

# Report on Cybersecurity Tools: Carrot2 and Zapmeta

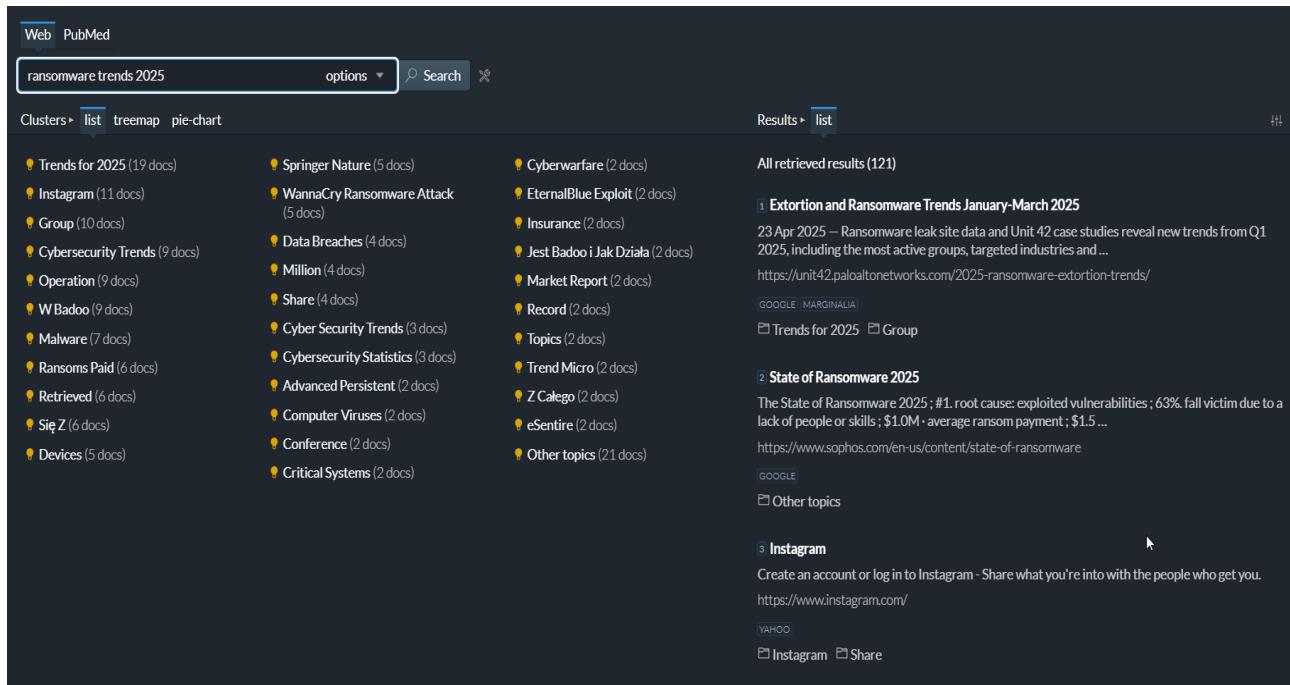
Vishwas S Adhikari | intern id : 135 | [vishwas.s.1002@gmail.com](mailto:vishwas.s.1002@gmail.com)

## Tool 1 : carrot2

- **What's the Story Behind Carrot2?**
  - Started as a Master's thesis project by Dawid Weiss in 2001, testing a clustering algorithm on Polish search results.
  - Went open-source in 2002, with the first official release (v1.0) in 2006.
  - Evolved through versions: v2.0 improved the UI, v3.0 (2009) boosted clustering quality, and v4.0 (2020) simplified the API.
  - The desktop Workbench was replaced with a web-based version in v4.1.0 (2021).
  - Now maintained by Carrot Search, a company founded in 2005 focused on text mining and visualization.
- **What's Carrot2 All About?**
  - It's an open-source **search results clustering engine** written in Java.
  - Takes messy, unstructured text (like search results or document abstracts) and organizes it into neat, thematic groups.
  - Think of searching "cybersecurity threats" and getting results clustered into "Malware," "Phishing," or "Ransomware" instead of a long, overwhelming list.
  - Helps you quickly grasp the big picture and zoom in on what matters.
- **Key Features That Make It Cool**
  - **Automatic Clustering:** Groups documents into labeled topics without needing predefined categories.
  - **Two Powerful Algorithms:** Lingo (uses math magic like singular value decomposition) and STC (Suffix Tree Clustering).
  - **Plays Nice with Others:** Works with search engines like Apache Solr, Lucene, Microsoft Bing, and PubMed.
  - **Cross-Platform Friendly:** Java-based but can connect with non-Java platforms (e.g., PHP, Ruby) via a REST interface. Also has a C#/.NET API.
  - **Visual Goodies:** Shows results as treemaps or pie charts, making data easy to digest.
  - **Language Support:** Handles text in seven languages with tools like tokenizers and stopword filters.
  - **Free and Open:** Uses the BSD license, so it's free to use and tweak.
  -

- **What's Inside Carrot2? (Modules)**
  - **Document Sources:** Pulls data from places like Bing, PubMed, Solr, or Lucene.
  - **Clustering Algorithms:** Lingo and STC do the heavy lifting to create clusters.
  - **Supporting Apps:**
    - **Document Clustering Server (DCS):** A REST service for easy integration.
    - **Web-based Workbench:** A user-friendly GUI to experiment with clustering.
- **How Does It Help in Cybersecurity?**
  - Perfect for **Open Source Intelligence (OSINT)**—think sifting through tons of articles, forum posts, or reports.
  - Organizes data into clusters like “Threat Actors,” “Malware Types,” or “Targeted Industries,” saving you hours of manual sorting.
  - Visualizations highlight trends and connections, like linking a malware to a specific vulnerability.
  - Speeds up research by letting you focus on the most relevant clusters.
- **Proof of Concept (PoC) – Investigating ransomware trends of 2025**

In the following activity I used carrot2 to cluster different ransomware trends and its sources



- The list view shows the source documents supporting the following search results and its walkarounds
- The right side of the search results shows the related documents from various sources like Google , Yahoo , Bing etc
- This search shows Carrot2 taking a huge pile of PubMed articles about ransomware trends for 2025 and turning it into an easy-to-explore map. The clusters tell you the hot topics—

future trends, social media buzz, cybersecurity angles, and specific threats like "Wladoo." The highlighted summaries give solid stats, like the rising ransom costs and why companies are struggling (skills or vulnerabilities). It's like having a research assistant who sorts everything for you, which is perfect for your internship if you're digging into ransomware threats or OSINT



### Interpretation of Your Treemap for "ransomware trends 2025"

#### 1. Center Cluster: Trends for 2025 (19)

- This is your **main theme**, with **19 articles grouped** around predictions and developments expected in 2025.
- Likely contains reports, expert forecasts, and security firm whitepapers.

#### 2. Blue Clusters

- Instagram (11)**: Could relate to how ransomware groups or scams are spreading on social media (specifically Instagram).
- Operation (9)**: May include operational tactics, named ransomware groups (e.g., REvil), or major case studies.
- Malware (7)**: General discussion on malware evolution, often intersecting with ransomware topics.

### 3. Brown/Olive Clusters

- **Group (10)**: Possibly focuses on **APT groups**, ransomware gangs, or threat actors.
- **W Badoo (9)**: May be a result of misindexed or unrelated social network content (possible noise — needs further inspection).
- **Ransoms Paid (6)**: Focused on statistics and cases where organizations paid ransom.

### 4. Purple Clusters

- **Cybersecurity Trends (9)**: Covers the broader field — may include ransomware but also phishing, data privacy, etc.
- **Springer Nature (5)**: Academic and journal-based results — valuable for sourcing technical insights.

### 5. Green Clusters

- **Devices (5)**: Possibly IoT or mobile devices being targeted.
- **Cyberwarfare (2)**: Discusses ransomware as a nation-state weapon.

### 6. Red Clusters

- **Retrieved (6)**: May refer to recovered files, decryptors, or leaked data.
- **Data Breaches (4)**: Links ransomware to breach incidents.



### • When to Use It?

- **Early Research Phase**: When you're drowning in data (articles, reports, social media posts) and need a quick overview.
- **Best Scenarios**:

- Reviewing tons of threat intelligence reports to spot key themes.
  - Analyzing dark web posts or X posts for OSINT investigations.
  - Summarizing customer feedback or incident reports for patterns.
- **What's Not Perfect?**
  - **Text-Only:** Can't handle images, videos, or other media.
  - **Cluster Quality:** Sometimes needs tweaking, which can be tricky for beginners.
  - **Not a Search Engine:** Relies on other tools to fetch data.
  - **Suggestions:**
    - Add a simpler interface for tweaking clusters.
    - Connect directly to social media APIs (X, Reddit, Telegram) for OSINT.
    - Use AI to make cluster labels more descriptive.
- **Why It's Awesome**
  - Saves time by organizing chaos into clear, actionable insights.
  - Visualizations are intuitive and make complex data approachable.
  - Open-source and free, with tons of flexibility for custom workflows.
  - Great for anyone dealing with large text datasets in cybersecurity.

## Tool 2: Zapmeta

- **What's the Story Behind Zapmeta?**
  - Launched as a meta-search engine by a team of college students, emerging from an eight-month beta in 2004.
  - Designed to aggregate results from multiple search engines for a comprehensive search experience.
  - Operates out of Texas, USA, and is maintained by Zapmeta.com LLC, focusing on user-friendly search capabilities.
  - Has faced criticism as a potential browser hijacker, but remains a tool for gathering diverse web results.[zoominfo.com/searchenginewatch.com2-spyware.com](http://zoominfo.com/searchenginewatch.com2-spyware.com)
- **What's Zapmeta All About?**
  - A **meta-search engine** that pulls results from multiple sources like Yahoo, Microsoft Bing, YouTube, Wikipedia, and Entireweb.
  - Instead of searching one engine, it queries several at once, delivering a broader, more varied set of results.
  - Aims to provide a “one-stop shop” for web searches, including web pages, videos, news, and products.
  - Emphasizes user privacy with functional cookies and non-personalized content.[zapmeta.comzoominfo.com](http://zapmeta.comzoominfo.com)
- **Key Features That Make It Stand Out**
  - **Multi-Source Search:** Combines results from AOL, Google (indirectly via AOL), Yahoo, Bing, YouTube, Wikipedia, and more.
  - **Advanced Search Options:** Filter results by keywords (must include, exact phrase, exclude), region, domain, or max results per page.
  - **Visualization Tools:** Offers “Quickview” to preview websites in a small window and “result snapshots” for thumbnail images of pages.
  - **Sorting Flexibility:** Sort results by relevance, domain, popularity, title, or source.
  - **Deduplication:** Removes duplicate results when grouping is enabled for cleaner output.
  - **Additional Insights:** Links to Alexa Rank for site traffic/ratings and the Wayback Machine for archived page versions.
  - **User-Friendly Design:** Clean interface with gradients and menus, distinct from plain HTML search engines.[metamend.comwebdeveloper.comsearchenginewatch.com](http://metamend.comwebdeveloper.comsearchenginewatch.com)
- **What's Inside Zapmeta? (Modules)**
  - **Web Search:** Core feature aggregating results from multiple search engines.

- **Advanced Search Interface:** Allows precise queries with Boolean operators and domain/region filters.
- **Directory Search:** Uses Open Directory Project (DMOZ) data for structured listings.
- **Shopping Search:** Powered by PriceGrabber for product searches.
- **Preferences Panel:** Customizes result display, grouping, and number of results per page.[metamend.com/searchenginewatch.com](http://metamend.com/searchenginewatch.com)
- **How Does It Help in Cybersecurity?**
  - **OSINT Powerhouse:** Ideal for gathering intelligence from diverse online sources in one go.
  - Queries multiple platforms (e.g., YouTube, Wikipedia, Bing) to collect data on threat actors, vulnerabilities, or incidents.
  - Advanced filters help narrow down results to specific regions, domains, or keywords, crucial for targeted OSINT.
  - Visual previews and snapshots speed up assessing the relevance of sources without visiting each site.
  - Useful for discovering obscure or archived content via Wayback Machine links, aiding historical threat research.
- **Proof of Concept (PoC) –**

### **1. Using zapmeta for searching the Trends and topics across different browsers**

Web results for **Ai In Cybersecurity 2025 Trends**

10 Cyber Security **Trends** For 2025 - SentinelOne

[www.sentinelone.com](http://www.sentinelone.com)

Jul 14, 2025 ... Robust endpoint encryption and zero trust segmentation are among the top cyber security **trends** driving this sector, which has been stimulated by ...

**Cybersecurity** considerations 2025 - KPMG International

[kpmg.com](http://kpmg.com)

The ever-evolving role of the CISO · The power of the people · Embed trust as **AI** proliferates · Harness **AI** for cyber: Racing ahead vs racing safely · Platform ...

The State of **Cybersecurity: 2025 Trends** Report® - Arctic Wolf

[cybersecurity.arcticwolf.com](http://cybersecurity.arcticwolf.com)

Both a source of concern and a growing tool **in** organizations' tech stacks, **AI** is poised to become a transformative technology **in cybersecurity**. Checkmark icon.

**2025 Cybersecurity** Predictions - Palo Alto Networks

[www.paloaltonetworks.com](http://www.paloaltonetworks.com)

In 2025, the **cybersecurity** landscape will undergo a transformative shift toward a unified data platform encompassing everything from code development to cloud ...

Top 12 Cyber Security **Trends** And Predictions For 2025 - Splashtop

[www.splashtop.com](http://www.splashtop.com)

Jul 7, 2025 ... In 2025, artificial intelligence will play a pivotal role **in** transforming **cybersecurity** by enhancing both detection and mitigation efforts.

**5 Cybersecurity Trends** to Watch in 2025

[sps.wfu.edu](http://sps.wfu.edu)

May 16, 2025 ... **AI** **Cybersecurity** Threats and Defenses ... Unsurprisingly, artificial intelligence (**AI**) is one of the major trends **in cybersecurity** today. It is ...

## 2. Google Dorking to find sensitive and critical info

The screenshot shows the Zapmeta search engine interface. At the top, the Zapmeta logo is displayed with its characteristic multi-colored arrow icon. Below the logo, there is a search bar with the text "Web More" and a magnifying glass icon. To the right of the search bar is a vertical toolbar with several icons. The main area of the interface displays a list of search results. The first result is a link to "CREDENTIAL SEARCH RESOURCES - Pastebin.com". The second result is a link to "LEAKED COLLEGE CREDENTIALS. - Pastebin.com". The third result is a link to "Pemiblanc Credential Stuffing archive download - Pastebin.com". The fourth result is a link to "List of Google ASE Queries/Dorks - Pastebin.com". The fifth result is a link to "pw - Pastebin.com". Each result includes the URL and a brief description.

Web More

site:pastebin.com leaked credentials

site credentials

leaked credit cards pastebin

hotmail com site pastebin com

site pastebin com credit card

login credentials site

update credentials site

Web results for **Site:pastebin.com Leaked Credentials**

CREDENTIAL SEARCH RESOURCES - Pastebin.com  
pastebin.com  
Oct 31, 2019 ... leaked.site. breachalarm.com. breachaware.com. www.avast.com. intelx.io. psbdmp.ws. ashley.cynic.al. breachrecon.com. databases.today. leak- ...

16BilPasswordsSOURCES - Pastebin.com  
pastebin.com  
Jun 22, 2025 ... https://cybernews.com/security/billions-credentials-exposed-infostealers-data-leak/ ... leaked-across-apple-google-more-what-to-know-and ...

LEAKED COLLEGE CREDENTIALS. - Pastebin.com  
pastebin.com  
Jun 24, 2018 ... 1 x1 2018-04-25 19:55:13 sramadhoss@yahoo.co.in 1 9443713531  
KEERTHANA R 0 · 2 x1 2018-04-26 02:18:48 lakhmigaja21@gmail.com 2 9600093458 ...

Pemiblanc Credential Stuffing archive download - Pastebin.com  
pastebin.com  
Nov 26, 2021 ... The M Record Pemiblanc Credential Stuffing List bltadwin.ru Lorem Ipsum is simply dummy text of the printing and typesetting industry.

List of Google ASE Queries/Dorks - Pastebin.com  
pastebin.com  
Nov 30, 2018 ... [Did I miss interesting ones which are useful for recon? Reach out to @payloadartist to get them added!] \_news/news.php?id= filetype:swf inurl: ...

pw - Pastebin.com  
pastebin.com

You used **Zapmeta**, a metasearch engine, to conduct a focused OSINT (Open Source Intelligence) search using the query:

***site:pastebin.com leaked credentials***

This query instructs the search engine to find results specifically on **Pastebin.com** that are related to **leaked credentials**.

1.	1	x1	2018-04-25	19:55:13	sramadhoss@yahoo.co.in	1	9443713531	KEERTHANA R 0
2.	2	x1	2018-04-26	02:18:48	lakhmigaja21@gmail.com	2	9600093458	HARIHARAN 0
3.	3	x1	2018-04-26	08:44:40		3	9840739334	Vidyalakshmi 0
4.	4	x1	2018-04-26	10:46:01		4	7845548468	DHVAKAR J 0
5.	5	x1	2018-04-26	10:51:21	preethirass2305@gmail.com	5	9597789789	Preethi 0
6.	6	x1	2018-04-26	10:56:37		6	9710584266	SANDHIYA N 0
7.	7	x1	2018-04-26	11:20:57	gopig472@gmail.com	7	6381003713	David 0
8.	8	x1	2018-04-26	12:01:26	babujii1185@gmail.com	8	9962185446	Bhakiyalakshmi 0
9.	9	x1	2018-04-26	13:44:53		9	9940067788	SWETHA V 0
10.	10	x1	2018-04-26	14:36:23	newstarxeroxs2010@gmail.com	10	9087941112	Lakshmi 0
11.	11	x1	2018-04-26	23:43:25	gg234366@gmail.com	11	9566096407	D. Ganesan 0
12.	12	x1	2018-04-26	23:46:39	gg234366@gmail.com	12	9566096407	D. Ganesan 0
13.	13	x1	2018-04-27	09:41:15		13	9087941112	Lakshmi 0
14.	14	x1	2018-04-27	10:52:47	electrotech.ind2009@gmail.com	14	9952073590	selvanathan 0
15.	15	x1	2018-04-27	12:11:13	SRIJI46@GMAIL.COM	15	8190920783	Nivedha R 0
16.	16	x1	2018-04-27	13:58:45		16	9943180124	Bharakavi M 0
17.	17	x1	2018-04-27	13:59:58		17	8680877422	Narmadha M 0
18.	18	x1	2018-04-27	14:40:41		18	9884606228	AKASH V 0
19.	19	x1	2018-04-27	15:07:05	rajrose30@gmail.com	19	8958654545	rajesh 0
20.	20	x1	2018-04-27	15:38:07		20	9941358554	KALAISELVI P 0
21.	21	x1	2018-04-27	19:28:34		21	9710365101	b.devi 0

Zapmeta showed **related auto-suggestions** like:

- “leaked credit cards pastebin”
- “hotmail com site pastebin com”
- “site pastebin com credit card”

This indicates that the engine is aggregating results potentially from multiple engines (Google, Yahoo, Bing, etc.) and highlighting common dark-web-style queries.

## 1. Confirms Data Exposure

You verified that **actual credential leaks** (even if old) are still **accessible** through search engines using simple dorks. This proves the surface web still hosts highly sensitive material.

## 2. Highlights Search Engine Vulnerabilities

Zapmeta aggregates results from sources like Google , Yahoo , Bing , Firefox . This makes it easier to discover sensitive documents in **one place**, increasing the **risk of exposure** for breached credentials.

- **When to Use It?**
  - **Initial Reconnaissance:** When starting an OSINT investigation to gather broad data from multiple sources.
  - **Best Scenarios:**
    - Researching a new threat actor across web, video, and news sources.
    - Finding archived versions of defaced websites or old threat reports via Wayback Machine.
    - Filtering results to specific domains (e.g., .gov or .edu) for credible intelligence.
- **When to Use in an Investigation?**
  - Use during the **intelligence-gathering phase** of a cybersecurity investigation.
  - Collects diverse data to build a comprehensive picture of a threat, campaign, or actor.
  - Helps verify information by cross-referencing multiple sources in one interface.
- **Who Should Use It and What Skills Do They Need?**
  - **Best Users:** OSINT analysts, threat intelligence researchers, or cybersecurity interns.
  - **Skills Needed:**
    - Familiarity with search engine mechanics and Boolean operators for advanced queries.
    - Analytical skills to evaluate the relevance and credibility of results.
    - Basic cybersecurity knowledge to contextualize findings (e.g., identifying phishing-related content).
- **What's Not Perfect?**
  - **Browser Hijacker Concerns:** Some users and AV tools flag Zapmeta as a potentially unwanted program (PUP) that may alter browser settings or show ads.[2-spyware.com](#)
  - **Slower Speed:** As a meta-search engine, it's slower than single engines like Google due to querying multiple sources.[webdeveloper.comresearchgate.net](#)
  - **Sponsored Links:** Sponsored results appear first, which can clutter the SERP, though they're clearly marked.[metamend.com](#)
  - **Limited Source Agreements:** Lacks direct access to Google; relies on AOL for partial Google results.[searchenginewatch.com](#)
  - **Suggestions:**
    - Improve transparency and user consent to avoid PUP classification.
    - Optimize speed by caching more results or streamlining source queries.

- Enhance filtering to prioritize non-sponsored results by default.
  - Secure direct agreements with major engines like Google for broader coverage.
- **Why It's Awesome**
    - Pulls from multiple sources for a wider net in OSINT research.
    - Advanced search and visualization tools make data exploration intuitive.
    - Free to use with a clean, visually appealing interface.
    - Links to tools like Wayback Machine add unique value for historical research.