



Wyszukiwanie informacji w Internecie

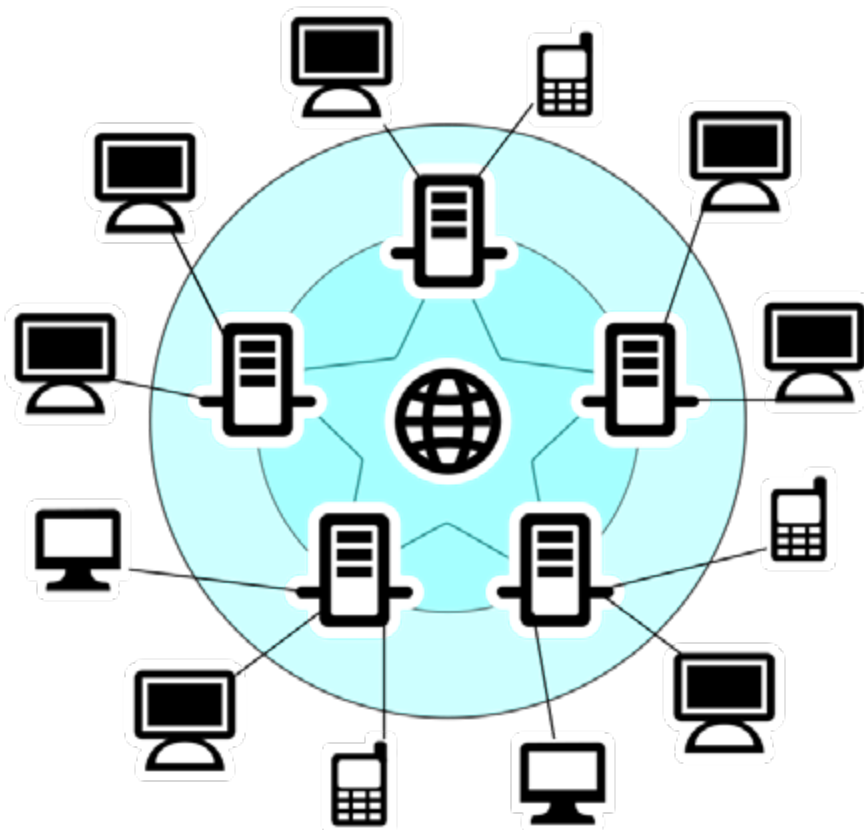
TEMAT 2: WYSZUKIWANIE INFORMACJI W INTERNECIE

2.2.2.1 Wprowadzenie do Internetu

Internet to rzeczywistość, która powstała w naszym życiu w stosunkowo niedawno. W tej sekcji zostanie omówionych kilka wstępnych koncepcji Internetu.

2.2.2.1.1 Co to jest Internet

Pojęcie Internet pochodzi od słów Sieć połączona (**Inter**connected **Network**). Oznacza to etymologicznie, że jest to sieć sieci, czyli wiele mniejszych sieci połączonych ze sobą, tworzących większą, Internet. Innymi słowy, każdy komputer lub inne urządzenie (np. Drukarka, telefon komórkowy, tablety, a ostatnio telewizory, klimatyzacja, zegarki itp.) podłączone do Internetu są w zasadzie jego częścią. Połączenie jest ustanawiane bezprzewodowo (tj. za pośrednictwem anten i satelitów) lub przewodowe (np. przez sieć telefoniczną). Internet powstał jako projekt wojskowy w USA w latach 60. XX wieku.



Rysunek 1. Internet to duża sieć składająca się z mniejszych sieci i urządzeń połączonych ze sobą.

2.2.2.1.2 Usługi internetowe

Internet oferuje różne usługi, takie jak World Wide Web (strony internetowe), usługi komunikacyjne (poczta elektroniczna, Telnet, VoIP itp.), FTP (protokół przesyłania plików) i IRC (Internet Relay Chat).

Należy zauważyć, że sieć WWW to tylko usługa internetowa. Jednak okazało się, że jest to usługa dominująca i opracowano na niej wiele aplikacji, aby pomieścić inne usługi (np. E-mail i FTP), wiele osób błędnie postrzega World Wide Web jako identyczny z Internetem.

2.2.2.1.3 Strony internetowe w sieci WWW

Sieć WWW została wynaleziona przez Tima Bernersa Lee w 1989 roku i składa się z dokumentów i innych plików (np. obrazów i filmów) połączonych ze sobą za pomocą hiperłączy. Te dokumenty i pliki znajdują się na komputerach podłączonych do Internetu. Strony internetowe to dokumenty napisane w określonym formacie opartym na języku HTML (Hypertext Markup Language). Pierwsza strona została opublikowana dla ogółu społeczeństwa w sierpniu 1991 roku.

Aby wyświetlić stronę internetową, potrzebujemy specjalnego programu komputerowego o nazwie przeglądarka internetowa i adresu strony internetowej. Przeglądarka używa tego adresu do jego zlokalizowania. Przykładami przeglądarek internetowych są Mozilla Firefox, Google Chrome, Opera, Safari i Edge.

2.2.2.1.4 Adresy w sieci WWW

Adres strony internetowej lub pliku w sieci WWW, który można również zobaczyć jako URL (Uniform Resource Locator), ma następujący format:

http://www.sitename.extension

Http to protokół, który czasem można zobaczyć jako *https* (s dla bezpieczeństwa). Możemy pominąć *http://*, a przeglądarka internetowa rozumie, że należy użyć protokołu HTTP.

Następną częścią jest *www* i czasami można go również pominąć, jeśli administrator strony dokonał odpowiedniego ustawienia.

Część *sitename.extension* to nazwa domeny. Rozszerzenie to standardowe słowo o nazwie Domena Najwyższego Poziomu (TLD), które deklaruje - ale niekoniecznie - typ lub lokalizację geograficzną witryny. Na przykład rozszerzenie *com* jest przeznaczone do celów komercyjnych, *org* dla organizacji, *bu* dla Bułgarii, *cy* dla Cypru, *gr* dla Grecji, *it* dla Włoch, *pl* dla Polski itd. Możesz zobaczyć pełną listę dostępnych TLD na <https://www.icann.org/resources/pages/tlds-2012-02-25-en>

Nazwa witryny to nazwa wybrana przez jej właściciela i została zarejestrowana w oficjalnej organizacji (innej w każdym kraju) na podstawie określonej TLD. Zauważ, że

sitename.com różni się od *sitename.org* i prowadzi do różnych stron internetowych. Niemniej jednak, jeśli właściciel witryny zarejestrował obie domeny, może skierować obie domeny do tej samej strony.

2.2.2.1.5 Adresy E-mail

Poczta elektroniczna lub e-mail to usługa internetowa, za pomocą której możemy wymieniać listy elektroniczne. Ktoś, kto wyśle wiadomość do innej osoby za pomocą e-maila, powinien mieć adres e-mail i wysłać go na inny adres e-mail, adres odbiorcy. Format adresu e-mail to:

user_selected_name@email_provider.extension

Część *user_selected_name* użytkownik może wybrać sam. Może to być pseudonim lub jego imię/nazwisko, po którym zawsze występuje znak @, który jest wymawiany jako „at”.

Część *email_provider.extension* jest nazwana domeną. Rozszerzenie to TLD (przypomnienie z akapitu o adresach w sieci World Wide Web).

Ważne jest, aby zauważyć, że dwa adresy e-mail, które mają tę samą *user_selected_name*, są całkowicie różne, jeśli znajdują się w innej domenie. Przykładowo johnsmith@hotmail.com jest innym adresem niż johnsmith@gmail.com

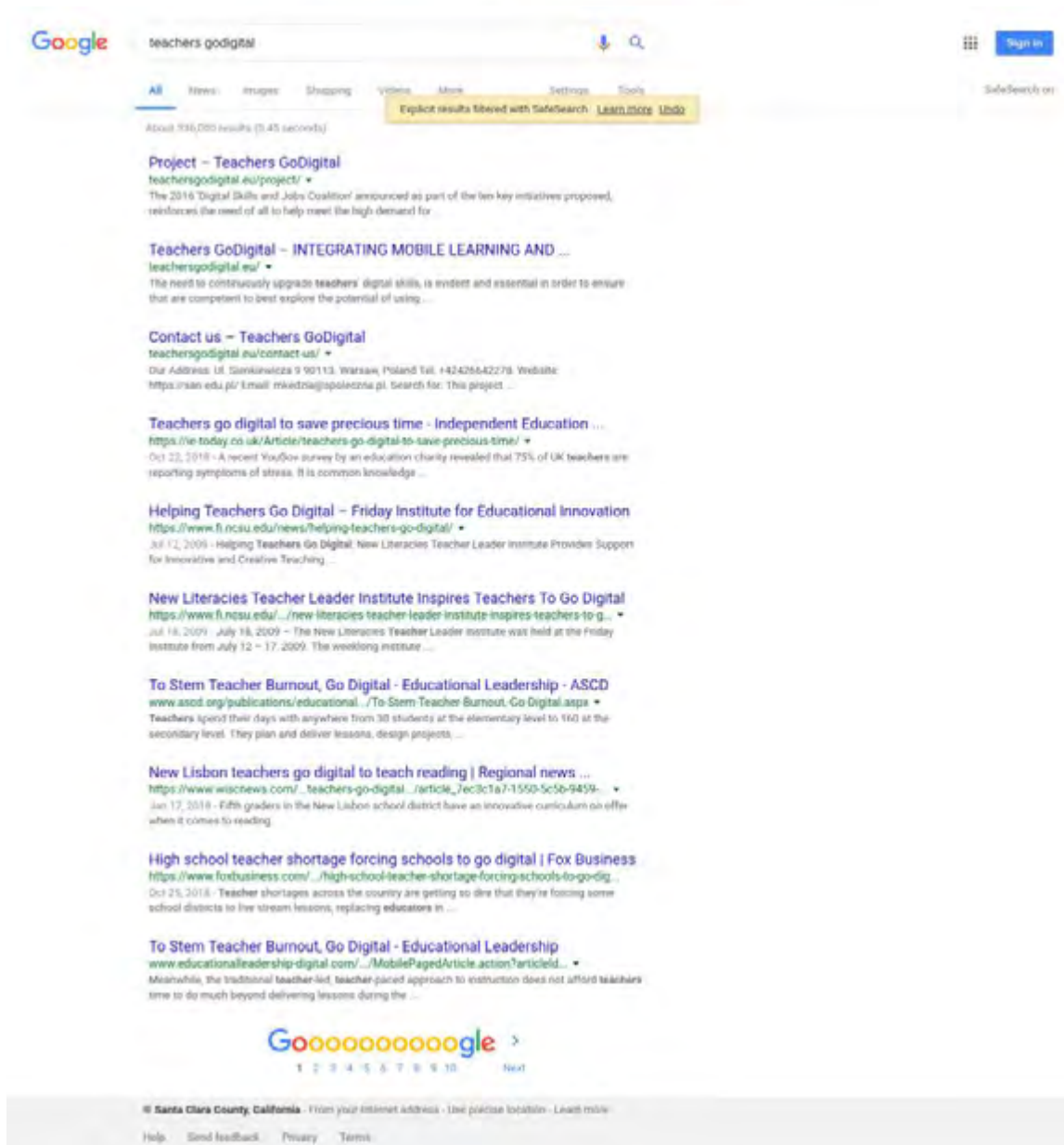
Zauważ też, że johnsmith.hotmail.com to adres strony internetowej, a nie adres e-mail. Adres e-mail powinien zawsze zawierać znak @.

2.2.2.2 Wyszukiwanie informacji

World Wide Web to ocean informacji. Każdy może publikować informacje w różnych formach, tj. teksty, obrazy, filmy, mapy itp. Ogromna ilość opublikowanych informacji oraz fakt, że nie mamy kontroli nad ich aktualnością, stwarza odpowiednio wyzwania, jak skutecznie wyszukiwać informacje w sieci i jak upewnić się, że odkryte informacje są prawdziwe, a nie fałszywe lub niedokładne.

2.2.2.2.1 Co to jest wyszukiwarka internetowa

Wyszukiwarka internetowa jest naszym sprzymierzeńcem, aby zmierzyć się z wyzwaniem skutecznego wyszukiwania informacji w sieci WWW. W swoim zapleczu znajduje się program komputerowy złożony ze skomplikowanych algorytmów, który przechowuje bazę danych ze szczegółami na temat zawartości stron internetowych. Interfejs to formularz internetowy, za pomocą którego użytkownik może wprowadzać słowa kluczowe lub wyrażenia, które chce wyszukać. Wyszukiwarka przeszukuje te słowa kluczowe i frazy w swojej bazie danych i tworzy ranking na podstawie trafności wyników. Wyniki są przedstawiane użytkownikowi w formie stronicowania. Rys. 2 to zrzut ekranu z wynikami wyszukiwania w wyszukiwarce Google po użyciu słów kluczowych *nauczyciele godigital*.



Rysunek 2: Wyniki wyszukiwania ze słowami kluczowymi nauczyciele godigital w wyszukiwarce Google.

2.2.2.2.2 Kategorie wyszukiwarek internetowych

Możemy wyróżnić następujące kategorie wyszukiwarek internetowych:

- *Wyszukiwarki ogólnego przeznaczenia*

Wyszukiwarka ogólnego przeznaczenia wyszukuje i pobiera informacje z każdej strony internetowej zarejestrowanej w jej bazie danych. Strona internetowa jest rejestrowana w bazie danych, gdy właściciel witryny rejestruje ją samodzielnie lub gdy specjalny program komputerowy wyszukiwarki (zwany przeszukiwaczem lub botem) odkryje

stronę internetową, która nie jest jeszcze zarejestrowana w tej bazie danych.

Przykładami wyszukiwarek ogólnego przeznaczenia są:

- a. www.google.com
- b. www.duckduckgo.com
- c. www.yandex.com
- d. www.bing.com
- e. www.yahoo.com
- f. www.ask.com
- g. www.aol.com
- h. www.baidu.com

- *Wyszukiwarki specjalnego przeznaczenia*

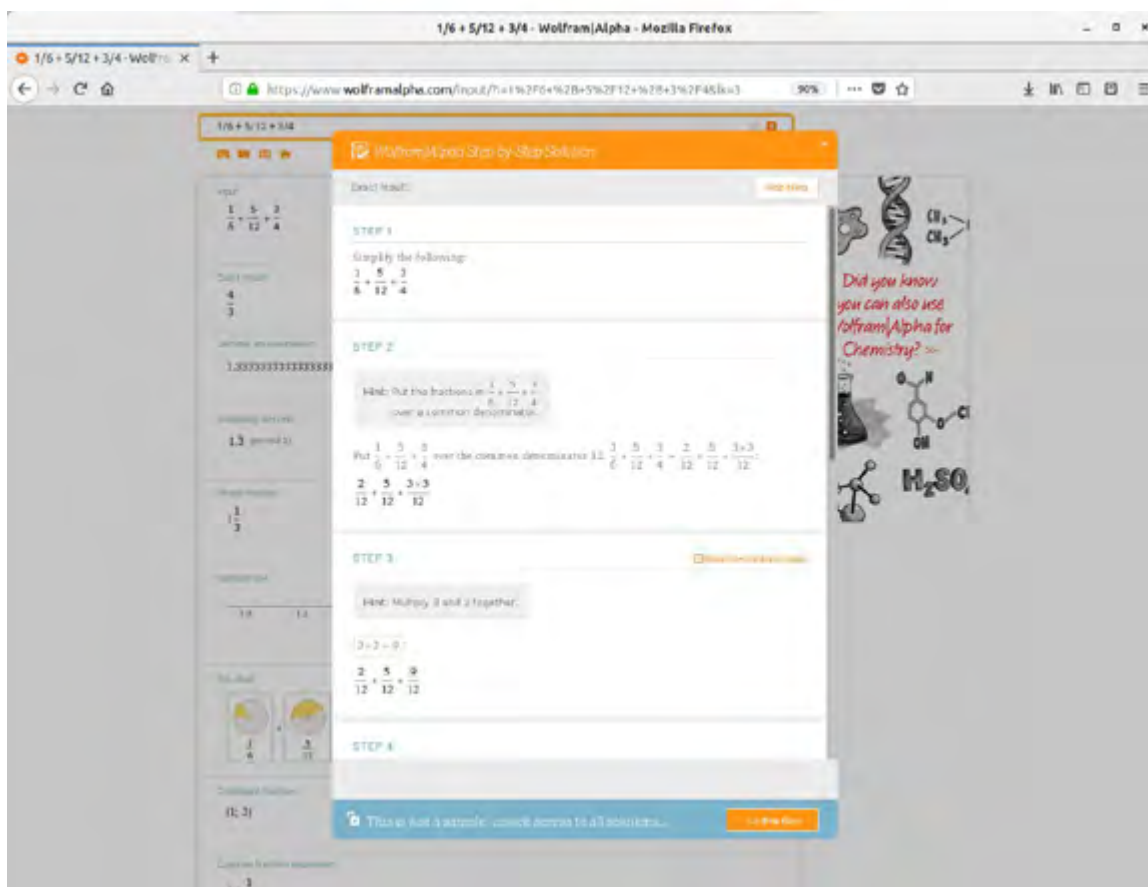
Wyszukiwarka specjalnego przeznaczenia to wyszukiwarka, w której można kontrolować zawartość przechowywaną w jej bazie danych. Przykładem jest wyszukiwarka zawarta w określonej witrynie, na przykład jak wikipedia.com. Wikipedia.com jest encyklopedią online, dlatego jej treść powinna być specyficzna dla tematów, które może zawierać encyklopedia. Na przykład nie znajdziemy informacji dotyczących prognozy pogody na następny dzień. Z drugiej strony strona internetowa z prognozami pogody ma wyszukiwarkę do wyszukiwania prognoz pogody dla określonego obszaru lub danych historycznych na temat pogody na danym obszarze.

- *Metawyszukiwarki*

Metawyszukiwarka to taka, która gromadzi wyniki wyszukiwania z wielu wyszukiwarek i przedstawia je użytkownikowi. Na przykład skyscanner.com to metawyszukiwarka, która przeszukuje wyszukiwarki firm turystycznych, zbiera ich odpowiedzi, a następnie przedstawia je użytkownikowi.

- *Wyszukiwarka wiedzy obliczeniowej*

Wyszukiwarka wiedzy obliczeniowej nie przedstawia użytkownikowi wyników wyszukiwania, ale oblicza odpowiedź wyszukiwania użytkownika lub pobiera odpowiedź z wybranych źródeł. Może to być bardzo przydatne dla nauczycieli i uczniów. Wolframalpha.com to taka wyszukiwarka wiedzy obliczeniowej. Rysunek 3 to zrzut ekranu przedstawiający proces krok po kroku działania $1/6 + 5/12 + 3/4$, natomiast rys. 4 to zrzut ekranu wyszukiwania słynnego greckiego pisarza Nikosa Kazantzakisa.



Rysunek 3: Wynik wyszukiwania krok po kroku działania $1/6 + 5/12 + 3/4$ w WolframAlpha.com Wyszukiwarce Wiedzy Obliczeniowej.

2.2.2.2.3 Proste wyszukiwanie

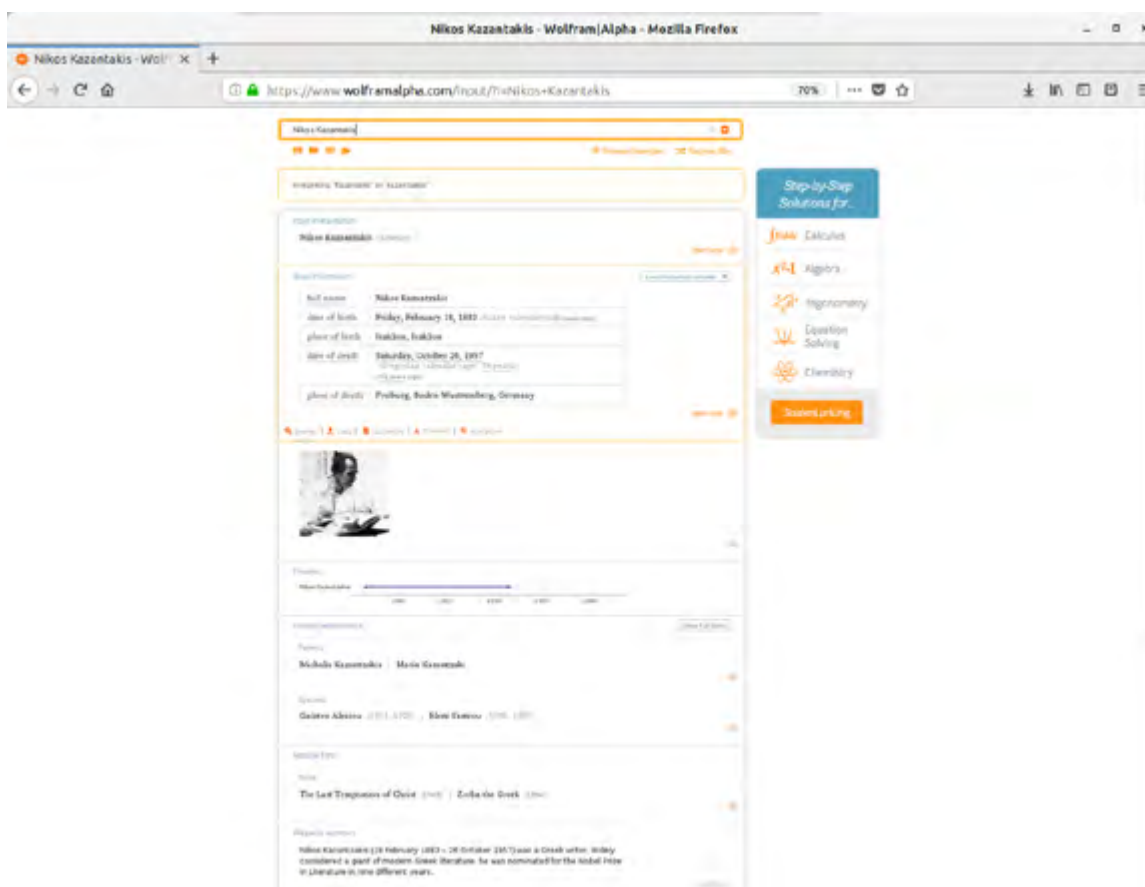
Aby zobaczyć w praktyce, jak możemy wyszukiwać informacje w sieci WWW, założmy, że chcemy wyszukiwać informacje o Projekcie GoDigital, którego produktem jest obecny tekst.

Pierwszym krokiem jest otwarcie przeglądarki internetowej (np. Mozilla Firefox, Google Chrome, Opera itp.) i paska adresu (patrz rysunek 5) w celu wpisania adresu internetowego wyszukiwarki, której chcemy używać (przypomnijmy adresy wyszukiwarka ogólnego przeznaczenia, o której wspomnieliśmy w punkcie 2.2.2.2). Skorzystajmy z wyszukiwarki duckduckgo.com. Podstawą filozofii wyszukiwarki duckduckgo.com jest to, że nie śledzi ona naszych wyszukiwań i chroni nasze dane osobowe (więcej informacji można znaleźć w duckduckgo.com/spread).

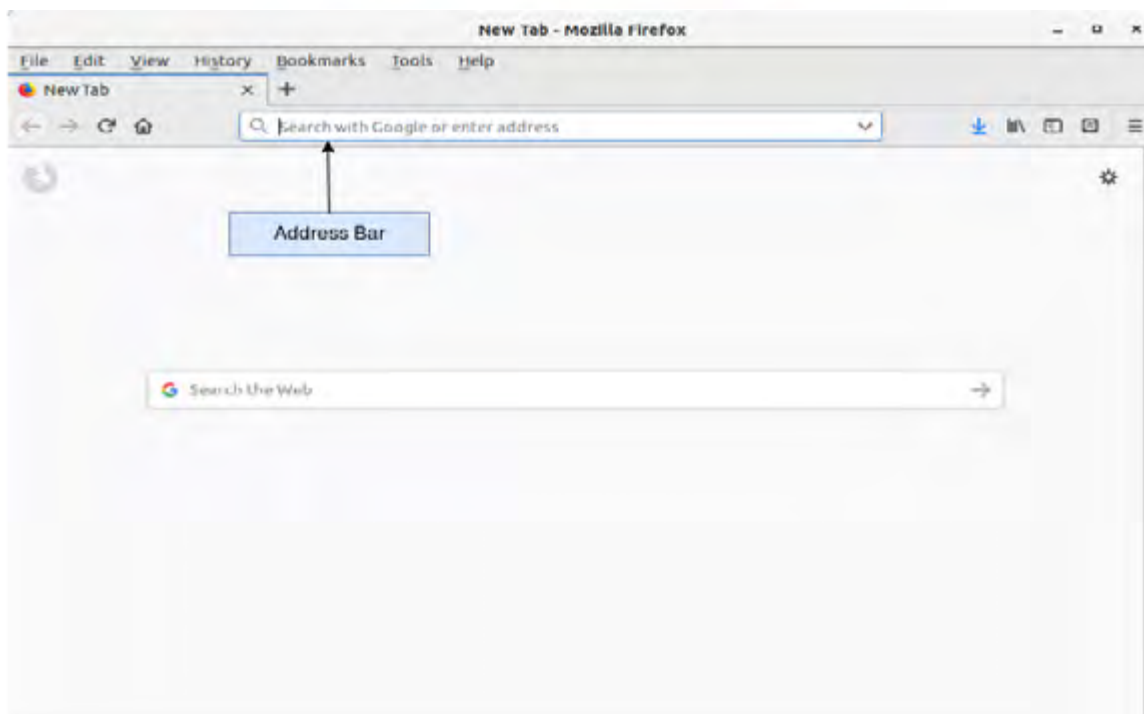
Drugim krokiem jest wstawienie słów kluczowych w celu przeprowadzenia wyszukiwania. W każdej wyszukiwarce znajduje się pole tekstowe (zwykle na środku strony), w którym wstawiamy słowa kluczowe naszego wyszukiwania. Bardzo ważne jest stosowanie poprawnych słów kluczowych, aby uzyskać najbardziej trafne wyniki. Do tego wyszukiwania

użyjemy słów kluczowych *godigital* i *project*, które później pokażemy, że nie jest to najlepszy wybór. Zrzut ekranu z tego wyszukiwania pokazano na rys. 6, a jego wyniki na rys. 7.

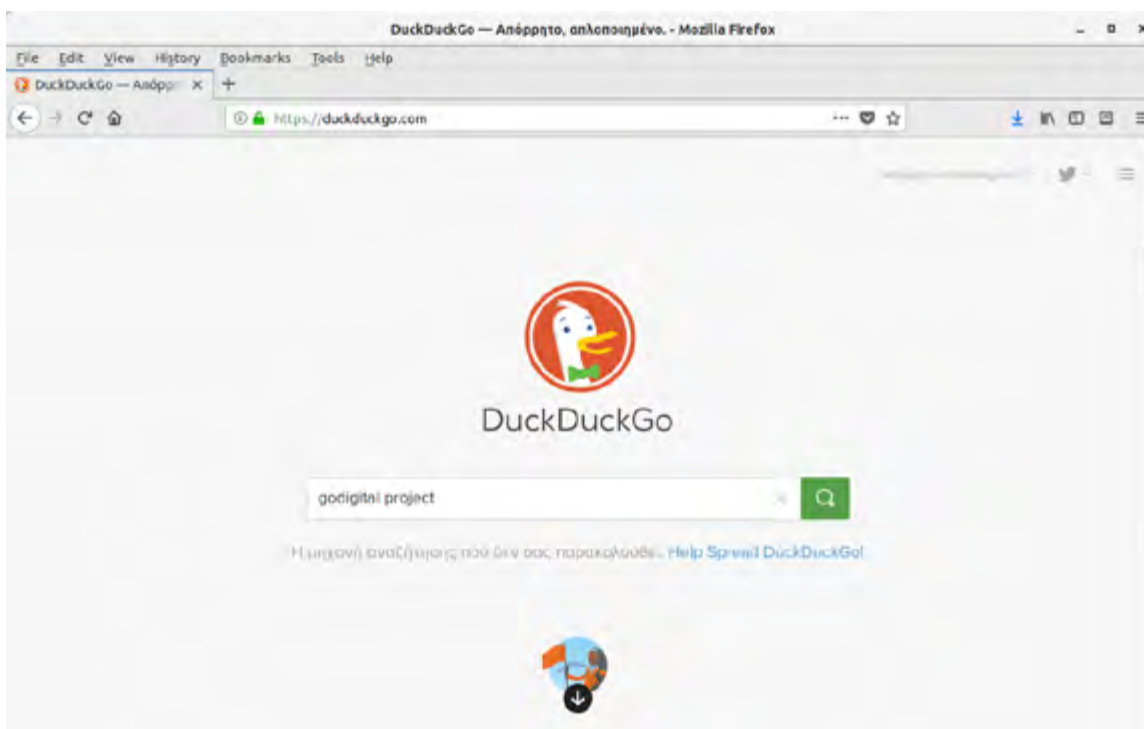
Należy również zauważyć, że zamiast odwiedzać stronę internetową wyszukiwarki w celu wykonania żądanego wyszukiwania, możemy alternatywnie wstawić słowa kluczowe wyszukiwania bezpośrednio w pasku adresu przeglądarki (patrz rysunek 9). Przeglądarka użyje predefiniowanej wyszukiwarki (zwykle google.com) i przekieruje nas na stronie wyników. Wyszukiwarke, z której domyślnie korzysta przeglądarka, można zmienić w sekcji Ustawienia / Preferencje przeglądarki.



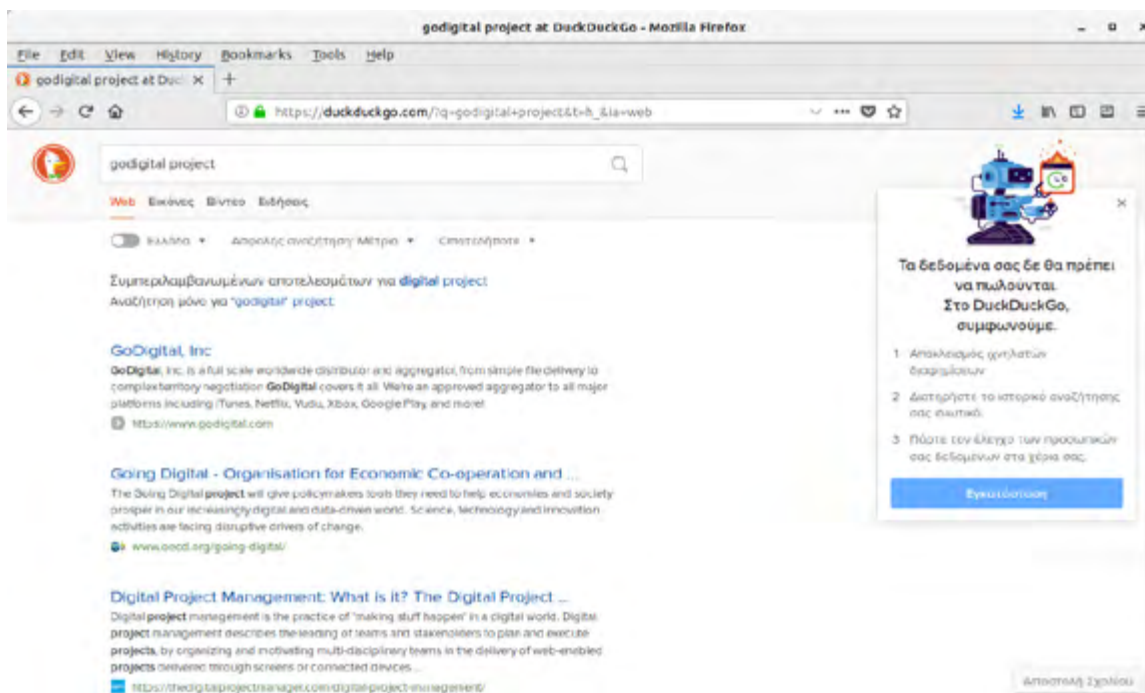
Rysunek 4: Wyszukanie słynnego greckiego pisarza Nikosa Kazantzakisa w wyszukiwarce obliczeniowej Wolframalpha.com.



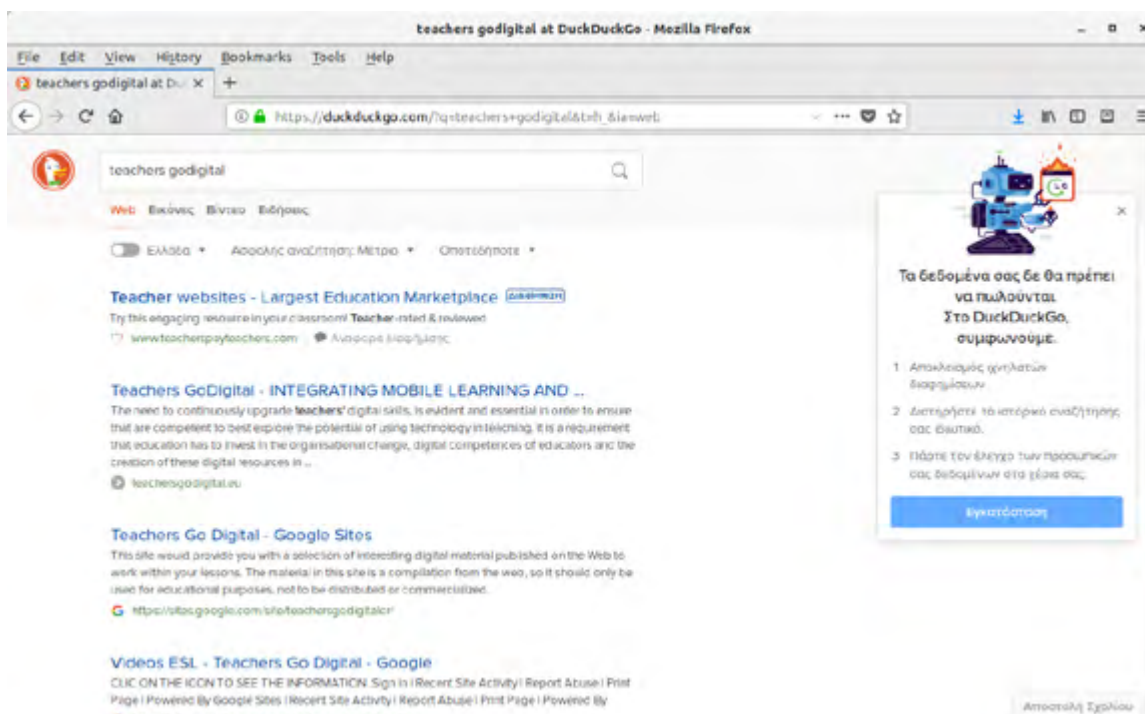
Rysunek 5: Pasek adresu w przeglądarce internetowej to pole tekstowe (zwykle w górnej części okna przeglądarki), w którym wpisujemy adres strony internetowej, którą chcemy odwiedzić.



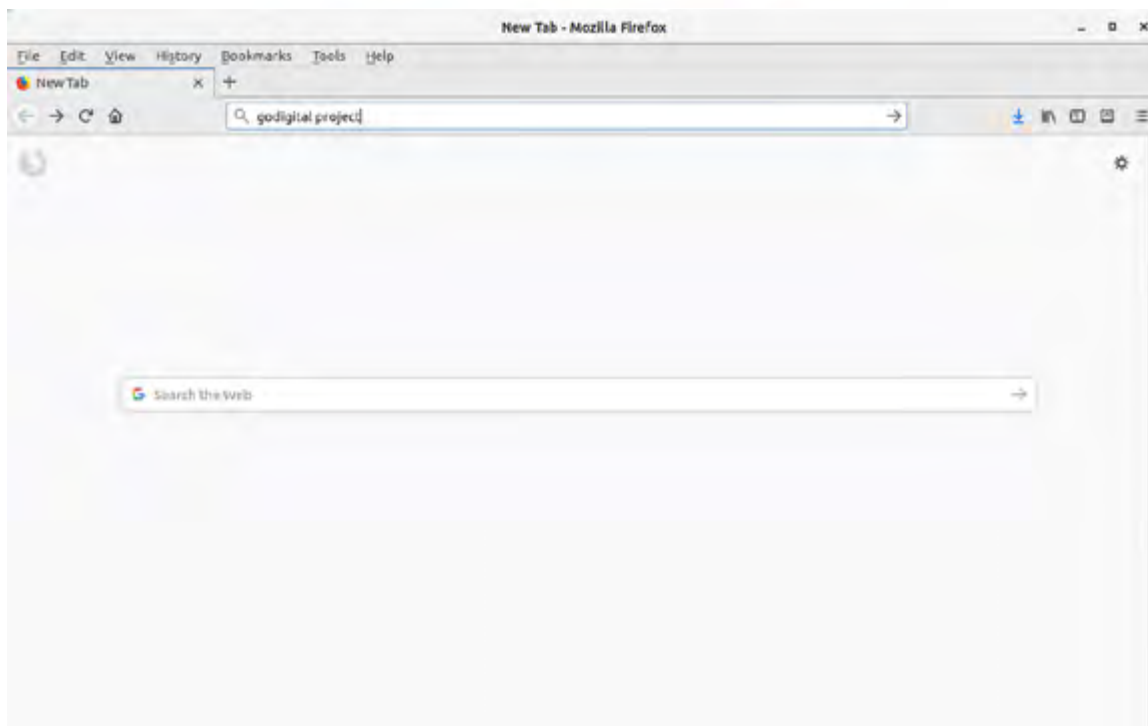
Rysunek 6: Wyszukanie w wyszukiwarce DuckDuckGo przy pomocy słów kluczowych godigital projekt



Rysunek 7: Wyniki wyszukiwania pokazane na rysunku 6



Rysunek 8: Zmiana słów kluczowych w wyszukiwaniu różnicuje wyniki wyszukiwania (patrz także rysunki 6 i 7)



Rysunek 9: Możemy przeprowadzić wyszukiwanie, wstawiając słowa kluczowe w pasku adresu przeglądarki internetowej. Przeglądarka internetowa wykorzystata predefiniowaną wyszukiwarkę, którą możemy zmienić z poziomu opcji ustawień / preferencji przeglądarki.

2.2.2.2.4 Jak działa wyszukiwarka

Wyszukiwarka to złożony program wykorzystujący zaawansowane algorytmy:

- a) aby szybko wyszukać podane słowa kluczowe,
- b) aby utworzyć ranking wyników z najbardziej względnym na górze,
- c) aby przedstawić je użytkownikowi.

Aby wyszukiwarka mogła szybko „odpowiedzieć” na nasze wyszukiwania, nie wyszukuje każdego słowa zawartego we wszystkich tych dokumentach, aby sprawdzić, czy istnieją tam słowa kluczowe. Zajmie to wieczność. Przeciwnie, wyszukiwarka wykorzystuje bazę danych do przechowywania informacji o każdej stronie internetowej, tworząc indeks, którego używa do szybkiego wyszukiwania słów kluczowych. Następnie algorytm rankingu decyduje, które wyniki są najbardziej odpowiednie do tego, czego szukamy, aby przedstawić nam najbardziej odpowiednie na górze.

Poniższy film w przyjemny sposób wyjaśnia działanie wyszukiwarki:

https://www.youtube.com/watch?v=LVV_93mBfSU

2.2.2.2.5 Wyniki wyszukiwania

Jak pokazano na zrzutach ekranu z rysunków 2, 7 i 8, każdy wynik wyszukiwania w wyszukiwarce ogólnego przeznaczenia ma nagłówek, który jest hiperłączem, który możemy kliknąć, aby odwiedzić stronę internetową. Pod tym nagłówkiem znajduje się krótki tekst, w którym znaleziono wyszukiwane słowa kluczowe. Ten mały tekst pomaga nam zrozumieć, czy ta strona internetowa / dokument zawiera informacje dotyczące tego, czego szukamy.

2.2.2.2.6 Zróżnicowanie wyników wyszukiwania

Jeśli wyniki wyszukiwania nie zawierają poszukiwanych informacji, możemy rozróżnić nasze wyszukiwanie. Prosty sposób na to jest zmiana słów kluczowych. Podczas wyszukiwania pokazanego na rys. 7 zauważamy, że nie widzimy wielu wyników dotyczących poszukiwanego projektu godigital. Jedynym względnym wynikiem jest strona projektu na Facebooku, a także nie na pierwszych pozycjach. Aby zobaczyć trafniejsze wyniki, powinniśmy przejść do następnych stron wyników. Słowa kluczowe „godigital” i „projekt” są tak powszechnymi słowami, że uzyskujemy wiele wyników niezwiązanych z tym, czego szukamy. Użyjmy teraz różnych słów kluczowych, a konkretnie „nauczycieli” i „godigital”. Wyniki zwrócone przez wyszukiwarkę są pokazane na rysunku 8. Jak widzimy, pierwszym rezultatem (zignorujmy reklamy) jest strona internetowa programu. Ten przykład pokazuje, że sukces wyszukiwania zależy od tego, jak udany jest wybór używanych przez nas słów kluczowych.

Innym sposobem na wyróżnienie naszego wyszukiwania jest wyszukiwanie całej frazy. Jeśli uwzględnimy w cudzysłów dwa lub więcej słów, wyszukiwarka przeszuka całą frazę, a nie każde słowo kluczowe osobno. Na przykład, jeśli nasze wyszukiwanie obejmuje „projekt godigital”, wyszukiwarka zwróci dokumenty zawierające frazę „projekt godigital”. Jeśli nie uwzględnimy cudzysłowu, wyszukiwarka pokaże dokumenty zawierające słowa godigital i project, ale mogą znajdować się w różnych częściach dokumentu.

2.2.2.2.7 Zaawansowane wyszukiwanie przy użyciu operatorów logicznych

Możemy użyć operatorów logicznych do bardziej wyszukanych wyszukiwań, aby zawęzić wyniki. Operatory logiczne to AND, OR i - (oznacza NIE).

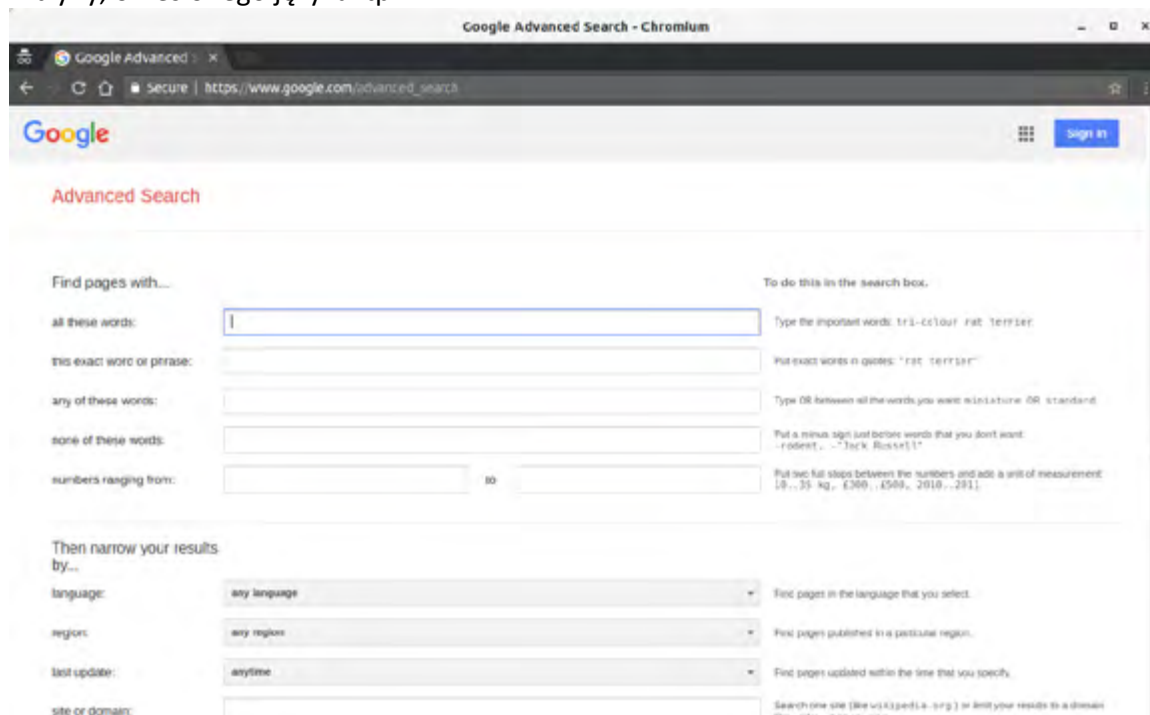
Użycie AND między dwoma słowami kluczowymi oznacza, że chcemy wyników zawierających oba te słowa. OR oznacza, że chcemy wyników zawierających co najmniej jedno z tych dwóch słów kluczowych. Jeśli użyjemy operatora przed słowem kluczowym, oznacza to, że nie chcemy wyników zawierających to słowo kluczowe.

Możemy przeprowadzać jeszcze bardziej zaawansowane wyszukiwania, jeśli użyjemy nawiasów. Na przykład wyszukiwanie *(tygrys lub dżungla) AND -Afryka* wyszukuje strony / dokumenty zawierające co najmniej jedno słowo tygrys, dżungla i jednocześnie nie zawiera słowa Afryka. To wyszukiwanie zwraca zupełnie inne wyniki niż wyszukiwanie *(tygrys lub dżungla) AND Afryka* w której usunęliśmy - operatora.

Warto zauważyć, że niektóre wyszukiwarki, takie jak Google, używają AND jako domyślnego operatora logicznego między słowami kluczowymi. Na przykład dla google.com używanie *nauczycieli godigital* jest takie samo jak *nauczycieli AND godigital*.

Google.com, która jest obecnie najbardziej zaawansowaną wyszukiwarką internetową, oferuje jeszcze kilku operatorów (patrz <https://support.google.com/websearch/answer/2466433?hl=en>) jak również wyszukiwanie

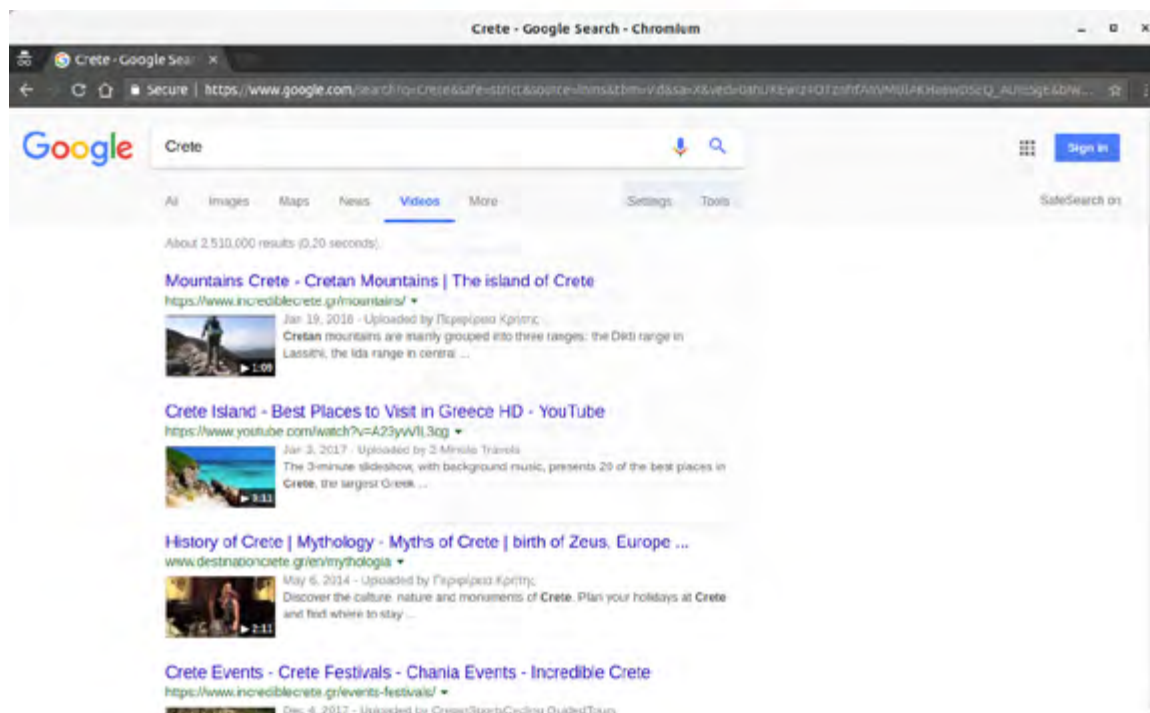
zaawansowane (www.google.com/advanced_search, zobacz Rysunek 10) które jest formularzem internetowym, za pomocą którego możemy dokonywać zaawansowanych wyszukiwań podobnych do tych, które możemy wykonywać przy użyciu operatorów logicznych, bez potrzeby pisania operatorów logicznych. Tam możemy zdefiniować także inne funkcje, które mają mieć wyniki wyszukiwania, takie jak typ pliku, wyniki z określonej witryny, określonego języka itp.



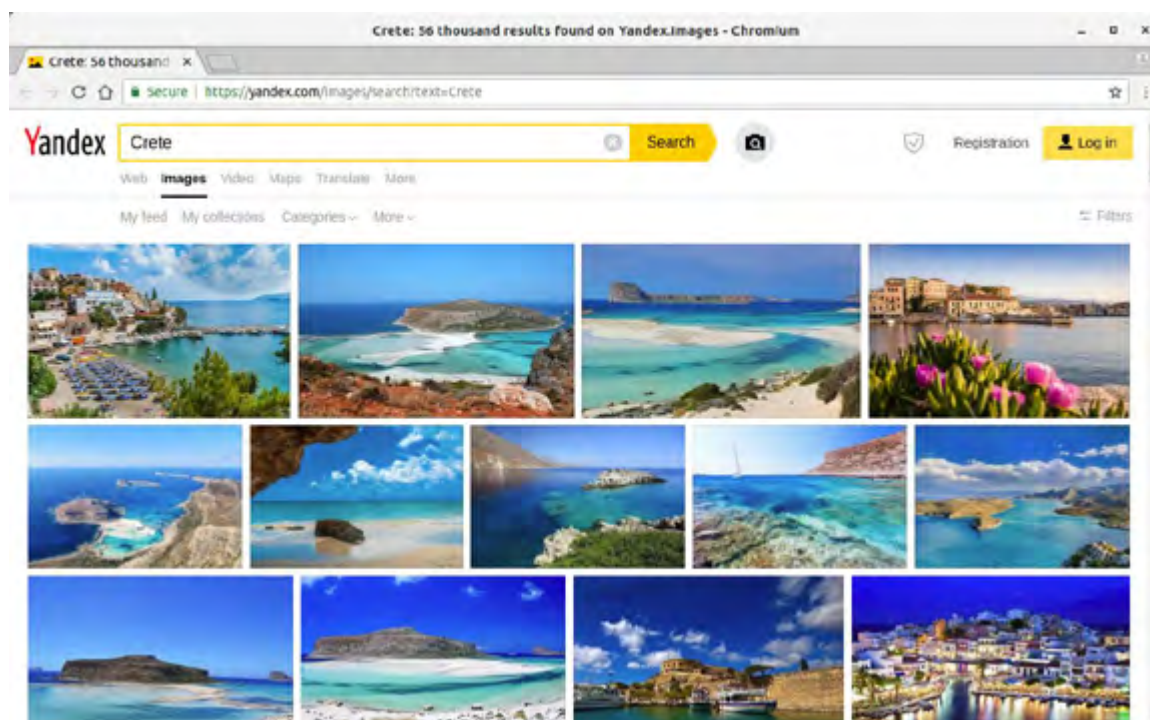
Rysunek 10: Zaawansowane wyszukiwanie w Google.com

2.2.2.2.8 Rodzaj wyników wyszukiwania

Fajną funkcją głównych wyszukiwarek internetowych jest możliwość wyboru rodzaju wyników wyszukiwania, jakie chcemy. Na przykład możemy poprosić wyszukiwarkę, aby przedstawiła nam tylko obrazy lub tylko mapy lub tylko filmy związane ze wstawionym słowem kluczowym. Rysunek 11 to zrzut ekranu z wynikami wideo w wyszukiwarce Google.com, która użyła Krety jako słowa kluczowego wyszukiwania. Rysunek 12 to zrzut ekranu z wynikami obrazów w wyszukiwarce Yandex.com dla tego samego wyszukiwania. Zauważ, że w obu wyszukiwarkach wybór rodzaju wyników wyszukiwania znajduje się dokładnie pod polem tekstowym, w którym użytkownik wstawia swoje słowa kluczowe wyszukiwania. W przypadku Google.com dostępne są następujące opcje: „Wszystkie”, „Obrazy”, „Mapy”, „Wiadomości”, „Filmy” i „Więcej” (w przypadku „Książek”, „Loty”, „Finanse”, „Osobiste”)



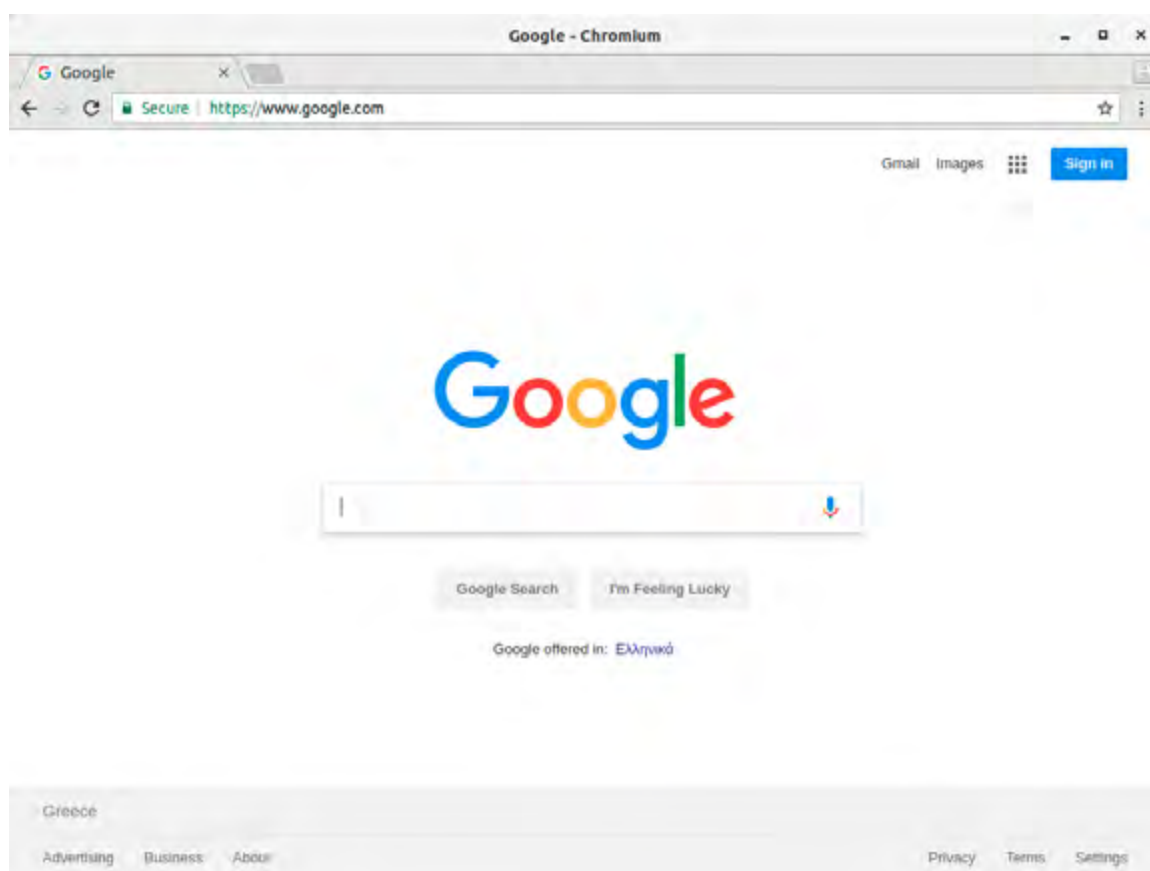
Rysunek 11: Wyniki wyszukiwania wideo dla Krety w wyszukiwarce Google.com.



Rysunek 12: Wyniki wyszukiwania obrazów dla Krety w wyszukiwarce Yandex.com.

2.2.2.2.9 Bezpieczne wyszukiwanie - Filtrowanie wyników wyszukiwania

Google.com oferuje wiele innych funkcji na swojej platformie wyszukiwarki internetowej i od czasu do czasu tworzy więcej. Dla nauczyciela ważne jest jednak, aby wspomnieć o ustawieniu SafeSearch. Jest on aktywowany lub dezaktywowany w opcji Ustawienia, którą można znaleźć na jego stronie internetowej (patrz Rysunek 13). Aktywując to ustawienie, wyniki są filtrowane, aby wykluczyć obraźliwe lub jednoznacznie seksualne wyniki. Jest to bardzo przydatna funkcja pozwalająca uniknąć niepożądanych sytuacji w klasie.



Rysunek 13: Filtr SafeSearch można aktywować za pomocą opcji Ustawienia w prawym dolnym rogu strony Google.com.

2.2.2.2.10 Fałszywe informacje w sieci

Główny problem, gdy znajdujemy informacje w sieci WWW, wiąże się z ważnością odkrytych informacji. Fakt, że każdy może publikować informacje na ten temat, nieuchronnie oznacza, że wiele fałszywych informacji jest publikowanych umyślnie lub nie. Dlatego wszelkie informacje należy traktować krytycznie i ostrożnie przyjmować. Oto kilka prostych zasad, których możemy przestrzegać, aby uniknąć stania się ofiarą dezinformacji:

- Używaj zdrowego rozsądku (np. Wszyscy wiemy, że ziemia nie jest płaska)



- Uzyskaj informacje z prawidłowych stron internetowych (np. Strona internetowa uniwersytetu lub organizacji rządowej daje nam pewność, że informacje tam są aktualne, w przeciwieństwie do osobistego bloga)
- Sprawdź ostatnią aktualizację strony internetowej zawierającej informacje
- Znajdź potwierdzenie informacji na wielu stronach internetowych