



## Internet and Information Searching

## 2.2.2 TOPIC 1: INTRODUCTION TO INTERNET

### 2.2.2.1 Introduction to Internet

Internet is a reality that has been established in our lives in the relatively recent years. In this section several introductory concepts of internet are discussed.

#### 2.2.2.1.1 What is Internet

Internet as word comes from the the words **Inter**connected **Network**. Thus, it etymologically means that it is a network of networks, that is many smaller networks connected with each other creating a bigger one, the Internet. In other words every computer or other device (e.g. printer, mobile phone, tablets and lately televisions, air conditions, watches etc) connected to Internet is essentially part of it. The connection is established wirelessly (i.e. through antennas and satellites) or wired (e.g. through the telephone network). Internet started as military project in USA in 1960s.

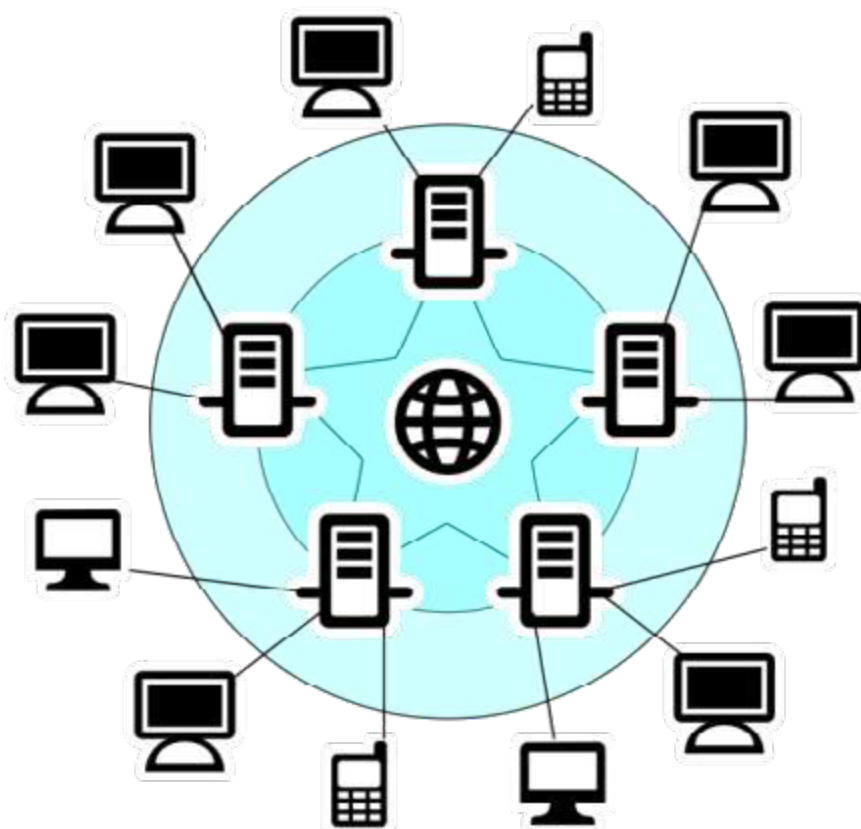


Figure 1. Internet is a big network consisting of smaller networks and devices connected to each other.

### 2.2.2.1.2 Internet services

Internet offers various services such as World Wide Web (the web pages), Communication services (Electronic Mail, Telnet, VoIP, etc.), FTP (File Transfer Protocol) and IRC (Internet Relay Chat).

It is important to point out that World Wide Web is just a service of Internet. Since though, it has been ended up to be the dominant service and many applications have been developed on it to accommodate other services (e.g. E-mail and FTP), many people erroneously perceive World Wide Web as identical to Internet.

### 2.2.2.1.3 Web pages in World Wide Web

The World Wide Web was invented by Tim Berners Lee in 1989 and consists of documents and other files (e.g. images and videos) linked to each other through hyperlinks. These documents and files are located in computers connected to Internet. The web pages are documents written in a specific format based on the Hypertext Markup Language (HTML). The very first webpage was published to the general public in August 1991.

To view a web page we need a special computer program that is called web browser and the web address of the web page. The browser uses this address to locate it. Examples of web browsers are Mozilla Firefox, Google Chrome, Opera, Safari and Edge.

### 2.2.2.1.4 Addresses in World Wide Web

The address of a web page or file in World Wide Web, which you can also see it referred as URL (Uniform Resource Locator), has the following format:

*http://www.sitename.extension*

*Http* is a protocol which sometimes you can see it as *https* (the *s* for secure). We can omit *http://* and the web browser understands that the *http* protocol should be used.

The next part is *www* and sometimes can be omitted too if the website administrator has done the appropriate setting.

The part *sitename.extension* is called domain name. The extension is a standard word called Top-Level-Domain (TLD) that declares - but not necessarily - the type or the geographic location of the web site. For example *com* extension is for commercial, *org* for organization, *bu* for Bulgaria, *cy* for Cyprus, *gr* for Greece, *it* for Italy, *pl* for Poland etc. You can view a full list of the available TLDs at <https://www.icann.org/resources/pages/tlds-2012-02-25-en>

The sitename is a name selected by the website owner and registered to an official organization (different in each country) under a specific TLD. Notice that *sitename.com* is different than *sitename.org* and lead to different web sites. Nevertheless, if the web site holder has registered both domains then he can drive both domains to the same web site.

### 2.2.2.1.5 E-mail Addresses

Electronic Mail or e-mail is the internet service with which we can exchange electronic letters. Someone to send a message to another person using e-mail should have an e-mail address and send it to another e-mail address, the address of the receiver.

The format of an e-mail address is

*user\_selected\_name@email\_provider.extension*

The part *user\_selected\_name* is the part that the user chooses. It may be a nickname or his/her name/surname and is always followed by @ character which is pronounced “at”.

The part *email\_provider.extension* is called domain. The extension is a TLD (recall from paragraph about Addresses in World Wide Web).

It is important to notice that two email addresses that have the same *user\_selected\_name* is completely different if they are at a different domain. For example [johnsmith@hotmail.com](mailto:johnsmith@hotmail.com) is a completely different address to [johnsmith@gmail.com](mailto:johnsmith@gmail.com)

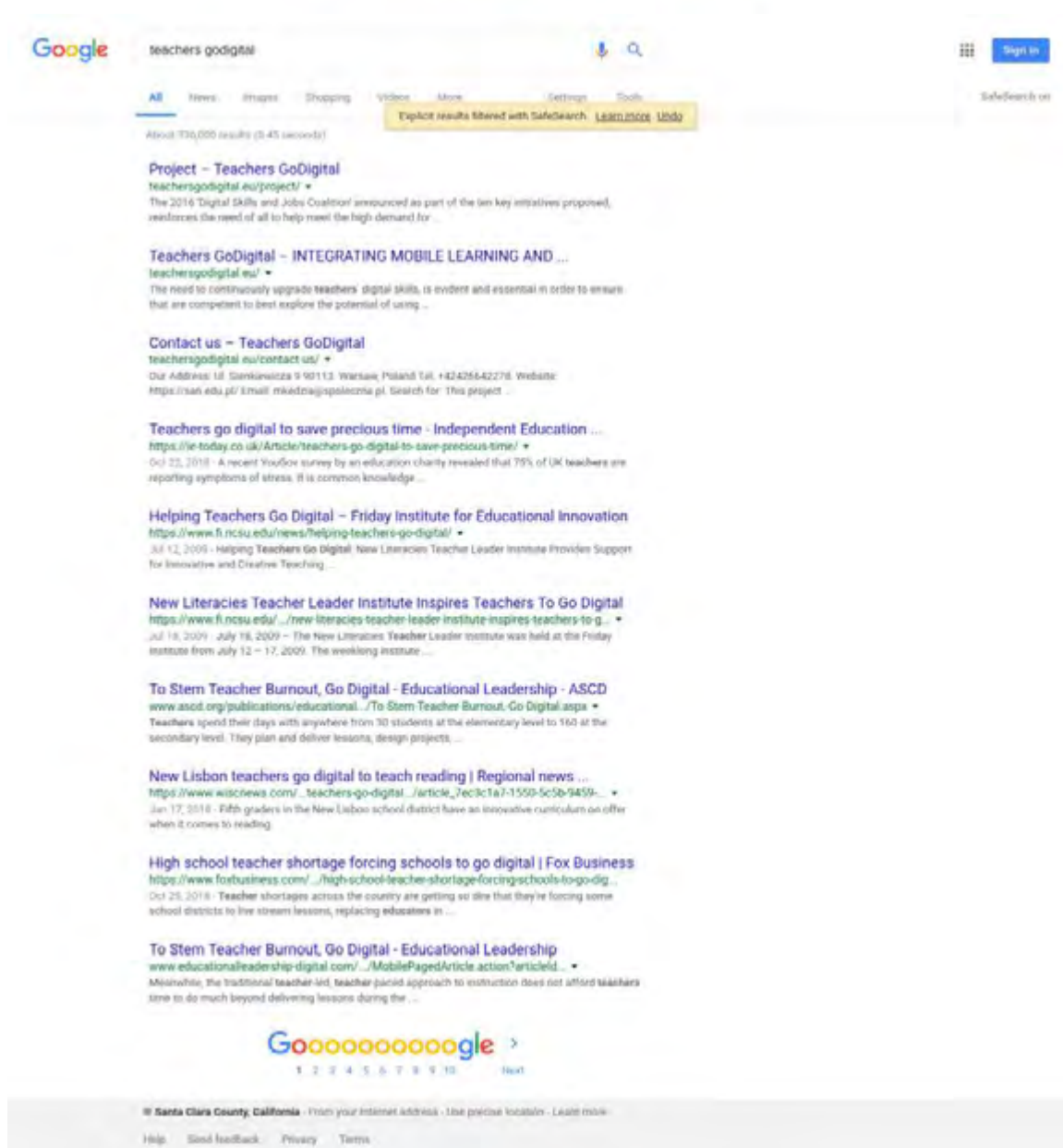
Notice also that johnsmith.hotmail.com is a web page address and not e-mail address. An email address should always contain the @ character.

### 2.2.2.2 Information searching

World Wide Web is an ocean of information. Everyone can publish information in various forms i.e. text, images, videos, maps etc. The huge amount of the published information and the fact that there is no control in its validity creates correspondingly the challenges of how to effectively search information on the web and how to ensure that the discovered information is valid and not fake or inaccurate.

#### 2.2.2.2.1 What is a Web Search Engine

A web search engine is an ally of us to confront the challenge of effective information search on World Wide Web. In its back end is a computer program consisting of complicated algorithms that preserves a database with details about the content of the web pages. The front end is a web form through which the user can enter the keywords or phrases the user wants to search. The search engine searches these keywords and phrases in its database and creates a ranking based on the relevancy of the results. The results are presented to the user in a paged form. Figure 2 is a screenshot with the results of a search in Google’s search engine having used the keywords *teachers godigital*.



**Figure 2:** The results of the search with keywords teachers godigital in Google search engine.

### 2.2.2.2 Web Search Engine categories

We can distinguish the following categories of web search engines:

- **General Purpose Search Engines**

A general purpose search engine searches and retrieves information from every web page registered in its database. A web page is registered in the database either when the web site owner registers it by himself/herself or when a special computer program of the search engine (called crawler or bot) discovers a web page not already registered in its database. Examples of general purpose search engines are:

- a. [www.google.com](http://www.google.com)
- b. [www.duckduckgo.com](http://www.duckduckgo.com)
- c. [www.yandex.com](http://www.yandex.com)
- d. [www.bing.com](http://www.bing.com)
- e. [www.yahoo.com](http://www.yahoo.com)
- f. [www.ask.com](http://www.ask.com)
- g. [www.aol.com](http://www.aol.com)
- h. [www.baidu.com](http://www.baidu.com)

- *Special Purpose Search Engines*

A special purpose search engine is a search engine for which there is a control of what content is stored in its database. An example is the search engine included in a particular website for example like wikipedia.com. Wikipedia.com is an online encyclopedia and therefore its content should be specific to the topics that an encyclopedia may contain. For example we will not find information regarding the weather prediction of the next day. On the other hand, a website for weather predictions has a search engine to search weather predictions for a specific area or historical data about the weather of that area.

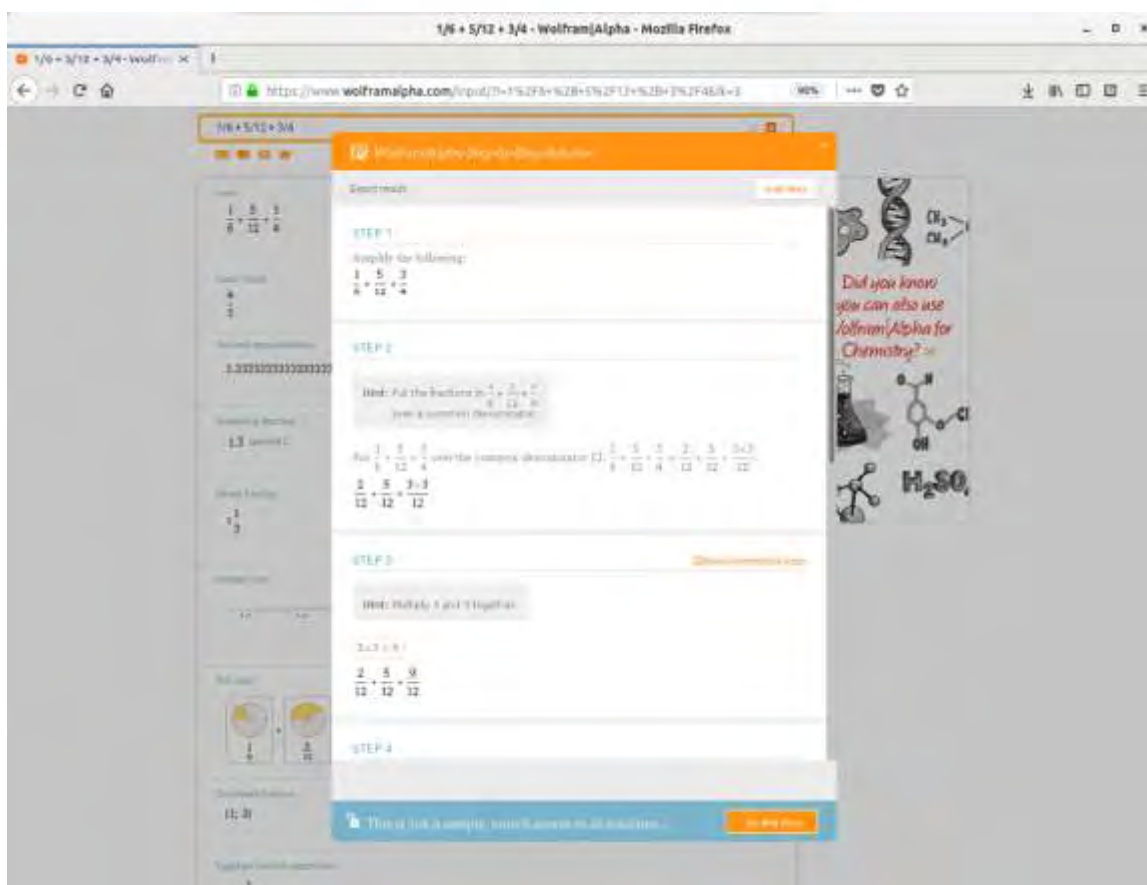
- *Meta-Search Engines*

A meta-search engine is a search engine that collects search results from multiple search engines to present them to the user. For example, skyscanner.com is a meta-search engine which queries the search engines of travel companies, collects their answers and subsequently present them to the user.

- *Computational Knowledge Engines*

A computational knowledge engine does not present search results to the user like a search engine does, rather computes the answer of the user's search or retrieves the answer from selected sources. That can be very useful for educators and learners. Wolframalpha.com is such a computational knowledge engine. Figure 3 is the screenshot for the search of what is the step-by-step process for the math  $1/6 + 5/12 + 3/4$ , while Figure 4 is the screenshot of the search for the famous Greek writer Nikos Kazantzakis.





**Figure 3:** Search result of what is the step-by-step process for the math  $\frac{1}{6} + \frac{5}{12} + \frac{3}{4}$  in WolframAlpha.com Computational Knowledge Engine.

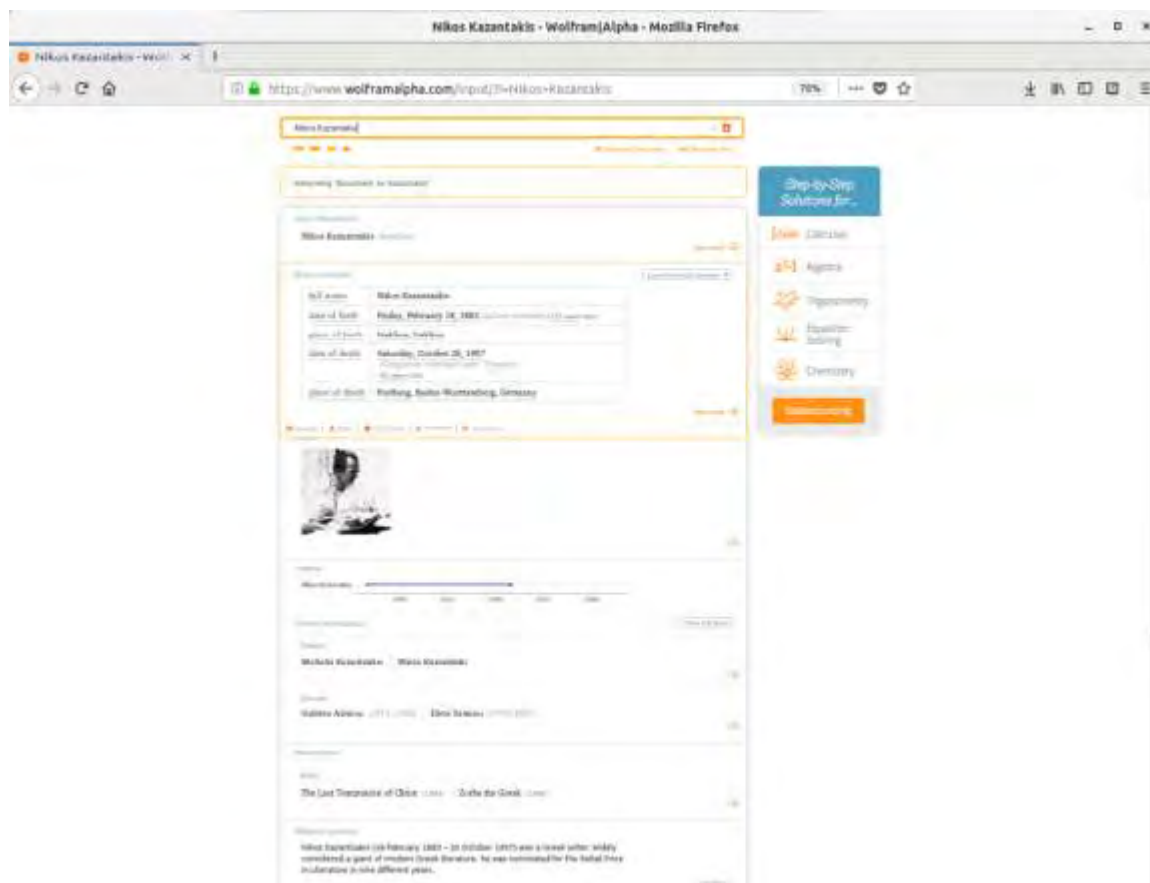
### 2.2.2.2.3 Simple searches

To see in practice how we can search information on the World Wide Web, let us suppose that we want to search information about the GoDigital Project, product of which is the present text.

The first step is to open a web browser (e.g. Mozilla Firefox or Google Chrome or Opera etc.) and in its address bar (see Figure 5) to insert the web address of the search engine we want to use (recall the addresses of a general purpose search engine we mentioned in 2.2.2.2.2 paragraph). Let us use the duckduckgo.com search engine. The cornerstone of the philosophy of duckduckgo.com search engine is that it does not track our searches and protects our personal data (more information you can find in [duckduckgo.com/spread](https://duckduckgo.com/spread)).

The second step is to insert the keywords to perform the search. In every web search engine there is a text box (usually in the center of the web page), where we insert the keywords of our search. It is very important to use the correct keywords in order to get the most relevant results. For this search we will use the keywords “godigital” and “project” that later we will show that it is not the best choice. Screenshot of this search is shown in Figure 6 and its results in Figure 7.

We should also point out that instead of visiting the web page of the search engine to perform the desired search, we can alternatively insert the search keywords directly in the address bar of the browser (see Figure 9). The browser will use a predefined search engine (usually google.com) and redirect us in the web page of the results. The search engine that the browser by default uses can be changed from the Settings/Preferences section of the browser.



**Figure 4:** search for the famous Greek writer Nikos Kazantzakis in Wolframalpha.com Computational Knowledge Engine.





**Figure 5:** The address bar in a web browser is a text box (usually on the top of the browser's window), where we type the web site address we want to visit.



**Figure 6:** Search using DuckDuckGo search engine using the keywords godigital project

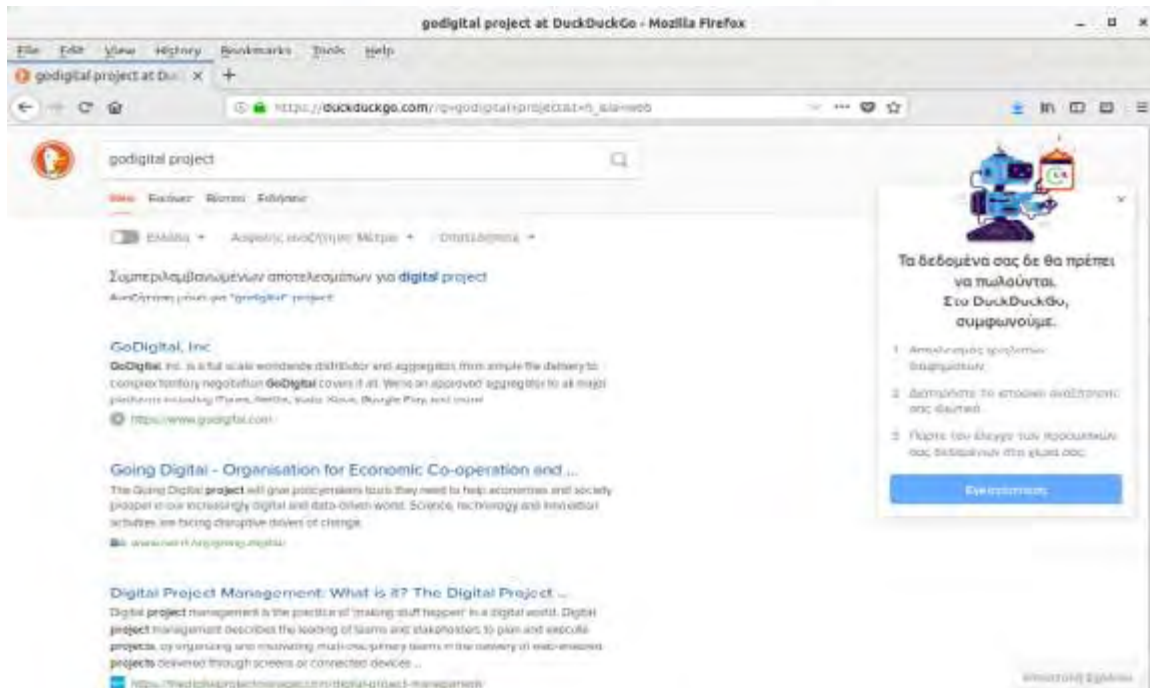


Figure 7: The search results of the search shown in Figure 6

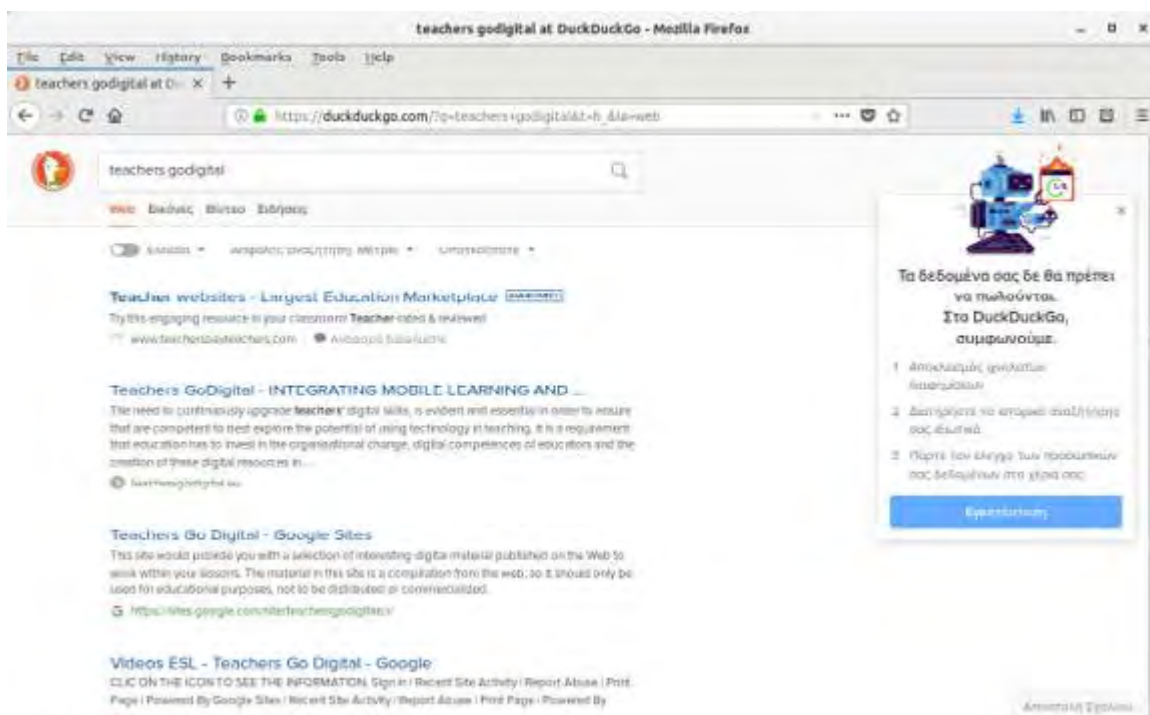
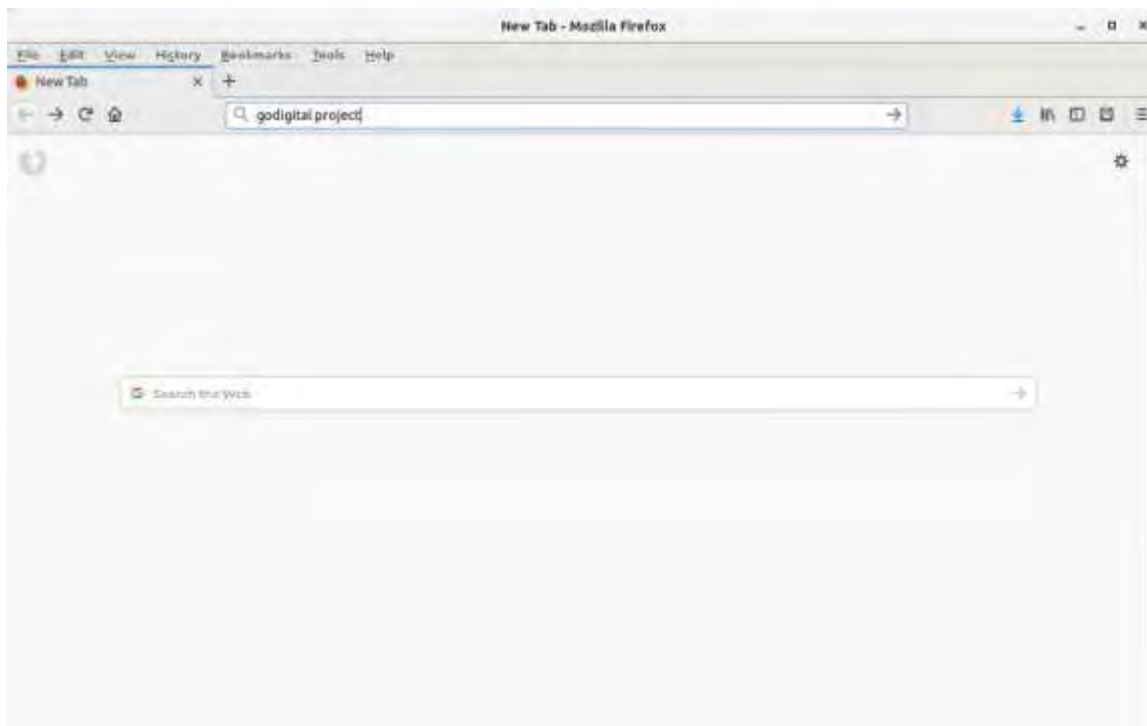


Figure 8: Changing the keywords in a search differentiates the search results (see also Figures 6 and 7)



**Figure 9:** We can do a search by inserting the keywords in the address bar of the web browser. The web browser used a predefined search engine which we can change from the settings/preference option of the browser.

#### 2.2.2.2.4 How a Search Engine works

A search engine is a complex software program that uses advanced algorithms:

- a) to quickly search the given keywords,
- b) to create a ranking of the results with the most relative on the top, and at last
- c) to present them to the user.

In order the search engine to be able to “answer” quickly to our searches, it does not search each word contained in all these documents to see if the keywords exist in there. This would be taking forever. On the contrary, the search engine uses a database to keep information about each webpage, creating an index that it uses to search the keywords quickly. Subsequently, a ranking algorithm decides which results are most relevant to what we are searching so that to present to us the most relevant on the top.

The following video explains in a nice way the way a search engine works:

[https://www.youtube.com/watch?v=LVV\\_93mBfsU](https://www.youtube.com/watch?v=LVV_93mBfsU)

#### 2.2.2.2.5 The search results

As shown in the screenshots of Figures 2, 7 and 8, each result of a search in a general purpose search engine has a header, which is a hyperlink we can click to visit the web page. Below this header there is a brief text within which the searched keywords were found. This

small text helps us realise if this webpage/document has information relative to what we are looking for.

#### 2.2.2.2.6 Differentiating the search results

In case the search results do not contain the information we are seeking then we can differentiate our search. A simple way to do it is to change the keywords. In the search shown in Figures 7 we notice that we do not see many results about the godigital project we are looking for. The only relative result is the Facebook Page of the project and also not in the first positions. In order to see more relevant results we should continue to the next pages of the results. The keywords “godigital” and “project” are so common words that we get many results unrelated to what we are looking for. Let us now use different keywords and specifically the keywords “teachers” and “godigital”. The results returned by the search engine are those shown in Figure 8. As we can see the very first result (ignoring the advertisement) is the web page of the program. This example shows that the success of a search depends on how successful is the selection of the keywords we use.

Another way to differentiate our search is to search a whole phrase. If we include two or more words in double quotes then the search engine searches the whole phrase and not each keyword separately. For example if our search is “*godigital project*” the search engine will return documents that contain the phrase “godigital project”. If we do not include the double quotes, the search engine will return documents that contain the words *godigital* and *project* but they may be in different parts of the document.

#### 2.2.2.2.7 Advanced searches using logical operators

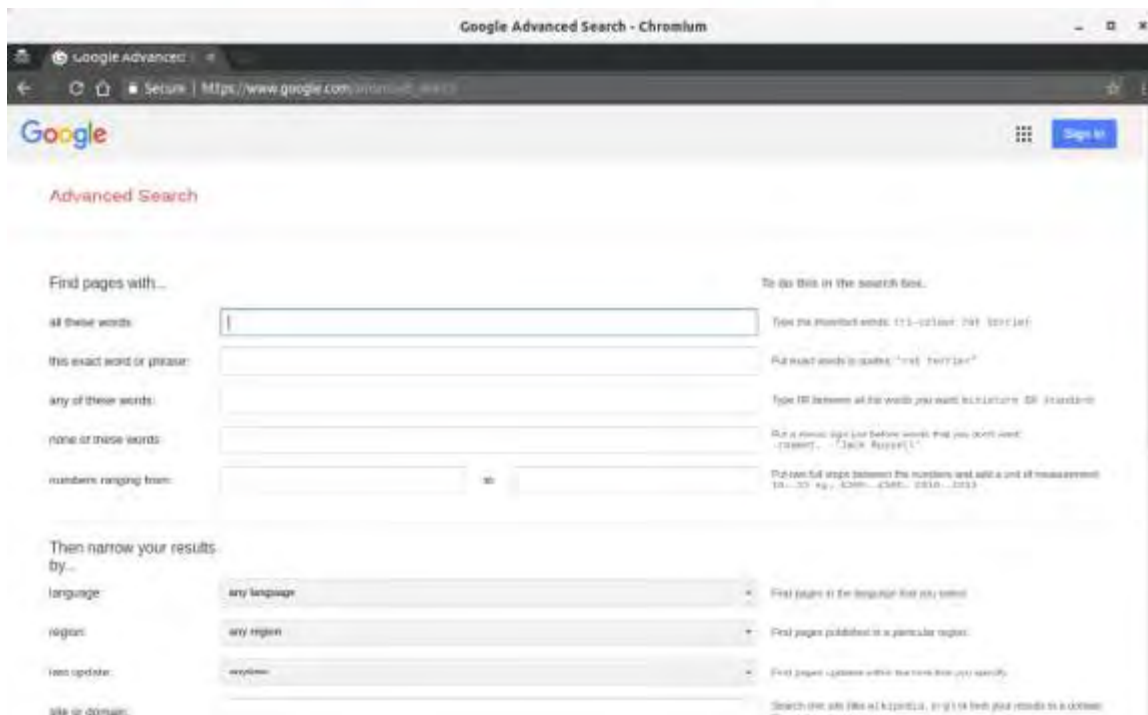
We can use logical operators to make more sophisticated searches so that to refine the search results. The logical operators are the AND, OR and - (means NOT).

Using AND between two keywords means that we want results containing both these two keywords. On the contrary, OR means that we want results containing at least one of these two keywords. If we use the - operator before a keyword, means that we do not want results containing this keyword.

We can do even more advanced searches if we use parenthesis. For example the search (*tiger or jungle*) AND -Africa searches for web pages/documents containing at least one of the words tiger, jungle and simultaneously do not contain the word Africa. This search returns completely different results than the search (*tiger or jungle*) AND Africa in which we have removed the - operator.

It worths to point out that some search engines like Google uses AND as default logical operator between keywords. For example, for google.com using *teachers godigital* is the same as *teachers AND godigital*.

Google.com which is currently the most sophisticated web search engine offers several more operators (see <https://support.google.com/websearch/answer/2466433?hl=en>) as well as the Advanced Search ([www.google.com/advanced\\_search](http://www.google.com/advanced_search), see Figure 10) which is a web form with which we can make advanced searches similar to the ones we can make using logical operators, without the need to write logical operators. There we can define also other features we want the search results to have like file type, results from a specific site, of specific language etc.

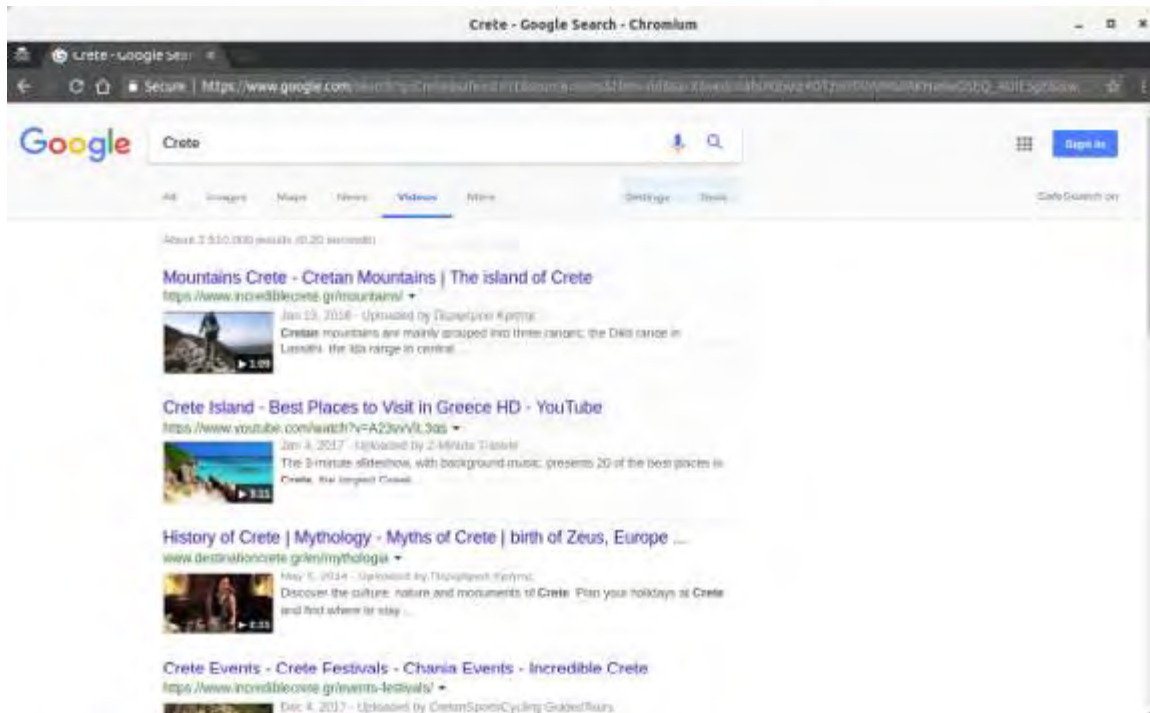


**Figure 10:** The Advanced Search of Google.com

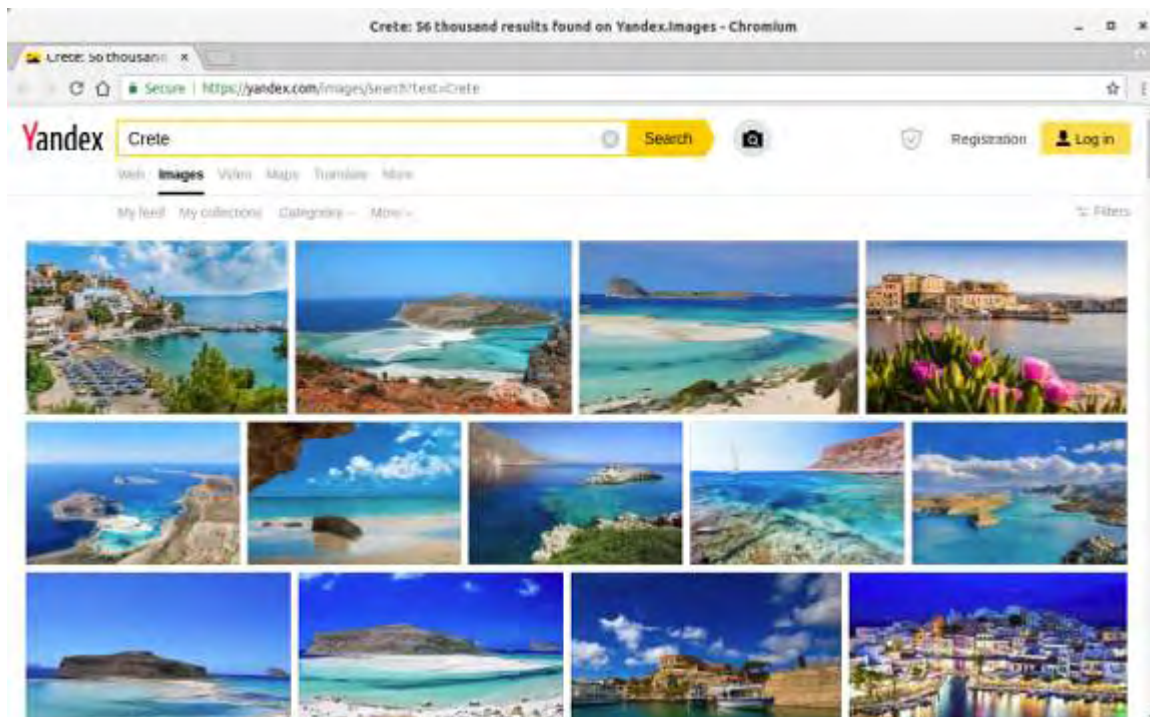
#### 2.2.2.2.8 Type of search results

A nice feature that the major web search engines have is to be able to choose the type of search results we want. For example we can ask from the search engine to present us only images or only maps or only videos relative to the keyword we inserted. Figure 11 is a screenshot with the video results in Google.com search engine having used *Crete* as search keyword. Figure 12 is a screenshot with images results in Yandex.com search engine for the same search. Notice that in both search engines the selection of the type of the search results is located exactly below the text box in which the user inserts his/her search keywords. For Google.com the options are “All”, “Images”, “Maps”, “News”, “Videos” and “More” (for “Books”, “Flights”, “Finance”, “Personal”)





**Figure 11:** Video search results for Crete in Google.com search engine.

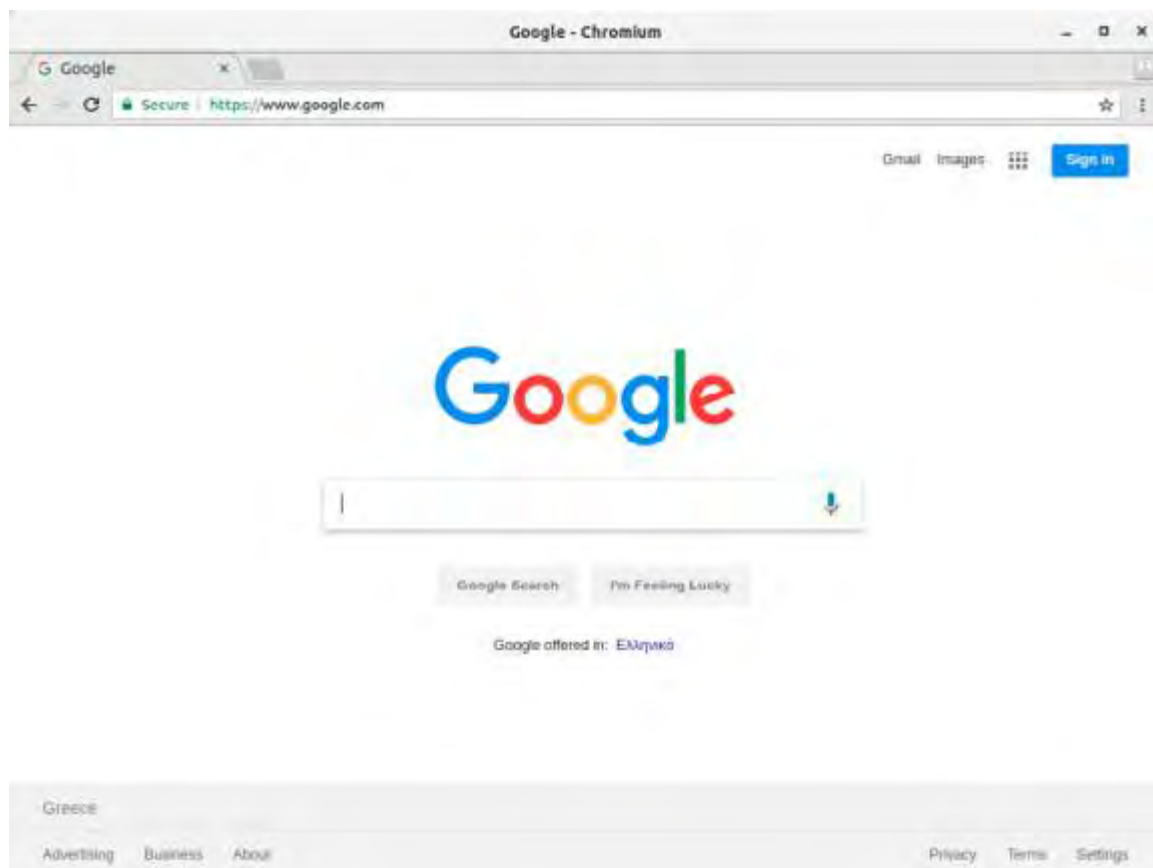


**Figure 12:** Image search results for Crete in Yandex.com search engine.



#### 2.2.2.2.9 Safe search - Filtering search results

Google.com offers plenty of more features in its web search engine platform and creates more every now and then. For a teacher though it is important to mention the SafeSearch setting. It is activated or deactivated from the Settings option you can find on its web page (see Figure 13). By activating this setting the results are filtered so that to exclude offensive or sexually explicit results. It is a very useful feature to avoid undesirable situations in the classroom.



**Figure 13:** SafeSearch can be activated from the Setting option on the bottom-right of Google.com web page.

#### 2.2.2.2.10 False Information on the Web

A main concern when we find information on the World Wide Web has to do with the validity of the discovered information. The fact that everyone can publish information on it, inevitably means that there is lot of false information published either intentionally or not. Therefore, any information should be faced critically and adopted carefully. Some simple rules we can follow so that avoid becoming victims of misinformation are the following:

- Use common sense (e.g. we all know that the earth is not flat)



- Get information from valid websites (e.g. the website of a university or a governmental organization gives us the confidence that the information there is valid, on the contrary to a personal blog)
- check the last update of the web page containing the information
- confirm an information in multiple websites