicating their views, and recruiting future members of terrorist groups, or by sharing material with plausible subversive use, including information for synthesizing homemade explosives (HMEs).

The recent advancements in Web search support LEAs in their effort to discover information related to terrorist activity. Most efforts have focused on the so-called *Surface Web* that constitutes the part of the Web indexed by conventional general-purpose search engines. However, such search engines are capable of indexing just a small portion of the available Web information; the rest lies in the so-called Deep Web which in general comprises content that cannot be detected by the crawlers employed by conventional search engines. There is also a part of the *Deep Web*, known as the *Dark Web*, formed by several darknets (e.g. Tor[1], I2P[2], and Freenet[3]), which provides anonymity both from a user and a data perspective and can only be reached if the appropriate setup of software, configuration or authorization is used [1]. For this reason, the Dark Web has become popular for sharing illegal material and disseminating extremist and terrorist-related information.

The challenge for law enforcement is to better understand and counter the ability for subversive use of the information related to the construction of HMEs being shared online both on the Surface and the Dark Web. The proliferation of information provides the subversive with the ability to easily research and get familiar with the manufacture of HMEs using everyday household goods, and easy to purchase materials. Therefore, LEAs need technologies that would allow them

[1] https://www.torproject.org/

[2] https://geti2p.net/en/

[3] https://freenetproject.org/

to tackle this threat through the automatic discovery and analysis of such HME information, so as to enable the interested stakeholders to access it in an effective and efficient manner.

To meet this challenge, this work proposes a novel application for the discovery and analysis of multimedia HME-related content shared on the Surface and the Dark Web. Such information can be found on various types of Web resources, such as Web pages, blogs, forums, social media posts that may express their content in in various modalities, such as text, images, and video. The proposed application is being developed in the context of the activities of the *HOMER*[4] (*HOmeMade Explosives and Recipes characterization*) EU FP7 project and enables the discovery and further analysis of Web resources containing information about HMEs through a user-friendly interface. It consists of three major components that provide several advanced functionalities:

1. an **HME Discovery** component that enables the identification of Web resources containing HME information based on a hybrid infrastructure which combines two different approaches: (i) a Web crawler focused on the HME domain, and (ii) the submission of HME domain-specific queries to general-purpose search engines; the discovered resources are re-ranked based on a classification process,
2. an **HME Multimedia Analysis** component that detects HME-related concepts on multimedia content and determines whether the visual content is relevant to the HME domain, and
3. a **Graphical User Interface** which facilitates the user communication with the application's main functionalities.

The main contribution of this work is the integration of advanced state-of-the-art technologies in a novel application for the discovery and analysis of heterogeneous Web resources containing multimedia HME information, developed in close collaboration with law enforcement agencies and domain experts. Additionally, this application provides a combination of domain-specific tools applied both on the Surface and the Dark Web.

The remainder of this paper is structured as follows: Section II reviews related work. Section III discusses the use cases provided by law enforcement agents. Section IV presents the architecture of the application, while Section V describes its major components. Section VI describes the core interaction modes of the application. Finally, Section VII concludes this work.

II. Related Work

Several research efforts towards the discovery, study and analysis of extremist and terrorist-related Web content have been conducted over the past years. The most comprehensive suite of text and Web mining tools for performing link, content, and video analysis has been developed in the context of the Dark Web

[4] http://www.homer-project.eu/

project at the University of Arizona [2]. However, this project deals with the whole breadth of terrorist content, rather than solely with HME information, as done here. Close to this work is also the framework developed in the context of the EU FP7 project CAPER [5] (Collaborative information, Acquisition, Processing, Exploitation and Reporting for the prevention of organized crime). Its goal is to build an information sharing platform for the detection and prevention of organized crime. Again, they deal with a wider scope of information, rather than only HMEs, emphasizing the analysis of social networks [3]. Our work thus proposes a set of novel tools that focus on the discovery and analysis of multimedia HME information.

III. Use Cases

This section discusses the use cases for the HME discovery and multimedia analysis, provided by law enforcement and security agents in the context of HOMER project who have offered guidance for a user-oriented development.

Use Case 1. Identify HME-related content in terrorist-related discussions: LEAs need to detect user posts in Web pages hosted on the Surface and the Dark Web containing discussions about the process of manufacturing HMEs, and providing instructions related to the use of such products for organizing terrorist operations. For example, consider the case where the members of an extremist/terrorist group, discuss their intention of synthesizing an HME and using it in a future terrorist attack.

Use Case 2. Search the Web for content related to an explosive: LEAs need to discover new sources containing information about the manufacture of HMEs by submitting keyword-based queries enhanced with a mechanism for their automatic formulation. They, also, need to get results ranked by their relevance to the HME domain. Such content is usually published online and indexed by general-purpose search engines of the Surface and/or the Dark Web. For instance, consider there is intelligence that a specific substance not previously used is currently being considered for subversive use.

Use Case 3. Perform multimedia analysis on the discovered Web resources: LEAs are interested in automatically analyzing the multimedia content hosted on the Web and in particular on multimedia sharing platforms, such as YouTube[6]. They aim at identifying videos with HME content among all available information, and at detecting specific objects related to the manufacture of HMEs within such videos for extracting useful information. For instance, the LEAs may obtain useful evidence for future investigations by exploiting the automatic identification of videos showing the particular chemical substances being used for the manufacture of an HME.

[5] http://www.fp7-caper.eu

[6] https://www.youtube.com/

IV. Application Architecture

This section provides a high level overview of our application for the discovery of HME Web resources and the analysis of the discovered multimedia content, which is designed based on the use cases described earlier. As shown in Figure 1, the application enables the users (i.e. police and anti-terrorism/intelligence officers) to perform customized searches both on the Surface and the Dark Web for discovering HME-related content, as well as to analyze the discovered multimedia content available on multimedia sharing platforms with the goal to identify HME-related objects and associate the visual content with the HME domain. The discovery component takes advantage of a hybrid model based both on *focused crawlers* and *general-purpose search engines*. The latter relies on an automatic query formulation approach based on *query patterns*. The focused crawler instead starts from a predefined set of seed Web resources and traverses the Web link structure with the goal to identify Web resources having HME-relevant content; the results of the discovery process are re-ranked based on their relevance to the HME domain through a classification process. On the other hand, the multimedia analysis component exploits the low-level visual features of the multimedia content discovered (i.e. images and videos) and aims to associate them with higher-level HME-related semantic *concepts* (i.e. objects) in order to determine whether the visual content is relevant to the HME domain and identify the HME-related objects it contains. The various functionalities of our application are accessible through a user-friendly and intuitive *Graphical User Interface (GUI).*

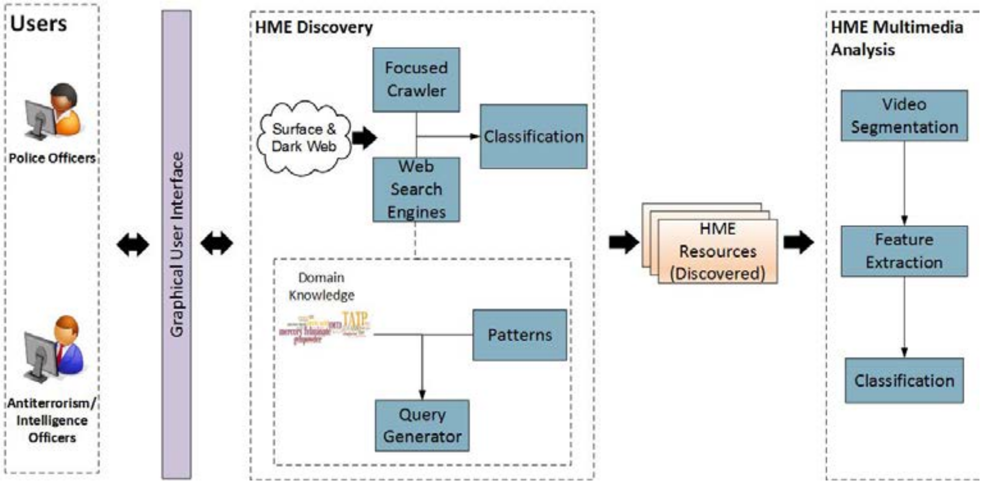


Figure 1: Discovery and Analysis Application architecture

V. Application Components

This section provides a detailed overview of the major components of our application for the discovery of HME information and the analysis of the multimedia content discovered.

A. Discovery Component

The discovery component of our application is based on hybrid architecture which combines a focused Web crawler module and a general-purpose Web search engine querying module, both followed by post-processing classification. Next, the functionality of the aforementioned modules is discussed in detail.

Focused Crawler

The focused Web crawler of our application is capable of traversing both the Surface Web and several darknets of the Dark Web (i.e. Tor, I2P, and Freenet) with the goal of discovering Web resources related to the HME domain. It is based on a customized version of Apache Nutch[7] (version 1.9).

The **focused crawler** gathers HME-related content by selecting to follow only the hyperlinks that lead to relevant resources. Specifically, the focused crawler starts from a set of predefined (seed) Web pages relevant to the HME domain, fetches (i.e. downloads) these pages, parses their content for extracting the hyperlinks they contain, and selects the ones that possibly point to other Web pages relevant to the HME domain. This process is iteratively repeated until a termination criterion is applied (e.g. a desired number of pages are fetched).

For predicting the benefit of fetching an unvisited Web page, our focused crawler takes advantage of the "*topical locality*" phenomenon observed on the Web, which dictates that most Web pages link to other pages with similar content. To this end, our focused crawler adopts a classifier-guided crawling strategy using supervised machine learning based on *Support Vector Machines (SVMs)*, for classifying the hyperlinks based on (i) their local context, i.e. the textual content appearing in the vicinity of the hyperlink in the parent page, such as its anchor text and its surrounding text, and/or (ii) global evidence associated with the entire parent page, such as its textual content and its hyperlink structure [4], [5]. The decision whether a hyperlink should be followed or not depends on the confidence score produced by the classifier.

Web Search Engine Querying Module

The **querying module** takes advantage of the indexes and search infrastructures of existing general-purpose search engines and submits domain-specific queries to them, which aim at discovering HME-related content on the Surface and the Dark Web. It supports submitting queries in several search engines both from the Surface and the Dark Web (i.e. Yahoo![8], Bing[9], DuckDuckGo[10], Ahmia[11], and

[7] http://nutch.apache.org/
[8] http://www.yahoo.com/
[9] https://www.bing.com/
[10] https://duckduckgo.com/

Torch[12]), and it also provides the option to search for HME-related content in Google Groups[13]. Regarding the Yahoo! and Bing search engines, the querying module takes advantage of the available APIs (i.e. Yahoo! BOSS API[14], and Bing Search API[15] respectively), whereas for DuckDuckGo, Ahmia, and Torch (where no API is available) the querying module submits a query on their Web interface and parses the returned results.

The automatic domain-specific query generation requires the availability of an initial set of seed queries that can be processed so as to mine abstract query patterns that can then instantiated into multiple query instances corresponding to sequences of domain-specific keywords [6], [7]. Once the keywords appearing in all queries are mapped to discrete concepts, then in every query in the initial seed set, the keywords are replaced by the respective concepts they are mapped to; for example, the query "*preparation anfo*" becomes "*action explosive*". This results in producing a set of discrete patterns that can be used for automatic query generation. For example, the pattern "*action explosive*", where "action" corresponds to keywords such as "*how to make*", "*preparation*", etc., may be instantiated to several different queries for each of the explosives of interest. Hence, once a user wishes to discover HME Web resources about a given explosive, they can select query patterns containing the concept "*explosive*", and these patterns will be automatically instantiated for that particular explosive, and will be submitted in parallel to the search engines of interest.

Post-processing classification module

The resources discovered through the focused crawling and the Web search engine querying are then classified based on their textual content. A text-based **classifier** is trained on a set of Web resources annotated as relevant or non-relevant to the HME domain. The training process includes parsing each resource, extracting its textual content, and performing textual processing (i.e. tokenization, stopwords removal and stemming), in order to define a set of features which will most efficiently and meaningfully represent the information that is important for analysis and classification. Then, the classifier is trained through a supervised procedure based on SVMs. The classification process is optional and can be initiated by the user on the results returned by a focused crawl or a Web search engine querying. Each returned Web resource acquires a confidence score that illustrates its relevance to the HME domain. The returned results are then re-ranked in descending order based on the confidence score produced. The

[11] https://ahmia.fi/

[12] http://xmh57jrzrnw6insl.onion/

[13] https://groups.google.com/

[14] https://developer.yahoo.com/boss/search/

[15] http://www.bing.com/toolbox/bingsearchapi

classifier is implemented using the libraries of the Weka[16] machine learning software.

B. Multimedia Analysis Component

The multimedia analysis component aims to associate the low-level visual features of visual content (i.e. images and videos) discovered with high-level HME-related semantic concepts, with the ultimate goal to detect HME-related objects and determine the relevance of the visual content to the HME domain. To this end, a state-of-the-art **multimedia concept detection** component for the automatic classification of visual content to a set of predefined semantic concepts (i.e. objects) is employed and consists of the following sequential steps [8]:

1. **video decoding** for extracting the most representative frames from videos for further processing,
2. **feature extraction** and representation for extracting a set of descriptors that effectively characterize the visual content, and
3. **classification** for training and subsequently applying the trained models so as to classify the multimedia content to a set of predefined HME-related semantic concepts.

Given the lack of available concept lexicons for the HME domain, a set of HME-related concepts was determined based on end-user requirements. This resulted in a list of HME concepts, including concepts related to the appearance of HME material (e.g. "*powder*", "*granules*", etc.), equipment used for the synthesis of HMEs (e.g. "*glassware*"), explosions resulting from using HMEs (e.g. "*fire*", "*smoke*", etc.), improvised explosive devices containing HMEs (e.g. "*device*"), as well as a generic concept which determines whether the visual content is related to the HME domain. For each of these concepts, a classifier has been trained based on machine learning techniques (i.e. logistic regression) and is available for annotating unlabelled multimedia content so as to support its retrieval.

C. Graphical User Interface

The two aforementioned components of our application, namely the discovery and the multimedia analysis components, are accessible via a Web-based **Graphical User Interface (GUI)** with a minimalist design aiming to provide a usable, flexible and consistent environment for the user.

The GUI provides a parametrized environment for running each one of provided facilities. Specifically, when using the focused crawling infrastructure of the discovery tool, the following parameters should be configured (see Figure 2): (i) the set of seed URLs constituting the starting points of the crawl (it may contain URLs representing Web pages from the Surface Web, the Tor network, the I2P network, and the Freenet), (ii) the crawler classifier's threshold (i.e. for every hyperlink encountered, the classifier produces a relevance score; the crawler will

[16] http://www.cs.waikato.ac.nz/ml/weka/

try to fetch each hyperlink having a relevance score greater than the threshold set), (iii) the crawl depth (i.e. the maximum distance allowed between the seed pages and the crawled pages), and (iv) the option to perform a domain-restricted crawl (i.e. the crawler is allowed to follow hyperlinks belonging only to the same domain name(s) of the URL(s) present in the seed URL list).

Additionally, the interface for the search engine querying tool (see Figure 4) provides the option to select one or more of the supported search engines both from the Surface and the Dark Web (i.e. Yahoo!, Bing, DuckDuckGo, Ahmia, and Torch), whereas it also supports searching for HME content in Google Groups. Two interaction modes are provided: (i) a free text mode for manually submitting queries, and (ii) an automatic mode for submitting queries taking advantage of 18 explosive-related or 16 ingredient-related query patterns (the pattern category selection is guided through a tabbed environment). Finally, an option of accessing the search history containing all the previous queries run along with the respective cached results is provided.

In both cases, a crawl or a querying run may be followed by a user-initiated post-processing classification which re-ranks the respective returned results based on their relevance to the HME domain.

Finally, the interface for supporting the multimedia analysis tool provides the option of submitting keyword-based queries to the Youtube API[17] for discovering relevant videos posted on the Youtube multimedia sharing platform and analyzing with the goal to determine their relevance to the HME domain and identify HME-related objects. The interface provides the option to parametrize the multimedia analysis score threshold which defines (i) whether a video file is considered as relevant or non-relevant to the HME domain (i.e. bigger threshold values entail applying a stricter policy on the detection of the HME concepts and/or considering the video file as relevant to the HME domain), and (ii) whether the HME-related concepts (i.e. objects) are detected on the video content.

VI. Modes of Interaction

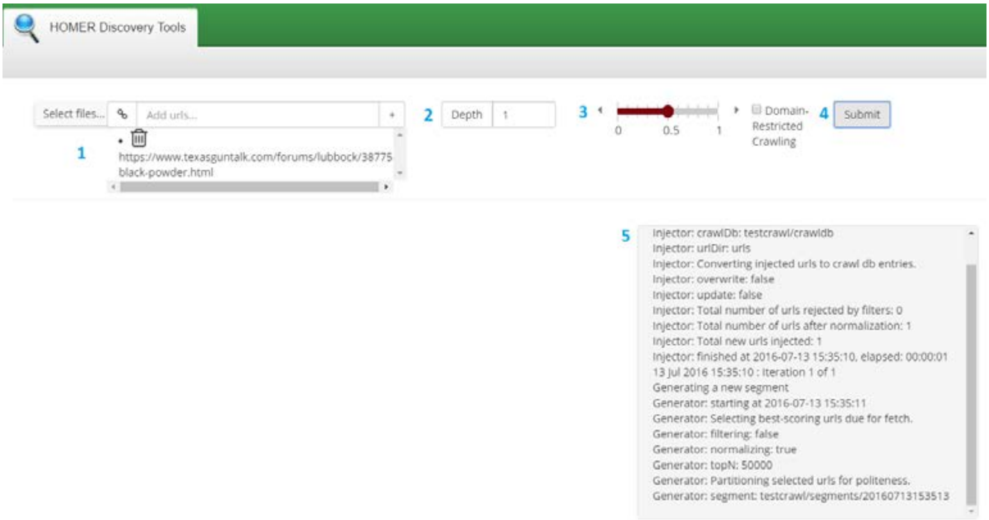
This section presents the modes of interaction when using our tools for performing the tasks described in the use cases discussed in Section II.

Interaction Mode 1. Search the Web for discussions about HMEs. In this case, the goal is to perform a crawl so as to discover posts in Web forums and discussion groups related to synthesizing HMEs. Figure 2 illustrates the focused crawling interface. For performing a crawl, the user should provide the seed URL(s) (in this case, a forum[18] Web page representing a section including topics related to HMEs) (1), select the crawl depth (it is set to 1) (2), set the crawler classifier's threshold (it is set to 0.5) (3), and initiate the crawl process (4). For the

[17] https://developers.google.com/youtube/v3/

[18] https://www.texasguntalk.com/forums/lubbock/38775-black-powder.html

whole duration of a crawl, a window providing continuous feedback messages regarding the crawling process appears on screen (5).



Depth 1

https://www.texasguntalk.com/forums/lubbock/38775 black-powder.html

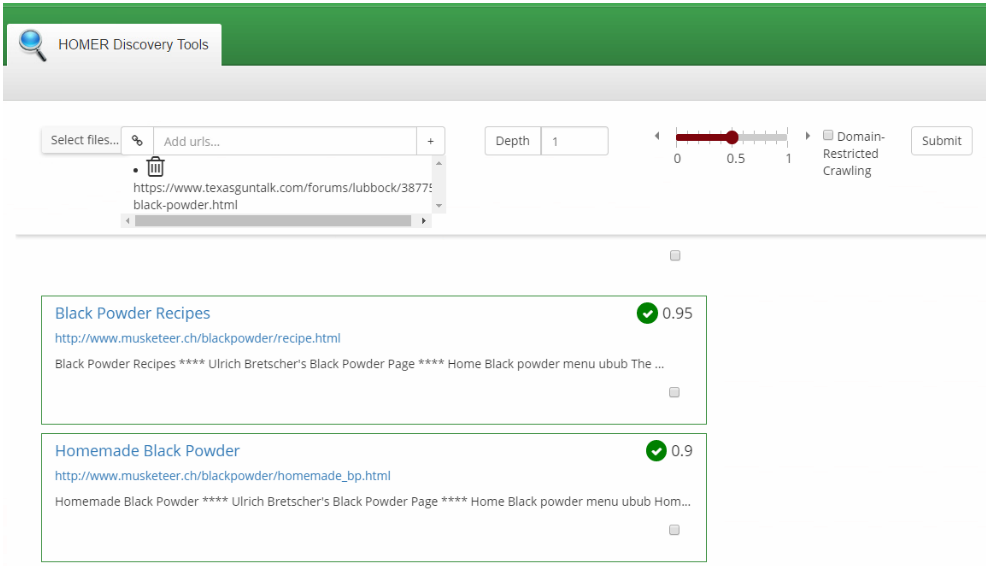
```
Injector: crawlDb: testcrawl/crawldb
Injector: urlDir: urls
Injector: Converting injected urls to crawl db entries.
Injector: overwrite: false
Injector: update: false
Injector: Total number of urls rejected by filters: 0
Injector: Total number of urls after normalization: 1
Injector: Total new urls injected: 1
Injector: finished at 2016-07-13 15:35:10, elapsed: 00:00:01
13 Jul 2016 15:35:10 : Iteration 1 of 1
Generating a new segment
Generator: starting at 2016-07-13 15:35:11
Generator: Selecting best-scoring urls due for fetch.
Generator: filtering: false
Generator: normalizing: true
Generator: topN: 50000
Generator: Partitioning selected urls for politeness.
Generator: segment: testcrawl/segments/20160713153513
```

Figure 2: Focused crawling interface when running a crawl. The feedback screen appears on the right part of the picture

After running a crawl, the respective results are presented on the screen. Figure 3 illustrates the list of the results returned for the aforementioned setup. Each returned result contains (i) the Web page title, (ii) the Web page URL, (iii) the Web page snippet (i.e. a short summary of the Web page based on its content and/or metadata tags), and (iv) the confidence score returned by the crawler classifier. A result is depicted in a colored box depending on whether the Web page it represents has been previously annotated by domain experts. There are two color variations: (a) green depicts relevant results and (b) red depicts non-relevant results.

Interaction Mode 2. Search the Web looking for an explosive. In this scenario, the intention is to discover Web resources containing information about HMEs by submitting pattern-based queries to general-purpose search engines. Figure 4 illustrates the querying facility interface. For performing a pattern-based querying on HMEs, the user should enter the desirable keyword(s) representing the explosive/ingredient of interest (in this case, the keyword provided is *ammonium nitrate*) (1), select either the Explosives or the Ingredients tab based on whether they wish to use the explosive-related or the ingredient-related query patterns respectively (in this case the Ingredients tab is selected) (2), choose the desirable patterns from the respective list (3), select the search engine(s) of their preference (Yahoo! and Bing search engines are selected)(4) and initiate the querying (5).

When the querying is completed, the results (after being merged) are presented to the user using the same format with the one used for the focused crawling results. Then, the user has the option to initiate the classification process. When the classification is completed, the results are re-ranked and presented along with their score of relevance to the HME domain. Figure 5 illustrates the re-ranked results after applying the classification process.



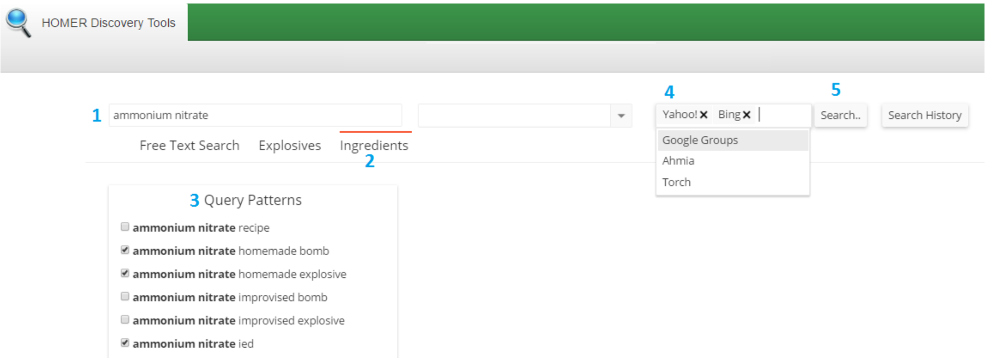
- https://www.texasguntalk.com/forums/lubbock/38775 black-powder.html

□

Black Powder Recipes ✓ 0.95
http://www.musketeer.ch/blackpowder/recipe.html
Black Powder Recipes **** Ulrich Bretscher's Black Powder Page **** Home Black powder menu ubub The ...
□

Homemade Black Powder ✓ 0.9
http://www.musketeer.ch/blackpowder/homemade_bp.html
Homemade Black Powder **** Ulrich Bretscher's Black Powder Page **** Home Black powder menu ubub Hom...
□

Figure 3: Results of a crawl run at depth=1 and threshold=0.5. The colored boxes indicate the relevance of the results to the HME domain.

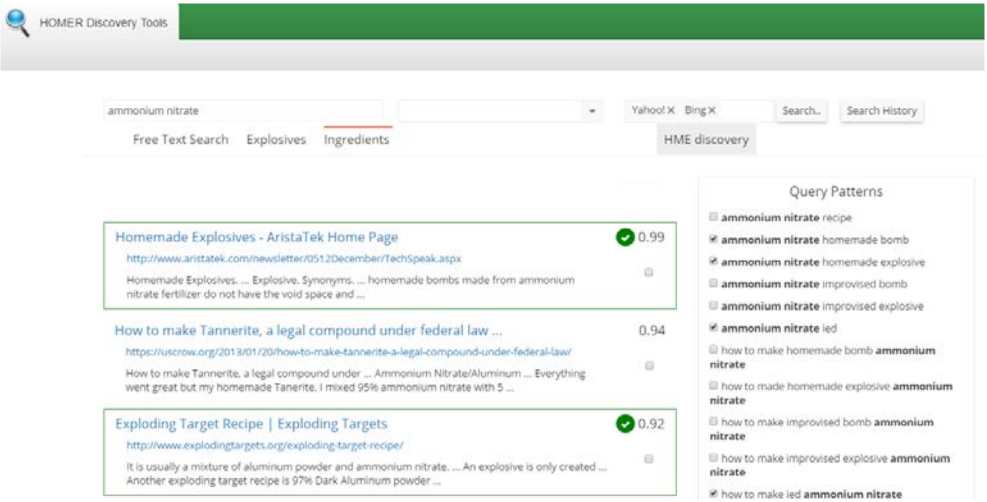


□ **ammonium nitrate** recipe
☑ **ammonium nitrate** homemade bomb
☑ **ammonium nitrate** homemade explosive
□ **ammonium nitrate** improvised bomb
□ **ammonium nitrate** improvised explosive
☑ **ammonium nitrate** ied

Figure 4: Querying interface when running a pattern-based search

Interaction Mode 3. Analyze multimedia content with possible HME content. In this scenario the goal is to analyze multimedia content (i.e. videos) hosted on the

Youtube platform for determining its relevance to the HME domain and identifying the HME-related objects contained. Figure 6 illustrates the multimedia analysis tool interface. The user should enter the desirable keyword(s) (in this case the keyword provided is *acetone peroxide*) (1), set the multimedia analysis score threshold (it is set to 0.5) (2), and initiate the multimedia analysis process (3). After running a search, a playable preview of the videos returned by the Youtube API appears on screen, along with their respective title, description and tags. Subsequently, the videos are automatically analyzed and the HME-related concepts detected (i.e. the ones having score higher than the threshold set) are presented on the screen, along with a sign indicating whether the video is relevant or non-relevant to the HME domain (i.e. a green tick sign appears below each video preview in case it is relevant, and a red x mark in case it is non-relevant to the HME domain).



ammonium nitrate

Homemade Explosives - AristaTek Home Page ✓ 0.99
http://www.aristatek.com/newsletter/0512December/TechSpeak.aspx
Homemade Explosives. ... Explosive. Synonyms. ... homemade bombs made from ammonium nitrate fertilizer do not have the void space and ... □

How to make Tannerite, a legal compound under federal law ... 0.94
https://uscrow.org/2013/01/20/how-to-make-tannerite-a-legal-compound-under-federal-law/
How to make Tannerite, a legal compound under ... Ammonium Nitrate/Aluminum ... Everything went great but my homemade Tanerite. I mixed 95% ammonium nitrate with 5 ... □

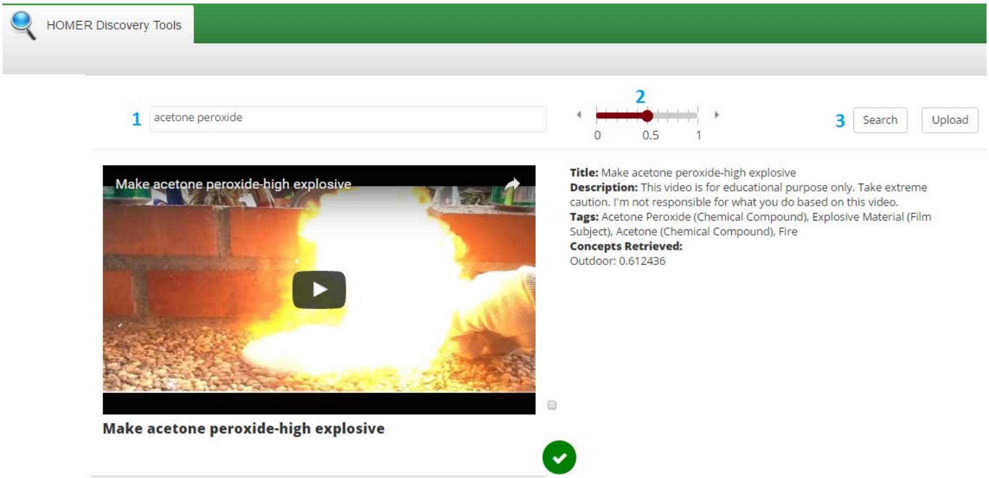
Exploding Target Recipe | Exploding Targets ✓ 0.92
http://www.explodingtargets.org/exploding-target-recipe/
It is usually a mixture of aluminum powder and ammonium nitrate. ... An explosive is only created ... Another exploding target recipe is 97% Dark Aluminum powder ... □

- □ **ammonium nitrate** recipe
- ☑ **ammonium nitrate** homemade bomb
- ☑ **ammonium nitrate** homemade explosive
- □ **ammonium nitrate** improvised bomb
- □ **ammonium nitrate** improvised explosive
- ☑ **ammonium nitrate** ied
- □ how to make homemade bomb **ammonium nitrate**
- □ how to made homemade explosive **ammonium nitrate**
- □ how to make improvised bomb **ammonium nitrate**
- □ how to make improvised explosive **ammonium nitrate**
- ☑ how to make ied **ammonium nitrate**

Figure 5: Querying results re-ranked after applying the classification process

VII. Conclusions

This work proposed an application that integrates several technologies for the discovery and analysis of multimedia Web resources containing HME information. It is envisaged that the tools developed will provide law enforcement and security agencies with additional means for tackling the threat from HMEs. The development of this application is currently ongoing and several further components will be integrated in the future, including a social media analysis component that will analyze the social media content in order to identify HME-related posts on social networking sites and detect user communities interested in the domain. Moreover, the application will be evaluated extensively in terms of its usability, effectiveness, and efficiency by law enforcement and security agency personnel, as well as by domain experts, in large-scale user studies that will take place in the context of the activities of the HOMER project.



Title: Make acetone peroxide-high explosive
Description: This video is for educational purpose only. Take extreme caution. I'm not responsible for what you do based on this video.
Tags: Acetone Peroxide (Chemical Compound), Explosive Material (Film Subject), Acetone (Chemical Compound), Fire
Concepts Retrieved:
Outdoor: 0.612436

Make acetone peroxide-high explosive

Figure 6: Multimedia analysis interface

Acknowledgements

This work was supported by the HOMER (312388) FP7 project partially funded by the European Commission.

References

[1] V. Ciancaglini, M. Balduzzi, R. McArdle, and M. Rösler, The Deep Web, Trend Micro, 2015.

[2] H. Chen, Dark Web: Exploring and Data Mining the Dark Side of the Web, Springer, 2011.

[3] C. Aliprandi et al., “CAPER: Crawling and analysing Facebook for intelligence purposes”, in IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM), Beijing, China, 2014, pp. 665-669.

[4] G. Kalpakis et al., “Interactive Discovery and Retrieval of Web Resources Containing Home Made Explosive Recipes”, in 18th International Conference on Human-Computer Interaction (HCI 2016), Toronto, Canada, 2016, pp. 221-233.

[5] C. Iliou, G. Kalpakis, T. Tsikrika, S. Vrochidis, and I. Kompatsiaris: “Hybrid Focused Crawling for Homemade Explosives Discovery on Surface and Dark Web”, in 11th International Conference on Availability, Reliability and Security (ARES 2016), Salzburg, Austria, 2016.

[6] G. Agarwal, G. Kabra, K.C.C. Chang, “Towards rich query interpretation: walking back and forth for mining query templates”, in 19th ACM

International Conference on World Wide Web (WWW 2010), Raleigh, North Carolina, USA, 2010, pp. 1-10.

[7] T. Tsikrika et al., “A Framework for the Discovery, Analysis, and Retrieval of Multimedia Homemade Explosives Information on the Web”, in 10th International Conference on Availability, Reliability and Security (ARES 2015), Toulouse, France, 2015, pp. 601 – 610.

[8] G. Kalpakis et al., “Concept Detection on Multimedia Web Resources about Home Made Explosives”, in 10th International Conference on Availability, Reliability and Security (ARES 2015), Toulouse, France, 2015, pp. 632 - 641.