e-book

OSINT: NOWY WYMIAR POSZUKIWAŃ W SIECI

Krzysztof Wosiński

securITum

Krzysztof Wosiński

OSINT
NOWY WYMIAR POSZUKIWAŃ W SIECI

BEZPŁATNY FRAGMENT

securitum

Projekt okładki: Krzysztof Kopciowski
Projekt typograficzny: Krzysztof Kopciowski
Redaktor: Tomasz Turba
Redakcja merytoryczna: Foxtrot Charlie

Redaktor prowadzący: Katarzyna Sajdak
Adiustacja językowa: Magdalena Anioł
Skład i łamanie: Krzysztof Kopciowski
Korekta: Magdalena Anioł, Ewa Budka, Paulina Lenar

ISBN: 978-83-968874-3-6

Wydanie I, w wersji elektronicznej [plik .pdf]
Kraków 2024

Securitum Wydawnictwo Sp. z o.o.
ul. Siostry Zygmunty Zimmer 5
30-441 Kraków
e-mail: wydawnictwo@securitum.pl
www.securitum.pl

Zastrzeżenia prawne

Bezpieczeństwo IT ma coraz większe znaczenie. Nie można profesjonalnie zabezpieczyć aplikacji czy systemu, nie znając technik ich atakowania. Omawiamy je w tej książce, ponieważ to bardzo skuteczny sposób podnoszenia wiedzy i świadomości użytkowników, administratorów i twórców aplikacji.

Wszelkie podawane przez nas informacje powinny jednak być wykorzystywane wyłącznie w granicach prawa, co z reguły oznacza zakaz wykorzystywania omawianej tu wiedzy bez zgody dysponenta systemu czy sieci. Wyjście poza te granice może skutkować zarówno odpowiedzialnością cywilną (np. obowiązkiem naprawienia wyrządzonej szkody), jak i odpowiedzialnością karną. Przykładowo, zgodnie z polskim kodeksem karnym nieuprawnione uzyskanie dostępu do systemu informatycznego lub jego części podlega grzywnie, karze ograniczenia wolności albo pozbawienia wolności do lat 2 (art. 267 §2 kodeksu karnego). Z kolei nieuprawnione zakłócenie w istotnym stopniu pracy systemu informatycznego, systemu teleinformatycznego lub sieci teleinformatycznej przez transmisję, zniszczenie, usunięcie, uszkodzenie, utrudnienie dostępu lub zmianę danych informatycznych podlega karze pozbawienia wolności od 3 miesięcy do lat 5 (art. 269a kodeksu karnego).

Zwracamy na to uwagę, ponieważ **nie jest naszym zamiarem wspieranie jakichkolwiek bezprawnych działań**. Dlatego zastrzegamy, że w najszerszym prawnie dopuszczalnym zakresie wyłączamy naszą odpowiedzialność za skutki takich działań.

Od Redakcji

Seria e-booków ukazująca się nakładem Securitum ma na celu przybliżanie wybranych zagadnień z szeroko rozumianego IT w jak najbardziej praktycznym podejściu.

Tworzymy ją dla osób, które potrzebują bodźca, żeby rozpocząć samodzielne zgłębianie jakiegoś tematu, ale albo trochę się wahają, albo nie mają pewności, że to droga dla nich.

Książki w tej serii będą też swoistymi poradnikami dla bardzo początkujących, szukających wprowadzenia w temat na poziomie podstawowym.

Trzy pierwsze e-booki wydawane przez Securitum są trzema różnymi ścieżkami prowadzącymi w świat bezpieczeństwa IT, na które zapraszamy Czytelników portalu sekurak.pl i słuchaczy Sekurak.Academy:

1. Adam Samson, *Bezpieczeństwo domowego routera Wi-Fi. Wprowadzenie*
2. Krzysztof Wosiński, *OSINT: nowy wymiar poszukiwań w sieci*
3. Grzegorz Trawiński, *Certyfikacje ofensywne w cyberbezpieczeństwie*

securITum

SEKURAK.ACADEMY

Miejsce, w którym rzetelna wiedza łączy się z dużą dawką pozytywnej energii i dobrego humoru. Świetna opcja na udaną inwestycję w siebie.

DOŁĄCZ DO NAS

>>> **https://sekurak.academy/** <<<

-25%

z kodem: **e-akademia**

Spis treści

Od Autora 15

Część I: OSINT – techniki 16

R1: OPSEC przede wszystkim 17

Błąd 1: Przyznanie, że konta w ogóle istnieją 17

Błąd 2. Powiązanie kont prywatnych z anonimowymi kontami 18

Błąd 3. Używanie nazwy użytkownika związanej z wyszukiwaną osobą 19

Błąd 4. Komentarze w jednym temacie 20

Na zakończenie 21

Podsumowanie i rady 22

R2: Google'a szkiełko i oko, czyli co nowego w wyszukiwaniu zawartością obrazów 23

Miejsca 24

Tekst 26

Samochody 28

Owoce 30

Logo 32

Twarze 34

Podsumowanie i rady 36

R3: Pasywny rekonesans kont Google i Microsoft 38

Google 38

Microsoft 42

Gravatar 42

Podsumowanie i rady 43

R4: Ćwierkając w czasie i przestrzeni, czyli analiza geolokalizacji wpisów na Twitterze/X 43

Narzędzia dostępne w sieci 45

Podsumowanie 46

R5: InVID – *In Video Veritas*. Ciekawe narzędzie przydatne w białym wywiadzie 46

Projekt InVID 47

Analiza materiałów wideo 48

Inne narzędzia 50
Podsumowanie 52
R6: Sherlock i doktor WhatsMyName – wyszukiwanie profili po nazwie użytkownika 53
Sherlock 53
WhatsMyName 54
Podsumowanie i rady 55
R7: Na OSINT-owym tropie tajnej niemieckiej organizacji 55
Akt I. Wykopki 56
Akt II. Wejście 59
Akt III. Namierzanie 60
Akt IV. Finał? 62
Post scriptum 62
Podsumowanie i rady 63
R8: OSINT w pandemii, czyli co wynikało z wyników testów 63
Przypadek Jacka Kurskiego 64
Przypadek Novaka Djokovicia 64
Podsumowanie i rady 71
R9: #OSINTForGood, czyli śledztwa w słusznej sprawie 71
Odczarowywanie OSINT-u 72
Działania na rzecz dzieci 74
Bellingcat 75
Każdy może pomóc 76
Podsumowanie i rady 77
R10: #F12IsNotACrime, czyli opowieść o zaglądaniu w kod źródłowy 78
Ku przestrodze 78
Identyfikatory Google Analytics 80
Podsumowanie i rady 82
R11: Wielki Brat patrzy, czyli OSINT z wykorzystaniem zdjęć satelitarnych 82
Tam sięgaj, gdzie wzrok nie sięga 82
Wizualizacja zmian 84
Mnogość danych 86
Podsumowanie i rady 86

R12: Analiza czasowa z wykorzystaniem (nieco) ukrytych danych ... 87
Analiza w Bing Maps ... 87
Analiza kodu strony ... 89
Podsumowanie i rady ... 91
R13: Metadane – Święty Graal czy ślepy zaułek? ... 92
Eksportowanie zdjęć z metadanymi z plików PDF ... 92
Metadane z filmów publikowanych w serwisach internetowych ... 94
Metadane pozyskane za pomocą narzędzia ExifTool ... 95
Podsumowanie i rady ... 97
R14: Co powie mapa, czyli o otwartych danych w serwisach mapowych ... 97
Nie wszystko, co widzisz, jest prawdziwe ... 97
Ograniczenia w Google Street View ... 100
Chiński pomysł na zabezpieczenie przed dokładną lokalizacją ... 101
Podsumowanie i rady ... 102
R15: OSINT^2, czyli kilka słów o weryfikacji danych ... 103
Poszukiwania Andrew Tate'a ... 103
Nowe oznaczenia kont na Twitterze/X ... 105
Podsumowanie i rady ... 107
R16: OSINT kontra dezinformacja, czyli jak zdemaskować trolla ... 108
Na początek: dlaczego w tytule nie znalazł się znak zapytania ... 108
Źródło ... 109
Autor ... 109
Czas i miejsce ... 111
Grafika komputerowa ... 112
Wyjście z bańki ... 114
Podsumowanie i rady ... 115
R17: Jak anonimowo pozyskać informacje o pracownikach firmy – przypadek LinkedIn i wzorce adresów e-mailowych ... 116
Pasywne przeglądanie prywatnych profili na LinkedIn ... 116
Od imienia i nazwiska do adresu e-mail ... 117
Podsumowanie ... 118

R18: Po słonecznej stronie OSINT-u – wykorzystanie cieni do weryfikacji zdjęć 119
Analiza cieni 119
Podsumowanie i rady 120
R19: Jedno zdjęcie, by znaleźć ich wszystkich – OSINT na podstawie materiałów wizualnych 120
Zdjęcie, które doprowadziło do zespołu odpowiedzialnego za programowanie rosyjskich pocisków 121
Przybliż... przybliż... przybliż... 123
Kategorie poszukiwanych wskazówek 124
Podsumowanie i rady 126
R20: Apki, słuchawki i inne wyciekające informacje 126
Podsumowanie i rady 131
R21: Śledzenie osób – jak czytanie z otwartej książki 132
Powrót problemów aplikacji do śledzenia tras biegów/wycieczek 132
Gdzie leży przyczyna? 133
Anonimowe dane także można przypisać 134
Podsumowanie i rady 135
R22: O jeden OSINT za daleko, czyli przykre skutki złych analiz 135
Poszukiwania zamachowców 136
Udostępnianie zdjęć wojskowych 138
Poszukiwanie sprawców ataku na Kapitol 139
Podsumowanie i rady 140

Część II: OSINT – narzędzia 141
R23: Jak stworzyć własny OSINT-owy lab do śledzenia samolotów i innych obiektów latających 142
Źródła danych 142
Trochę teorii 145
Własne laboratorium 147
Sprzęt 147
Software 149
Podsumowanie i rady 151

R24: GHunt, czyli jak wycisnąć jak najwięcej informacji z profilu Google 151
Czym jest GHunt? 151
ChromeDriver 153
Podsumowanie 153
R25: WIPO – wyszukiwarka patentów i znaków handlowych 153
Global Brand Database 153
PATENTSCOPE 154
Global Design Database 155
OpenCorporates 156
Podsumowanie i rady 156
R26: OSINT poza Google Dorks – operatory innych wyszukiwarek 157
Operatory wyszukiwania 159
Wyszukiwanie na Twitterze/X 159
Podsumowanie i rady 161
R27: Rozszerzenia do przeglądarek ułatwiające OSINT 162
Mitaka – wyszukiwarka informacji o IoC 162
Gotanda – wyszukiwarka informacji o IoC + media społecznościowe 163
Sputnik – kolejna wyszukiwarka informacji 163
EXIF Viewer – analiza metadanych w obrazach 164
BuiltWith – analiza budowy strony 165
User-Agent Switcher – zmiana identyfikacji przeglądarki „w locie" 167
Podsumowanie i rady 167
R28: Jak narzędzia AI zmieniają OSINT 168
ChatGPT 168
Midjourney 171
Podsumowanie i rady 174
R29: Czy geolokalizacja może zrobić się sama? 175
Pozyskiwanie danych z map cyfrowych 175
Więcej automatyzacji! 177
Podsumowanie 183

Część III: OSINT *hints* 184

R30: Mastodon – OSINT-owa analiza coraz popularniejszej alternatywy wobec Twittera/X 185

Witaj w fediwersum 186

OSINT w Mastodonie 187

Podsumowanie i rady 190

R31: Przekierowania z krótkich linków i jak je znaleźć 191

Modyfikacje linków przekierowujących 191

Google – *goo.gl* 191

Bitly – *bit.ly* 191

TinyURL 192

Serwis *is.gd* 192

Serwis *tiny.cc* 192

Serwisy „wydłużające" linki 192

Rozszerzenia/dodatki do przeglądarek 194

Zestawienie modyfikatorów adresów URL dla różnych serwisów 194

Podsumowanie i rady 194

R32: Skryptozakładki – jak z zakładki w przeglądarce zrobić narzędzie do OSINT-u 195

Analiza skryptów 195

Tworzenie narzędzi OSINT-owych 196

Podsumowanie i rady 200

R33: Kolekcjonowanie dowodów: recon a zrzuty ekranowe stron w sieci 200

WMN_screenshooter 200

EyeWitness 201

Dodatki 202

Podsumowanie i rady 203

Część IV: OSINT w biznesie 204

R34: Jak wykorzystać OSINT do ochrony biznesu? 205

Dokumenty 205

Adresy e-mail 206

Media społecznościowe – firmowe i prywatne 207

Zdjęcia 208
Kluczowe osoby 209
Podsumowanie i rady 209
R35: Podstawy obrony przed atakami socjotechnicznymi 209
Przygotowanie i przeprowadzenie ataku 211
Najczęstsze rodzaje ataków 212
Święta nie tylko od święta 212
Ludzie tacy jak my 213
Przysługa za przysługę 214
Uratuj mój świat 214
Czy jest na sali jakiś lekarz? 215
Owczy pęd 216
Powiedziałeś A, powiedziałeś B... 216
Już tylko pięć minut do końca promocji 216
Nieodpowiedzialni odpowiedzialni 217
Metody obrony 217
Jedno proste pytanie 217
Kontrola – najwyższą formą zaufania 218
Czy aby na pewno wiesz, gdzie jesteś? 219
Aktualizacje i backupy to podstawa 219
To samo hasło wszędzie to drzwi do raju dla przestępcy 220
Nawet najlepsze hasła nie pomogą, jeśli leżą na biurku 220
Taki strój można kupić w każdym sklepie 220
Szkolenie z przykładami, czyli pokaż mi, czego nie potrafię 221
Graj zgodnie z regułami gry 222
Podsumowanie i rady 222
Zakończenie 223

Krzysztof Wosiński. Certyfikowany tester bezpieczeństwa systemów IT (Certified Ethical Hacker) specjalizujący się w systemach militarnych oraz wywiadzie otwartoźródłowym (Open-Source Intelligence – OSINT, zwanym także „białym wywiadem").

Od kilkunastu lat zajmuje się zagadnieniami jakości oraz bezpieczeństwa sprzętu i oprogramowania o przeznaczeniu wojskowym, produkowanych dla polskich i amerykańskich odbiorców. Obecnie koncentruje się na cyberbezpieczeństwie systemów tworzonych dla Sił Zbrojnych RP.

Autor popularnych serii OSINT HINTS oraz Czwartki z OSINT-em w portalu sekurak.pl.

Współautor I tomu książki *Wprowadzenie do bezpieczeństwa IT:* rozdział *Bezpieczeństwo danych w spoczynku – szyfrowanie i usuwanie danych.*

W Securitum od 2020 roku prowadzi szkolenia z OSINT-u, bezpieczeństwa działań w sieci oraz z przeciwdziałania dezinformacji. Jest autorem serii szkoleń „OSINT master" (część 1: Poszukiwanie informacji o osobach, miejscach i pojazdach oraz część 2: Geolokalizacja osób – jak szukać informacji i jak nie dać się znaleźć) i OSINT dla każdego – podstawy poszukiwań informacji w Internecie.

Prowadzi także dwudniowe szkolenie OSINT/OPSEC – narzędzia, techniki śledcze, ochrona przed śledzeniem, podczas którego można bliżej zapoznać się z technikami i narzędziami, opisanymi w tej książce.

Jest również wykładowcą Akademii Sekuraka, gdzie przybliża uczestnikom tematy związane z OSINT-em i bezpieczeństwem pracy w Internecie.

W 2023 roku obronił doktorat z zakresu OSINT-u, który zwieńczyło wydanie książki *Bezpieczeństwo osób i systemów IT z wykorzystaniem białego wywiadu. OSINT* (Warszawa 2024).

Po godzinach zgłębia zagadnienia związane z socjotechniką i białym wywiadem, a także tworzy narzędzia usprawniające rozpoznanie otwartoźródłowe.

OD AUTORA

Internet. Ostateczna granica. Oto zredagowane na nowo, uzupełnione i poszerzone zapiski z serii artykułów ukazujących się na przestrzeni ostatnich kilku lat w portalu *sekurak.pl* dotyczących **OSINT-u (Open-Source Intelligence)**, czyli wywiadu prowadzonego z wykorzystaniem narzędzi otwartoźródłowych (ang. *open source*). W przypadku opisywanych technik, narzędzi i śledztw są one głównie źródłami internetowymi, chociaż warto pamiętać, że OSINT może korzystać także z klasycznych kanałów informacyjnych, takich jak np. prasa, radio czy telewizja.

Teksty opublikowane w tym e-booku zostały uzupełnione i zaktualizowane do realiów Internetu z chwili oddawania tej książki do publikacji, choć może się okazać, że kiedy ją czytasz, jej otoczenie jest już w innym miejscu.

W pierwszej części, *OSINT – techniki* (rozdziały R1–R22), znaleźć można opis technik wykorzystywanych w ramach działań OSINT-owych, w drugiej, *OSINT – narzędzia* (rozdziały R23–R29), przedstawione zostały wybrane narzędzia najczęściej wykorzystywane w białym wywiadzie, a w trzeciej: *OSINT hints* (rozdziały R30–R33) – wskazówki, które w artykułach publikowanych w portalu *sekurak.pl* ukazywały się w ramach popularnej serii „OSINT Hints". Ostatnia, czwarta część e-booka, *OSINT w biznesie* (rozdziały R34–R35), dotyczy styku technologii z biznesem, a więc technik ataku i sposobów obrony przed działaniami wykorzystującymi OSINT oraz socjotechnikę.

Mam nadzieję, że Czytelnicy tego e-booka znajdą w nim wiele interesujących przypadków, dotyczących działań OSINT-owych, które spowodują zwiększenie apetytu na dalsze zgłębianie tej, bardzo ciekawej i coraz częściej wykorzystywanej w codziennym życiu i pracy, tematyki.

A tych, których przygoda z OSINT-em przyciąga jeszcze bardziej, zapraszam do lektury mojej najnowszej książki: *Bezpieczeństwo osób i systemów IT z wykorzystaniem białego wywiadu. OSINT*[*].

---

* Wosiński K., *Bezpieczeństwo osób i systemów IT z wykorzystaniem białego wywiadu. OSINT*, Warszawa 2024.

CZĘŚĆ I: OSINT – TECHNIKI

R1 OPSEC PRZEDE WSZYSTKIM

W tym rozdziale poruszę temat, który warto rozważyć jeszcze przed rozpoczęciem jakichkolwiek czynności OSINT-owych, dotyczący zapewnienia sobie odpowiedniego poziomu bezpieczeństwa w ramach prowadzonych działań (określanego często jako **OPSEC, czyli Operations Security**).

Jedną z historii, które dają pogląd na to, **jak ważne jest oddzielenie informacji na kontach wykorzystywanych do anonimowych działań od wszelkich elementów prawdziwej tożsamości**, jest przypadek domniemanego wyśledzenia konta dyrektora FBI, Jamesa Comeya. Domniemanego, gdyż nie udało się uzyskać potwierdzenia, że namierzone konto na Twitterze* naprawdę należy do niego, jednak wszystkie poszlaki na to wskazują, łącznie z komentarzem samego zainteresowanego. Przyjrzyjmy się zatem bliżej błędom, które zostały popełnione podczas zakładania i utrzymywania tego konkretnego konta.

Błąd 1: Przyznanie, że konta w ogóle istnieją

Wszystko zaczęło się od pewnego spotkania, a konkretnie od Intelligence and National Security Alliance Dinner, na którym pod koniec marca 2017 roku przemawiał właśnie James Comey. Podczas godzinnego wystąpienia, w którym przekazywał informacje o tym, jak ważne są różne aspekty bezpieczeństwa, a także starał się sprytnie unikać odpowiedzi na zbyt konkretne pytania, padło też stwierdzenie, że jemu „bardzo zależy na prywatności", więc jest, co prawda, na Twitterze oraz Instagramie, gdzie „ma jedynie dziewięciu obserwujących", jednak są to wyłącznie osoby należące do jego najbliższej rodziny i znajomych, gdyż nie lubi się dzielić prywatnymi zdjęciami z ludźmi spoza tego kręgu (dla zainteresowanych – dostępne jest nagranie z tego spotkania (rysunek 1)[1], a temat prywatności poruszany jest w 21. minucie).

* W książce będziemy używać obu nazw tej platformy: Twitter oraz X, w zależności od momentu, z którego pochodzą cytowane wpisy albo omawiane mechanizmy, traktując lipiec 2023 jako moment zmiany jej nazwy.

Rysunek 1. Dyrektor FBI zdradza informację o swoim koncie na Instagramie

Te skąpe informacje o posiadanych kontach, poparte niewielkim dodatkowym śledztwem dziennikarskim Ashley Feinberg[2], dotyczącym szczegółów życia i kariery dyrektora FBI, pozwoliły na odnalezienie informacji na temat jego syna, którego imię (niestety dla poszukiwaczy) jest takie samo, jak drugie imię jego ojca – Brien. Niemniej jednak, w końcu, poprzez wpisy na kontach szkolnych drużyn sportowych na Twitterze, możliwe było dotarcie do konta Briena na Instagramie.

Błąd 2. Powiązanie kont prywatnych z anonimowymi kontami

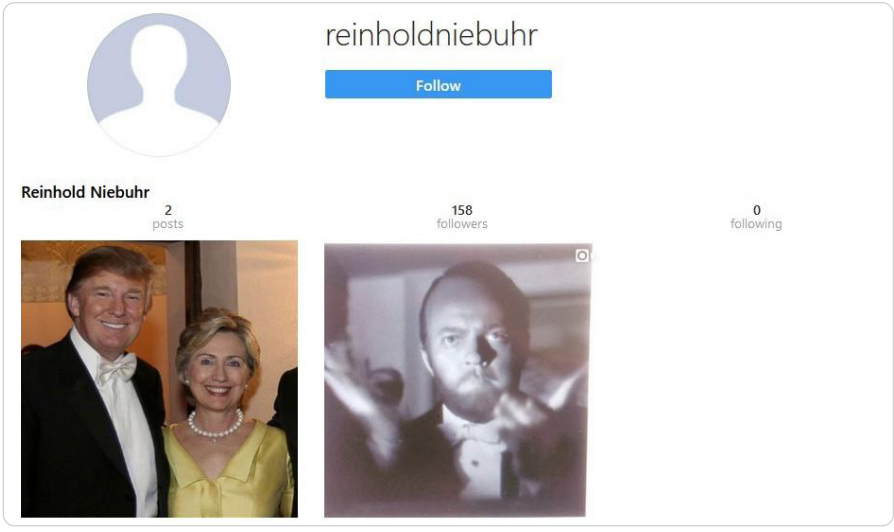
Znaleziony na Instagramie profil został jednak zamknięty, co znacznie utrudniło Ashley Feinberg dalsze poszukiwania. Tutaj jednak przydatna okazała się funkcjonalność, która umożliwiła odnalezienie innych powiązanych kont, a polegała ona na próbie śledzenia zablokowanego konta, a następnie na skorzystaniu z funkcjonalności podpowiedzi innych znajomości, które zostały dobrane na podstawie wcześniejszego wyboru.

Planując takie działania, najlepiej stworzyć zupełnie nowe, „czyste" konto (ta technika ma zastosowanie w wypadku wielu portali społecznościowych, nie tylko Instagrama) i pozwolić, by algorytmy zaproponowały dodanie kolejnych znajomych na podstawie jednej tylko osoby. Można tu wykorzystać „czysty" (nieskonfigurowany) telefon, z wpisanym jednym kontaktem, do którego oczywiście damy dostęp apce społecznościowej, a sugerowane kontakty wskażą najprawdopodobniej powiązane z nią osoby (albo jej dane osobowe, jeśli np. posiadamy tylko numer telefonu). Zamiast telefonu można wykorzystać także maszynę wirtualną lub emulator.

Błąd 3. Używanie nazwy użytkownika związanej z wyszukiwaną osobą

W omawianym przypadku wśród znalezionych powiązanych kont było jedno, które nie do końca pasowało do innych kont osobistych znajomych i członków rodziny ze względu na dość unikatową nazwę: @reinholdniebuhr (rysunek 2). Tutaj przydało się szybkie przestudiowanie historii edukacji Jamesa Comeya, który, jak się okazało, swego czasu napisał pracę dyplomową m.in. o teologu nazwiskiem Reinhold Niebuhr. Ten szczegół wskazywał, że to konto może faktycznie być poszukiwanym profilem dyrektora FBI, tym bardziej że posiadało ono dziewięć obserwujących je osób, co pokrywało się dokładnie z oświadczeniem Comeya.

Na rysunku 3 widoczne jest konto Jamesa Comeya na Twitterze – tutaj, tak samo jak na Instagramie, jako Reinhold Niebuhr.



Rysunek 2. Konto dyrektora FBI na Instagramie już po jego „odkryciu"

Rysunek 3. Konto Jamesa Comeya na Twitterze – tutaj, tak samo jak na Instagramie, jako Reinhold Niebuhr (stan na 12 kwietnia 2017 roku)

Próba przeskoku z konta na Instagramie na konto na Twitterze nie była łatwa, gdyż profil o tej samej nazwie użytkownika nie wyglądał jak należący do głównego bohatera, więc Ashley Feinberg pozostało szukanie po nazwie użytkownika (a nie loginie). Jednym z kont, które wyglądało obiecująco ze względu na dość skryty profil, było konto: @projectexile7 (rysunek 4). I tutaj kolejny raz dał o sobie znać błąd polegający na wykorzystaniu nazwy, którą można było dość jednoznacznie powiązać z osobą dyrektora FBI, gdyż Project Exile* był jednym z projektów, które współprowadził.

Błąd 4. Komentarze w jednym temacie

Na odnalezionym koncie z Twittera można zauważyć, że znaczna liczba polubionych wpisów dotyczy osoby Jamesa Comeya, a jedyną osobą śledzącą to konto jest jego przyjaciel, Benjamin Wittes. Także analiza kont, które są śledzone przez @projectexile7, wskazuje na powiązania z upodobaniami dyrektora FBI w zakresie źródeł informacji.

Rysunek 4. Mała wskazówka – wpis zretweetowany przez konto @projectexile7 wskazuje, że jednak śledztwo zaimponowało jego właścicielowi

Te wszystkie drobne okruchy oczywiście nie dają pewności co do właściciela znalezionych kont. Należy pamiętać, że w ramach OSINT-u ważnym elementem jest analiza zgromadzonych dowodów i uczciwe odpowiedzenie sobie na

* Project Exile dotyczył zwalczania przestępstw z użyciem broni poprzez eskalację procesów osób nielegalnie posiadających broń na poziom federalny i, co za tym idzie, zwiększenie wyroków oraz tytułowe „wygnanie", czyli osadzenie skazanych z dala od ich miejsc zamieszkania.

pytanie: czy na pewno mogę wskazać daną osobę jako powiązaną z badaną sprawą? Nie można w tym miejscu, niestety, poddać się podświadomej chęci poskładania wszystkich elementów w bardzo medialną i robiącą wrażenie całość. Nie jest żadną ujmą przyznanie w pewnym momencie, że nie ma się pewności i podstaw do wydania konkretnych osądów. Niestety, często podświadomie ulegamy emocjom i układamy historię tak, jak chcielibyśmy, aby się ułożyła – wszystkie złe cechy przypisujemy osobom z łatką „tych złych", a zbyt mocno próbujemy wybielić osoby, które wydają nam się sympatyczne i co do których jesteśmy nastawieni bardziej przychylnie.

Na zakończenie

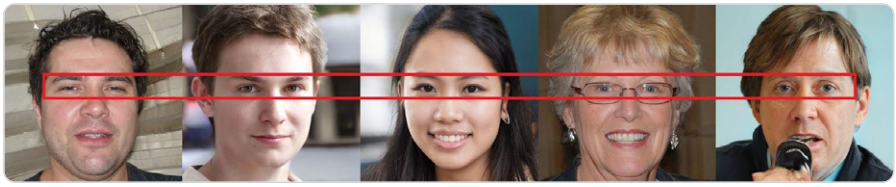
Przedstawiona historia pokazuje kilka błędów popełnianych w ramach przygotowywania środowiska i zasad bezpieczeństwa operacji, czyli OPSEC-u, a także bezpieczeństwa osobistego, czyli PERSEC-u. Po pierwsze, zawsze musimy sobie odpowiedzieć na pytanie, jaki zakres działań powinniśmy wykonać, aby uzyskać odpowiedni poziom bezpieczeństwa, gdyż nie każde śledztwo w Internecie wymaga takiego samego poziomu zabezpieczenia.

Musimy także powstrzymać się od informowania o tym, jakie konta są wykorzystywane do anonimowej pracy i w jakich działaniach biorą udział. Nawet dyrektor FBI nie powstrzymał się od uchylenia rąbka tajemnicy, ale taką pokusę ma też wiele innych osób, zwłaszcza pracujących w sferach życia dotyczących bezpieczeństwa, gdzie pojawia się wiele bardzo ciekawych historii. Niestety, lista osób, z którymi można na dany temat porozmawiać, niekiedy ogranicza się do wąskiego kręgu współpracowników lub – w skrajnych przypadkach – zaledwie jednej osoby.

Tworząc środowisko do pracy, powinniśmy pamiętać o **technologicznej stronie OPSEC-u**, szczególnie w przypadku śledztw o wysokim poziomie ryzyka (czyli np. dotyczących przestępstw, terroryzmu lub mogących wpłynąć na czyjeś bezpieczeństwo osobiste). Począwszy od kwestii typowo **sprzętowo-systemowych** (jakiego komputera, telefonu, łącza internetowego będę używać), poprzez **oprogramowanie** (jaka przeglądarka, dodatki anonimizujące czy VPN), aż po takie elementy, jak **budowanie swoich „kukiełek"** (ang. *sock-puppets*, czyli fałszywych tożsamości, których konta będziemy wykorzystywać podczas prowadzenia śledztwa). Ten ostatni element ma jeszcze wiele dodatkowych kwestii do dopięcia, takich jak np. zdjęcia profilowe, których dobre przygotowanie wymaga odrobinę większego nakładu pracy niż tylko pobranie obrazu z serwisu *thispersondoesnotexist.com*.

Na rysunku 5 widoczne są **zdjęcia** wygenerowane przez model sztucznej inteligencji **GAN (Generative Adversarial Network, generatywne sieci przeciwstawne)**. Niekiedy są one używane jako zdjęcia profilowe dla fałszywych kont w mediach społecznościowych. Używanie obrazów generowanych przez

GAN dla kont w serwisach społecznościowych jest jednak dość łatwe do wychwycenia (rysunek 6). Obecnie dużo lepsze zdjęcia dla kont profilowych można generować za pomocą narzędzi AI, o czym piszę szerzej w rozdziale *Jak narzędzia AI zmieniają OSINT*.



Rysunek 5. Zdjęcia wygenerowane przez model sztucznej inteligencji GAN niekiedy są używane jako zdjęcia profilowe dla fałszywych kont w mediach społecznościowych. Jednym z elementów ułatwiających ich rozpoznanie jest umieszczanie oczu zawsze w tym samym miejscu zdjęcia

Rysunek 6. Przykład fałszywego konta rzekomego ukraińskiego dziennikarza

Dla konta-kukiełki należy także wygenerować odpowiednie dane osobowe, przemyśleć sposoby ukrywania się, a w przypadku posiadania wielu kont także zarządzania ich tożsamościami i ich utrzymywania.

Podsumowanie i rady

- Nie ujawniaj istnienia kont, których „spalenie" może narazić całą operację na porażkę.
- Nie twórz powiązań pomiędzy kontami prywatnymi a tymi wykorzystywanymi do działań OSINT-owych (w momencie kiedy otrzymasz na swoim anonimowym koncie propozycję znajomych z kręgu twoich krewnych lub przyjaciół, już jest za późno).

- Nie wykorzystuj do tworzenia anonimowych kont informacji, które można łatwo powiązać z Tobą.
- Nie zdradzaj swojej prawdziwej tożsamości przez aktywność związaną z Twoimi prawdziwymi zainteresowaniami lub nawiązaniami do Ciebie.

R2 GOOGLE'A SZKIEŁKO I OKO, CZYLI CO NOWEGO W WYSZUKIWANIU ZAWARTOŚCIĄ OBRAZÓW

Kiedy w 2021 roku przeprowadziłem badanie *Jak wyszukiwarki radzą sobie z analizą zawartości obrazów*[3], najpopularniejsze z nich: **Google i Bing**, wypadły dość blado na tle wyszukiwarki **Yandex**. Od tego czasu Google zrobiło duży krok naprzód, integrując stopniowo narzędzie Obiektyw Google (Google Lens) ze swoją wyszukiwarką grafiki. Od listopada 2022 roku narzędzie to jest już domyślnie stosowane w przypadku korzystania z funkcjonalności wyszukiwania za pomocą obrazu w Google. Czas więc na sprawdzenie, jak zabieg ten zmienił układ sił wśród trzech najpopularniejszych wyszukiwarek.

> Google – wyszukiwarka od firmy Google, umożliwiająca m.in. wyszukiwania tekstowe oraz wyszukiwanie obrazem, realizowane z wykorzystaniem narzędzia Obiektyw Google (Google Lens), które zastąpiło Grafikę Google (Google Images).
>
> Bing – wyszukiwarka od firmy Microsoft, umożliwiająca m.in. wyszukiwanie tekstowe oraz wyszukiwanie obrazem (tzw. wyszukiwanie wizualne).
>
> Yandex – rosyjski portal, będący od lutego 2024 w pełni w posiadaniu rosyjskich inwestorów (wcześniej był częścią holenderskiego Yandex N.V.), oferujący m.in. wyszukiwanie tekstowe oraz na podstawie obrazu. W tym drugim zakresie przez długi czas przewyższał swoich zachodnich rywali.

Zanim zagłębimy się w świat analizy zawartości obrazów przez silniki wyszukiwania, chciałbym zaznaczyć, że **badanie to nie obejmuje serwisu TinEye**. Pomijam go celowo, gdyż – mimo bardzo dobrych wyników w wyszukiwaniu obrazów – nie analizuje on ich zawartości, a jedynie sprawdza, czy konkretny obraz nie został umieszczony gdzieś w Internecie (na tyle, na ile oczywiście pozwalają mu na to jego bazy obrazów). **TinEye polecam zatem do wyszukiwania konkretnych zdjęć czy grafik np. w celu odnalezienia ich źródła**, a pod lupę po raz kolejny wezmę trzy najpopularniejsze serwisy: Google, Bing i Yandex. Do badania posłużą te same obrazy, które zostały wykorzystane za pierwszym razem, w celu weryfikacji, czy i w jakim stopniu zmieniły się wyniki wyszukiwania. Dodatkowo wykorzystamy także inne, nowe obrazy, aby upewnić się co do otrzymanego wyniku. Rezultaty poszukiwań poszczególnych wyszukiwarek zostaną porównane, aby zobaczyć, która z nich obecnie radzi sobie najlepiej i która zanotowała największy progres lub regres.

Zasada wyszukiwania za pomocą **Google Lens** jest taka sama jak w przypadku tradycyjnej wersji Grafika Google – wskazujemy link do pliku graficznego, wybieramy plik z obrazem z dysku lub przeciągamy go do okna z otwartą wyszukiwarką Google (nawet z poziomu strony wyszukiwarki tekstowej, niekoniecznie zakładki: GRAFIKA) i upuszczamy we wskazanym polu, aby otrzymać wynik wyszukiwania. Jeśli chcemy skorzystać z tradycyjnej wersji wyszukiwania obrazem (a nie analizy jego zawartości przez Google Lens), musimy po wyszukaniu kliknąć przycisk ZNAJDŹ ŹRÓDŁO OBRAZU usytuowany nad przesłanym przez nas obrazem.

W poszczególnych kategoriach brane były pod uwagę następujące możliwości wyszukiwarek:

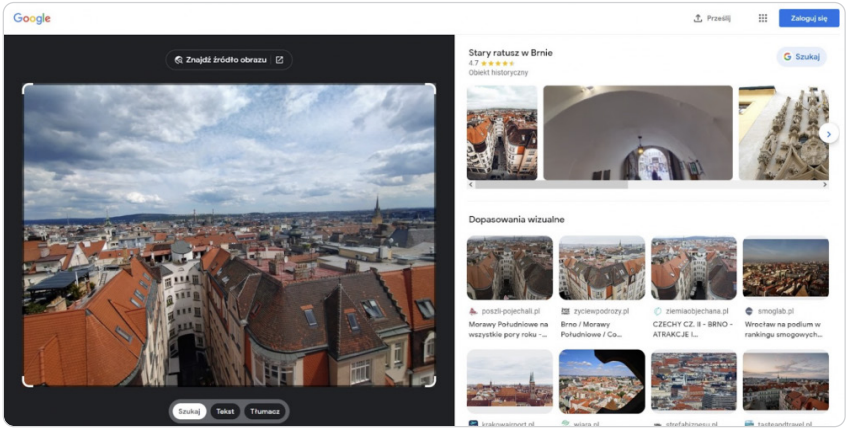
- identyfikacja miejsca na podstawie zdjęcia miasta,
- umiejętność odczytania tekstu w różnych językach na podstawie zdjęć zawierających napisy,
- rozpoznawanie marki i modelu na podstawie zdjęcia samochodu (w zakresie rzadko spotykanych marek samochodów osobowych oraz typowych busów),
- identyfikacja gatunków na podstawie zdjęć owoców,
- rozpoznawanie marek na zdjęciach zawierających logotypy oraz
- rozpoznawanie osób (wskazywanie imienia i nazwiska danej osoby) na zdjęciach, które zawierały twarze. W ostatniej kategorii brane były pod uwagę także zdjęcia przedstawione w skali szarości oraz obrócone o 90 i 180 stopni.

Miejsca

Jako przykład zdjęcia miejsca w poprzednim badaniu posłużyła fotografia paryskiej Statuy Wolności. Wyszukiwarka Google była wówczas w stanie wskazać inne obrazy przedstawiające to miejsce, chociaż trzeba było się im przyjrzeć, aby nie popełnić błędu. Bing nie poradził sobie w ogóle, a Yandex odnalazł miejsce bezbłędnie.

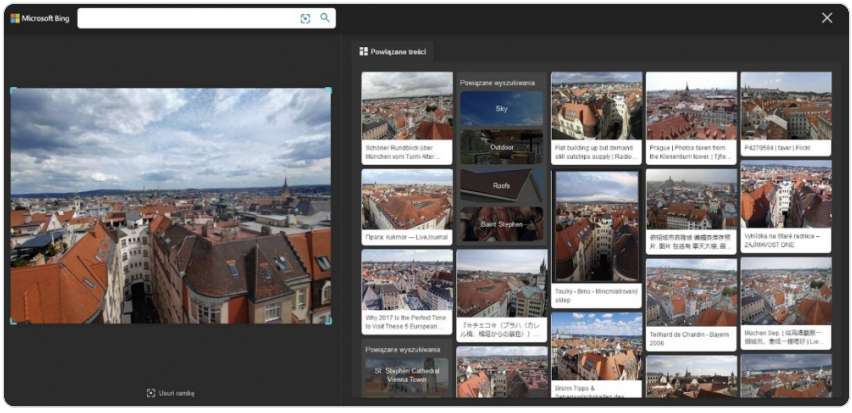
W 2023 roku **Grafika Google** (przy użyciu Google Lens) nie ma już najmniejszego problemu, aby rozpoznać, co znajduje się na zdjęciu, i wskazać od razu nazwę obiektu. W celu sprawdzenia, jak radzi sobie klasyczna wersja wyszukiwania graficznego od Google, kliknąłem w opisywany wcześniej przycisk ZNAJDŹ ŹRÓDŁO OBRAZU. Wśród wyszukanych wyników królowały linki do portalu *sekurak.pl* (ze względu na obecność tam zdjęcia zamieszczonego we wspomnianym już artykule *Jak wyszukiwarki radzą sobie…*, opisującym wzmiankowane na wstępie badanie, z 2021 roku), a zdjęcia podobne wizualnie pokazywały różne miasta nad wodą, jednak żadne z nich nie przypominało poszukiwanego paryskiego widoku.

Równie dobrze Google Lens poradziło sobie ze zdjęciem dachów kamienic na starówce w Brnie (rysunek 7): tutaj także bezbłędnie wskazane zostało miasto, a nawet konkretne miejsce wykonania zdjęcia.



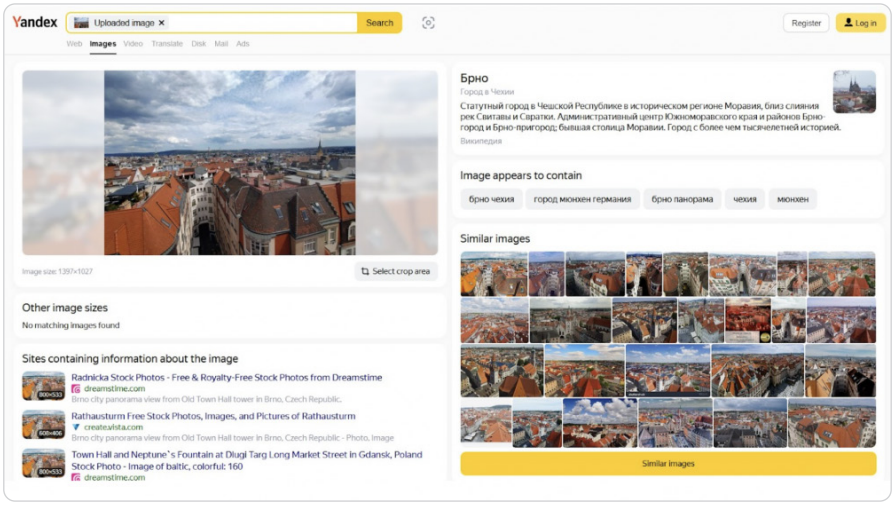
Rysunek 7. Wyszukiwanie w wyszukiwarce Google przy użyciu zdjęcia miasta

Bing (rysunek 8) poradził sobie lepiej niż poprzednio – tym razem wśród podobnych wizualnie obrazów znalazły się zdjęcia miejsc zgodnych z tym, co przedstawiają zdjęcia. Wyszukiwarka ta uzyskuje jednak lepsze wyniki, jeśli użyjemy opcji WYSZUKIWANIE WIZUALNE, dostępnej pod zdjęciem, które zostało przesłane. Mimo tego w przypadku zdjęcia brneńskiej starówki wyniki były zdecydowanie niezadowalające. Wśród wskazanych stron z obrazami znalazły się wprawdzie takie, które dotyczyły tego samego miasta, jednak na liście były także zdjęcia m.in. Monachium czy Pragi. Uzyskując takie wyniki, osoba poszukująca danej lokalizacji musi włożyć dodatkowy wysiłek w znalezienie poprawnych odpowiedzi, a nie zweryfikować jedynie otrzymane informacje.



Rysunek 8. Wyszukiwanie w Bing przy użyciu zdjęcia miasta

Yandex poradził sobie równie dobrze, jak Google – zidentyfikował miejsce łącznie z jego nazwą oraz pokazał zdjęcia poszukiwanego obiektu. W przypadku zdjęcia z Brna dodatkowo opis był dłuższy, jednak pomimo korzystania z anglojęzycznej wersji serwisu tekst był dostępny jedynie w cyrylicy, co może nieco utrudniać analizę osobom, które nie są obeznane z tego typu zapisem (rysunek 9), chociaż można zaznaczyć otrzymany opis i skopiować go do tłumacza online lub wykorzystać funkcje przeglądarki służące do tłumaczenia tekstów.



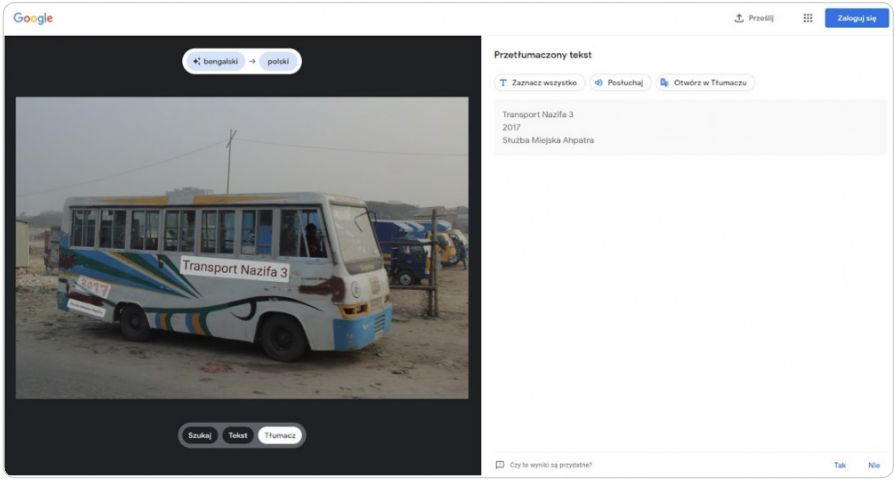
Rysunek 9. Wyszukiwanie w silniku Yandex przy użyciu zdjęcia miasta

Tekst

W analizie z 2021 roku rozpoznawanie tekstu badane było z wykorzystaniem zdjęcia tablicy informacyjnej zapisanej głównie tajskim alfabetem. Najlepiej poradził sobie wtedy Yandex, który był w stanie rozpoznać tekst i przetworzyć go tak, aby możliwe było jego skopiowanie i przetłumaczenie. Dość dobrze poradził sobie Bing, natomiast wyszukiwarka Google wypadła najsłabiej.

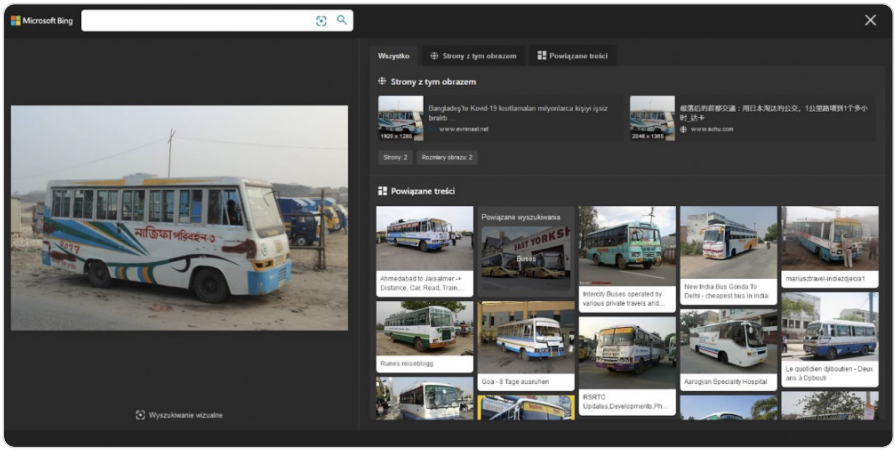
Tym razem **Google** nie miał już najmniejszych problemów z rozpoznaniem tekstu na zdjęciu. Jeśli obraz zawiera jakikolwiek tekst i chcemy go dalej przetwarzać (np. skopiować lub od razu przetłumaczyć), możemy skorzystać z przycisku TEKST znajdującego się poniżej przesłanego obrazu. Dodatkowo w przypadku badanego zdjęcia Google od razu wskazał, jakie to miejsce, dodając opinie i zdjęcia, jednak mogło to być głównie pokłosiem rozpoznanej nazwy, zapisanej w alfabecie łacińskim na zdjęciu. Niemniej skok jakościowy w tym przypadku jest naprawdę znaczny. W celu dodatkowej weryfikacji przeprowadziłem także test z rozpoznawaniem napisu w języku bengalskim (jako przykład dość rzadko spotykanego języka, rysunek 10) oraz umiejętnością odczytywania drogowskazów zapisanych pismem chińskim. W obu przypadkach

Google nie miał problemu z odczytaniem napisu, a dodatkowo dla drogowskazów od razu wskazywał dane miejsce na mapie lub wyświetlał wskazówki dotarcia do niego.



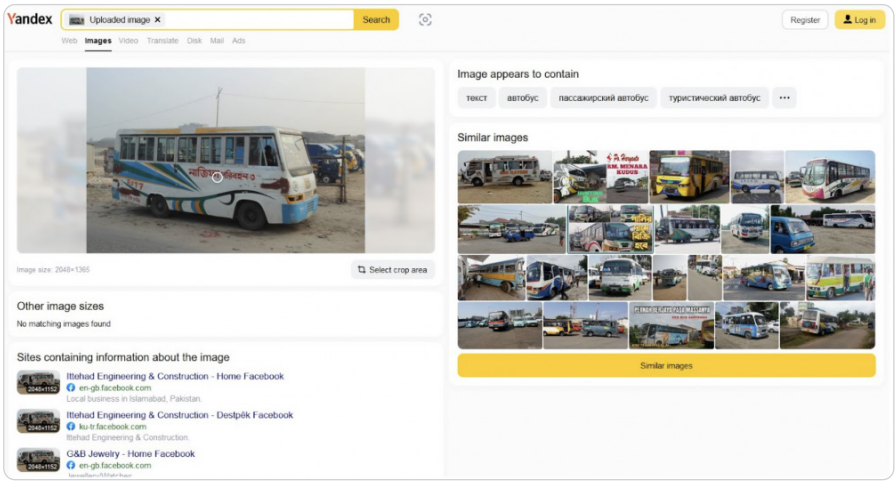
Rysunek 10. Analiza bengalskich napisów w wyszukiwarce Google

Bing, który dość dobrze radzi sobie z rozpoznawaniem tekstu w alfabecie łacińskim, niestety cały czas pozostaje w tyle, jeśli chodzi o inne systemy pisma. Biorąc pod uwagę postępy konkurencji, Microsoft powinien pomyśleć nad aktualizacją i rozwojem tej funkcji, tym bardziej że jej fundament ma już gotowy. Dla wskazanych zdjęć odnalezione zostały poprawnie ich wystąpienia na powiązanych stronach internetowych, co dało pewną wskazówkę odnośnie do poszukiwanej lokalizacji (rysunek 11). Tym razem Bing dostaje tylko mały plus.



Rysunek 11. Analiza bengalskich napisów w wyszukiwarce Bing

Dla wyszukiwarki **Yandex** alfabety inne niż łaciński nie stanowią zazwyczaj problemu. Dlaczego zazwyczaj? Otóż podczas badania odkryłem, że nie radzi sobie ona z alfabetem bengalskim (rysunek 12). Po dalszych dociekaniach okazało się, że wyszukiwarka ta ma problemy z wieloma systemami pisma z tego rejonu świata, ponieważ nie była w stanie rozpoznać także pisma dewanagari używanego w Indiach. To niestety zaniżyło jej ocenę w tym zestawieniu.



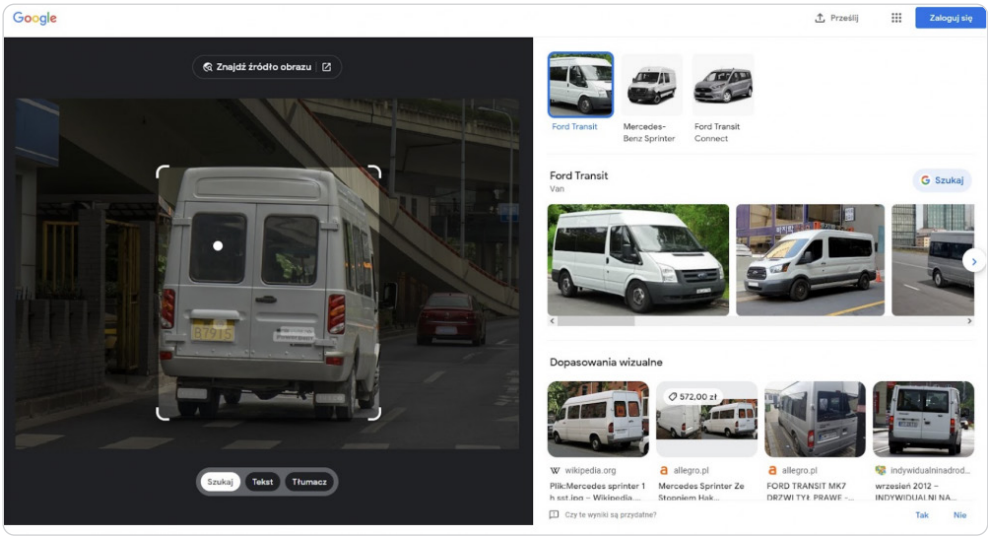
Rysunek 12. Analiza bengalskich napisów w wyszukiwarce Yandex

Samochody

W badaniu z 2021 roku z rozpoznawaniem samochodów najlepiej poradziły sobie wyszukiwarki Google i Yandex, a Bing pozostał w tyle. W tym zakresie wszystkie silniki wyszukiwania zanotowały duży progres i poprawiły swoje wyniki.

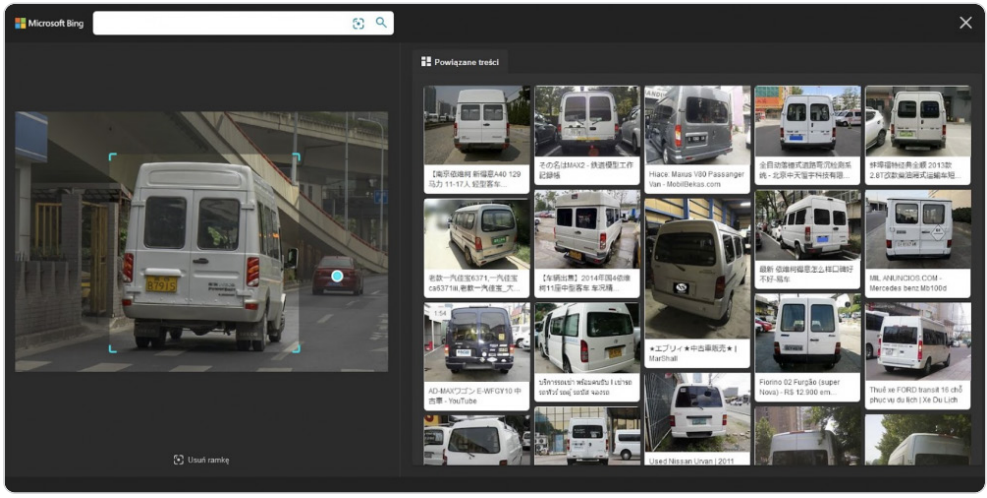
Zacznijmy standardowo od wyszukiwarki **Google**. Zdjęcie zawierające przód klasycznego samochodu zostało tym razem bezbłędnie rozpoznane, a dodatkowo w bocznym panelu wyświetliły się informacje dotyczące marki i modelu pojazdu wraz ze zdjęciami przedstawiającymi inne egzemplarze poszukiwanego samochodu. W celu dodatkowej weryfikacji umiejętności Google Lens zostały sprawdzone także na kilku innych egzemplarzach, m.in. na takich „klasykach", jak Fiat 126p, oraz na przykładzie innowacyjnej myśli brytyjskiej motoryzacji, jaki stanowi Bond Bug. W obu przypadkach wyszukiwarka nie miała problemu z rozpoznaniem marki i modelu oraz zaprezentowaniem informacji o nich. Kiedy już myślałem, że Google radzi sobie z rozpoznawaniem samochodów naprawdę dobrze, sprawdziłem jego mechanizmy na zdjęciu, na którym było widać tył busa Iveco Power Daily (rysunek 13). I tutaj niestety nastąpił moment rozczarowania, połączony z pewnym zdziwieniem, ponieważ pomimo odczytania marki na tylnych drzwiach wyszukiwarka podała, iż jest

to... Mercedes Sprinter albo Ford Transit (w zależności od kadrowania, gdyż ten zabieg potrafi zmienić wyniki, jakie otrzymamy).



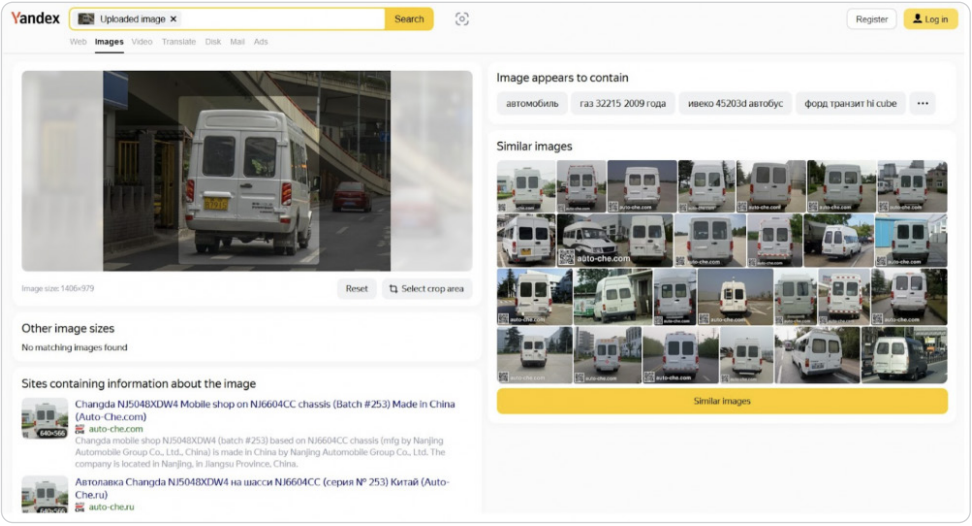
Rysunek 13. Analiza marki samochodu w wyszukiwarce Google

W wyszukiwarce **Bing** (rysunek 14) za każdym razem możliwe było odnalezienie na liście podobnych obrazów, linków do stron dotyczących danego modelu. Tylko w przypadku Fiata 126p jasno wskazane zostały marka i model. Rozpoznanie busa wyszło wyszukiwarce Microsoftu gorzej, gdy włączony został tryb wyszukiwania wizualnego, który teoretycznie powinien poprawić osiągane rezultaty.



Rysunek 14. Analiza marki samochodu w wyszukiwarce Bing

Yandex poradził sobie dość dobrze ze wszystkimi obrazami, poprawnie wskazując modele samochodów i dodatkowe informacje o nich, jednak, tak samo jak Google, poległ na przykładzie busa marki Iveco (rysunek 15). Co prawda wśród podobnych grafik znalazły się przykłady tego modelu, jednak w informacjach tekstowych na pierwszym miejscu Yandex wskazał, że może to być Gaz 32215 z 2009 roku, Iveco 45203d lub Ford Transit Hi Cube. Przy włączonym trybie wyszukiwania graficznego wyszukiwarka sugeruje, że na zdjęciu może znajdować się samochód dostawczy Mercedes.



Rysunek 15. Analiza marki samochodu w wyszukiwarce Yandex

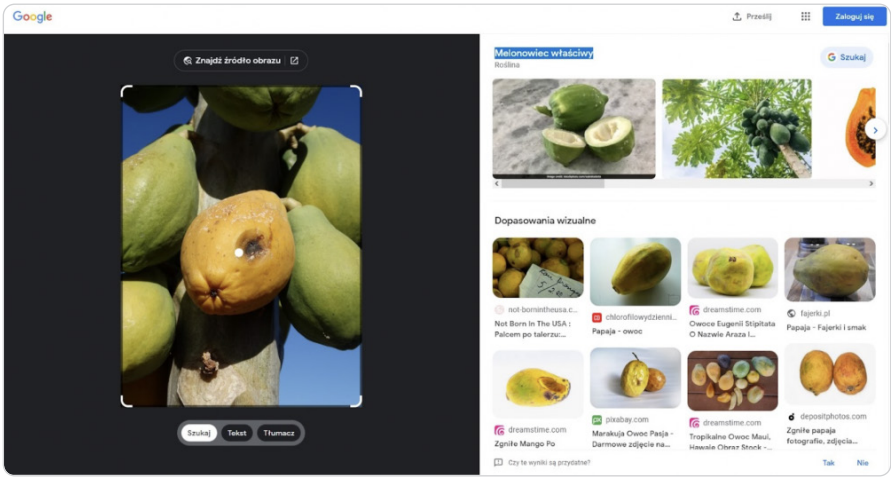
W związku z tym, że Bing nie wskazał konkretnego modelu samochodu dostawczego, nie pomylił się tak, jak Google czy Yandex. Być może takie podejście jest lepsze niż prezentowanie błędnej odpowiedzi przy niskim poziomie dopasowania?

Na pewno oznacza to, że nawet jeśli otrzymamy jakiś wynik podany „na tacy" przez wyszukiwarkę, powinniśmy go jeszcze zweryfikować, gdyż, jak widać, surowy efekt pracy samego narzędzia to jeszcze nie wszystko.

Owoce

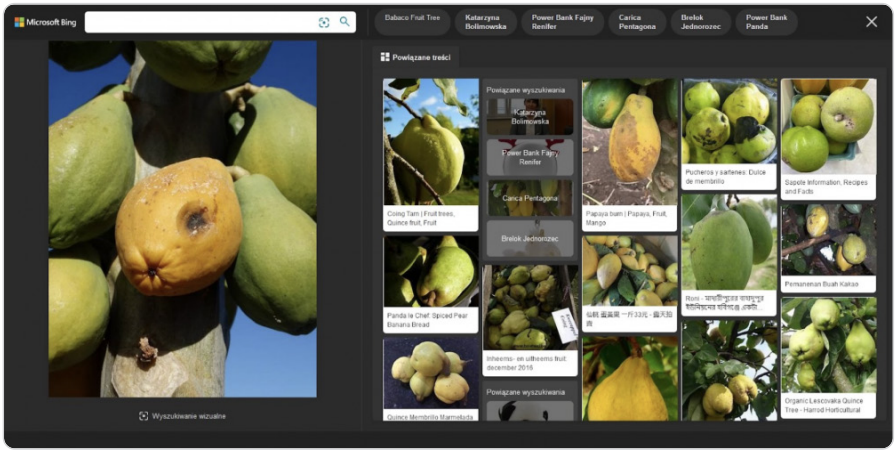
Zdjęcia owoców za pierwszym razem sprawiły trudność jedynie wyszukiwarce **Google**. Tym razem, tak jak i w innych punktach, wykorzystanie mechanizmów Google Lens znacznie poprawiło efektywność wyszukiwania, gdyż wskazane zostały nazwy owoców pokazanych na zdjęciach wraz z innymi przykładowymi zdjęciami danego owocu oraz (w niektórych przypadkach) nawet ich ceny. Wśród podobnych wizualnie obrazów znalazły się także inne, podobne owoce (rysunek 16), ale ze względu na fakt, że Google potrafi się mylić

co do obiektu przedstawionego na zdjęciu, może dobrze, że takie wyniki też się pojawiają.



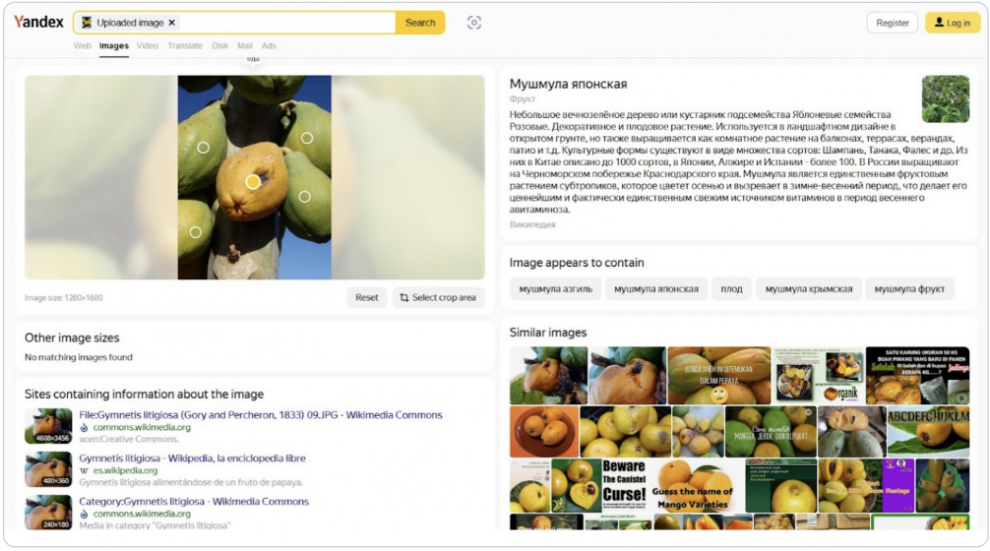
Rysunek 16. Rozpoznawanie owoców w wyszukiwarce Google

Dla **Bing** (rysunek 17) wyniki były niesatysfakcjonujące, np. w przypadku granatów powiązane strony zawierały poszukiwane owoce, a w dwóch przypadkach odnalezione zostały strony konkretnie ze zdjęciem, za pomocą którego wyszukiwano. Kiedy na zdjęciu znajdowała się papaja, wyniki nie były już takie dobre i Bing, oprócz jednego zdjęcia papai, wskazywał całą serię różnych owoców, podpowiadając przy okazji hasła takie jak: „babaco fruit tree" (blisko), „Power Bank Fajny Renifer" lub „Brelok jednorożec" (chyba dość daleko). Włączenie trybu wyszukiwania wizualnego praktycznie nic nie zmieniło.



Rysunek 17. Rozpoznawanie owoców w wyszukiwarce Bing

Yandex (rysunek 18) co prawda rozpoznał granaty, jednak nie wyświetlił w bocznym panelu informacji o tych owocach, a jedynie wskazał, że obrazy mogą je zawierać. W przypadku zdjęć papai taka szczegółowa informacja już została wyświetlona, jednak wystąpiła jedna zastanawiająca sytuacja. Wyszukiwarka Yandex dla rozpoznanych elementów obrazu wskazuje możliwość wyszukania informacji o nich. Niekiedy jest to informacja ogólna (np. budynek lub samochód), a niekiedy konkretne rodzaje przedmiotów (w tym przypadku był to konkretny owoc). Niestety, pomimo pierwotnego poprawnego rozpoznania owocu, kliknięcie w okrągły przycisk na samym zdjęciu, oznaczający wyszukiwanie informacji o danym obiekcie, przyniosło w efekcie wyświetlenie informacji o… niespliku japońskim. Szybka analiza wykazała wprawdzie pewne pokrewieństwo tej rośliny z papają, ale zbyt dalekie, by uznać wynik wyszukania za poprawny. Po raz kolejny więc doszło do sytuacji, kiedy wyszukiwarka dla tego samego zdjęcia wskazała różne wyniki.

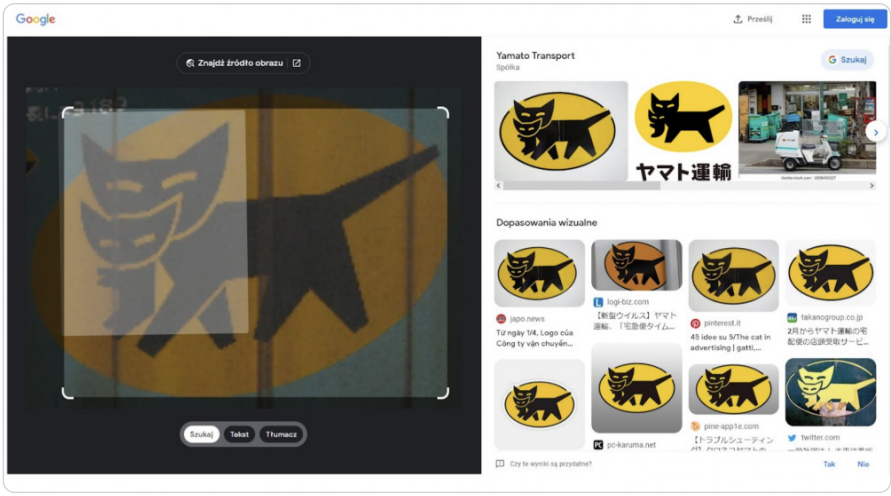


Rysunek 18. Rozpoznawanie owoców w wyszukiwarce Yandex

Logo

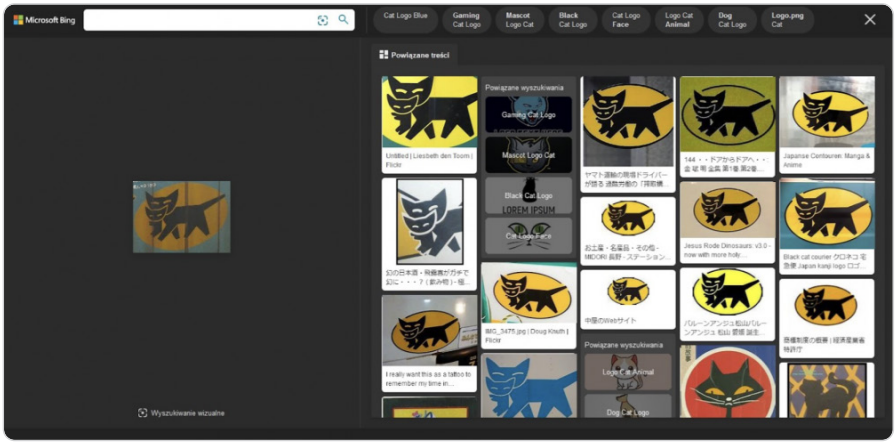
W rozpoznawaniu znaków graficznych znanych marek lub wydarzeń w poprzednim badaniu najgorzej zaprezentował się Bing, mający problemy z rozpoznawaniem tego typu obrazów. Google i Yandex wypadły podobnie do siebie i nieco lepiej od Binga.

W nowej odsłonie wyszukiwarka **Google** (rysunek 19) poradziła sobie najlepiej ze wszystkich. Bez problemu rozpoznała zarówno znak japońskiej firmy kurierskiej, igrzysk w Pekinie, jak i jednej z polskich cukierni (w formie z nazwą i bez niej). Nawet słaba jakość obrazu nie uniemożliwiła odczytania nazwy ze zdjęcia.



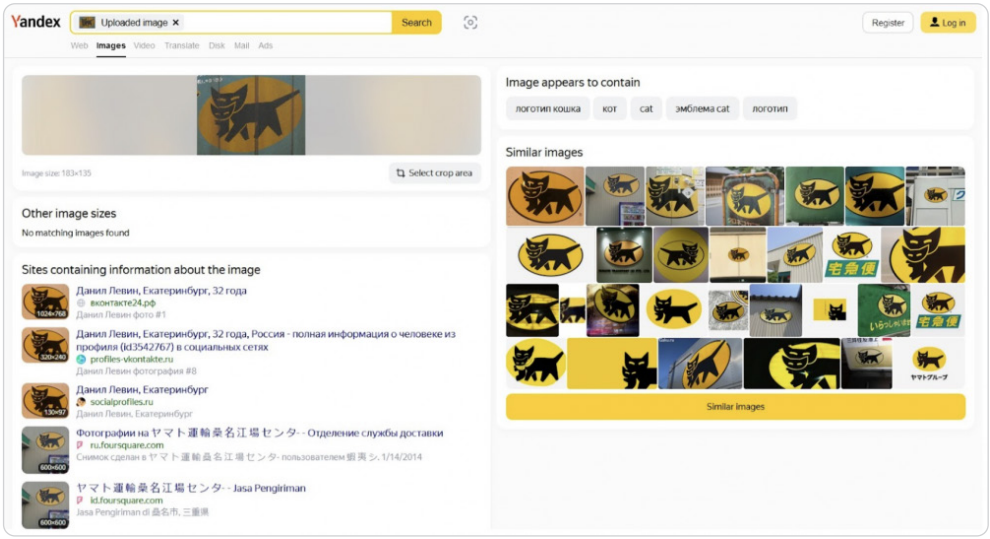
Rysunek 19. Rozpoznawanie logotypu w wyszukiwarce Google

Bing (rysunek 20) radził sobie dużo lepiej niż poprzednim razem, ale duża liczba błędnych wyników dla logo igrzysk, a także brak wyszukania powiązanych treści – nawet pomimo odczytania tekstu ze zdjęcia szyldu cukierni – sprawiły, że wynik tego badania w wykonaniu wyszukiwarki Microsoftu można uznać za niesatysfakcjonujący.



Rysunek 20. Rozpoznawanie logotypu w wyszukiwarce Bing

Dla wyszukiwarki **Yandex** (rysunek 21) nie stanowiło problemu odnalezienie stron z logo cukierni i igrzysk, ale w przypadku szyldu odczytanie napisu ją przerosło i nie była w stanie wskazać dobrego wyniku. Logo firmy kurierskiej zostało za to poprawnie dopasowane do innych obrazów z sieci.

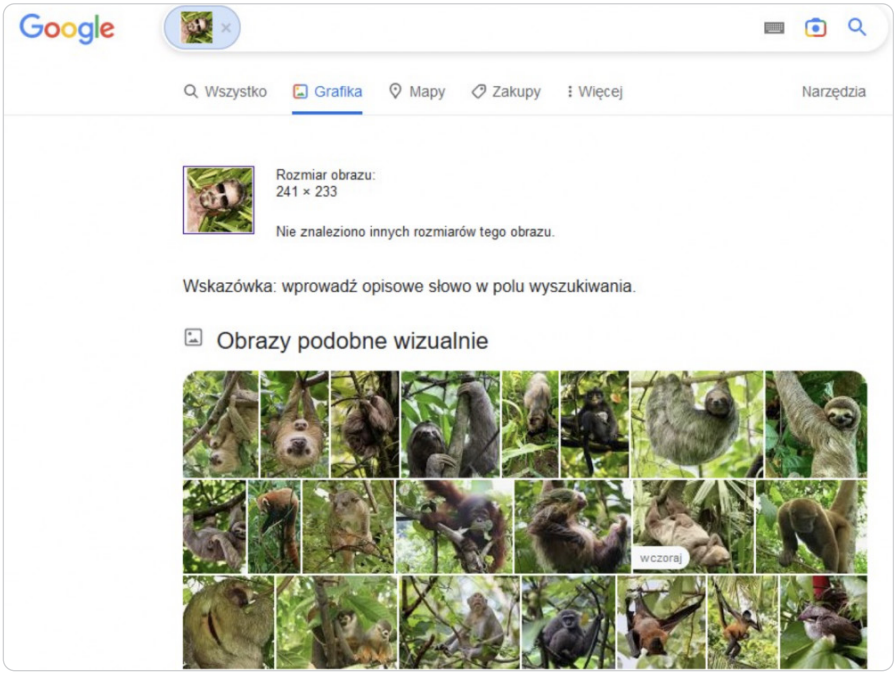


Rysunek 21. Rozpoznawanie logotypu w wyszukiwarce Yandex

Twarze

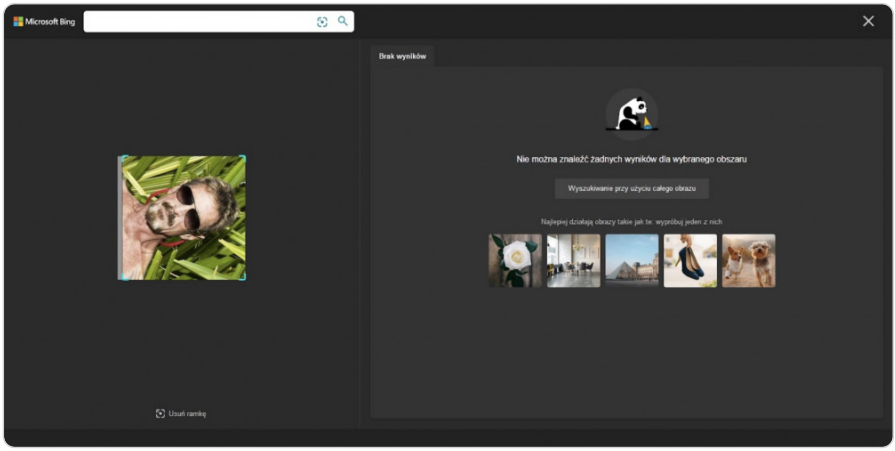
Ostatnim elementem badania było rozpoznawanie twarzy – i tutaj w badaniu z 2021 roku nastąpiło największe zaskoczenie, gdy Bing zwyciężył w tej kategorii, z dużą przewagą nad konkurencją.

Narzędzie **Google Lens** przyzwyczaiło użytkowników do dobrych wyników, jednak w zakresie rozpoznawania twarzy poległo. To zdecydowanie najsłabsza strona tego mechanizmu. Przy wyszukiwaniu zdjęciem Johna McAfee'ego przekonwertowanym do skali szarości Google Lens pochwaliło się, że znalazło na zdjęciu… okulary! Dopiero powrót do tradycyjnej **Grafiki Google** (Google Images) uratował honor tej wyszukiwarki, gdyż wskazana została prawidłowa osoba. Niestety, w przypadku obrócenia zdjęcia o 90 stopni było tylko gorzej – wyszukiwanie wizualne nadal twierdziło, że widzi tylko okulary, a tradycyjny silnik wyszukiwania tym razem zobaczył głównie… leniwce oraz różne gatunki małp i małpiatek (rysunek 22), a także pandę małą. Podobnie było z twarzami dwóch szwajcarskich skoczków narciarskich. Wizualnie byli oni (według wyszukiwarki Google) podobni do bluz sportowych, a Grafika Google nie mogła się zdecydować, czy na zdjęciu znajduje się ktoś z Uniwersytetu w Miami, czy piłkarz ręczny. Wyszukiwanie stockowego obrazu dziewczyny przyniosło zaskakująco mało wyników jak na tego typu grafikę.



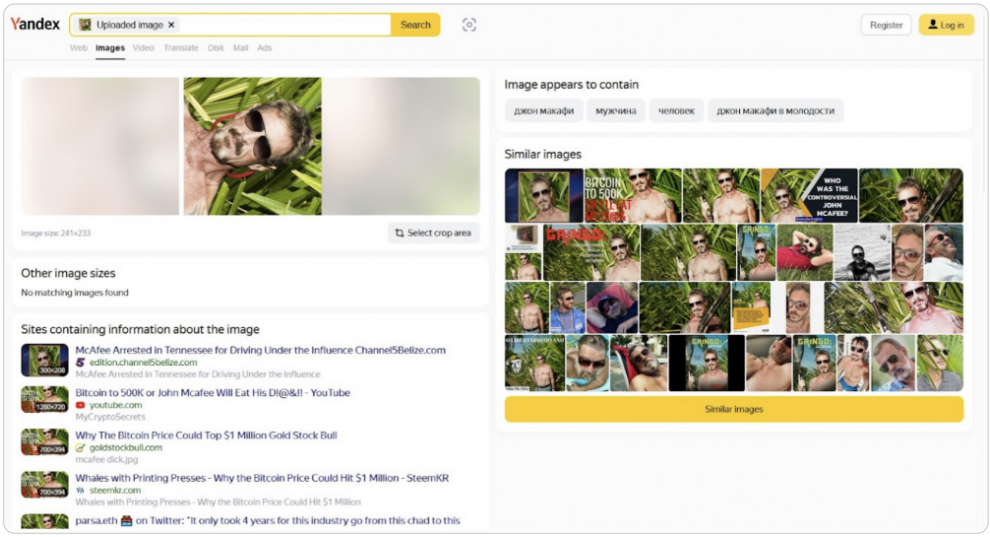
Rysunek 22. Rozpoznawanie twarzy w wyszukiwarce Google Grafika

Bing po raz kolejny dobrze poradził sobie z wizerunkami skoczków i jako jedyny wskazał z imienia i nazwiska, kto znajduje się na zdjęciu, a także rozpoznał, że na obrazie w skali szarości widnieje John McAfee (rysunek 23). Co do stockowego zdjęcia dziewczyny: wskazał więcej stron z wyszukiwaną grafiką niż Google. Najgorzej jednak wypadło wyszukiwanie zdjęciem obróconym o 90 stopni – tutaj wyszukiwarka się poddała i nie wskazała żadnych wyników.



Rysunek 23. Rozpoznawanie twarzy w wyszukiwarce Bing

Yandex nie rozpoznał wprawdzie żadnego ze skoczków, wskazując w obu przypadkach, że na zdjęciu mogą znajdować się osoby o nazwisku Schumacher, jednak był w stanie wyszukać dużą liczbę powiązanych stron, na których wykorzystano stockowe zdjęcie dziewczyny. Także wyszukiwanie zdjęciem Johna McAfee'ego, zarówno w skali szarości, jak i w przypadku obrócenia go o 90 stopni, dało poprawny wynik. Z tym ostatnim zadaniem Yandex poradził sobie jako jedyny (rysunek 24).



Rysunek 24. Rozpoznawanie twarzy w wyszukiwarce Yandex

W zakresie rozpoznawania twarzy dużo bardziej skuteczne będą chociażby takie narzędzia, jak PimEyes czy search4faces, tym bardziej jeśli np. zostaną użyte w połączeniu z narzędziami AI do postarzania twarzy, jak można było to zobaczyć w przypadku osób poszukiwanych od 30 lat w Niemczech.

Podsumowanie i rady

Aby podsumować badanie i przedstawić graficznie, jak duże zmiany zaszły w popularnych wyszukiwarkach graficznych w ciągu ostatnich niemal dwóch lat, musiałem nieco zmienić skalę ocen.

Za pierwszym razem brak umiejętności poprawnego wyszukania oznaczałem znakiem minus (–), poprawne wyszukanie – znakiem plus (+), a wyszukanie pełne, ze wskazaniem szczegółów, oznaczone było podwójnym plusem (++).

Tym razem, aby zwiększyć gradację, posłużyłem się skalą od 1 do 5, przy czym 1 oznacza brak umiejętności wyszukania, 3 – poprawne wyszukanie, a 5 – informację pełną, podaną ze szczegółami. Nie jest to rozwiązanie idealne, ale pozwala odnieść wyniki z 2023 roku do tych z 2021 roku (tabela 1).

Tabela 1. Różnice w wynikach działania wyszukiwarek pomiędzy badaniami z 2021 i 2023 roku

	GOOGLE			BING			YANDEX		
	2021	2023	ZMIANA	2021	2023	ZMIANA	2021	2023	ZMIANA
Miejsca	3	5	+2	1	3	+2	3	5	+2
Tekst	1	5	+4	3	2	-1	5	4	-1
Samochody	3	4	+1	1	3	+2	5	4	-1
Owoce	1	5	+4	3	3	0	3	4	+1
Logo	3	4	+1	1	3	+2	3	3	0
Twarze	1	2	+1	5	4	-1	3	4	+1
Suma	**12**	**25**	**+13**	**14**	**18**	**+4**	**22**	**24**	**+2**

Podsumowując wyszukiwanie za pomocą obrazów przy użyciu omawianych tu wyszukiwarek, warto przytoczyć ranking porównawczy wyszukiwarek oferujących funkcję wyszukiwania na podstawie zawartości obrazu (ang. *reverse image search*) (rysunek 25).

Rysunek 25. Ranking porównawczy wyszukiwarek oferujących funkcję *reverse image search* (im więcej kropek, tym lepszy wynik). Materiały własne

Jak widać, wdrożenie mechanizmów Google Lens do wyszukiwarki grafiki Google było dobrą decyzją. Dzięki tym usprawnieniom wyszukiwarka Google z ostatniego miejsca w poprzedniej klasyfikacji przeskoczyła na pierwsze. Nie jest to oczywiście rozwiązanie idealne, gdyż istnieje jeszcze sporo obszarów do poprawy, np. w zakresie rozpoznawania osób, ale skok jakości wyszukiwania jest naprawdę znaczący. W tyle została wyszukiwarka Bing, która cały czas najlepiej radzi sobie z rozpoznawaniem twarzy, jednak pozostałe wyniki są co najwyżej zadowalające lub średnie (ocena 4 lub 3). Yandex w stosunku do 2021 roku stracił najwięcej. Tracąc pozycję wiodącej wyszukiwarki graficznej, pozostał nadal do-