BRNO UNIVERSITY OF TECHNOLOGY

FACULTY OF ELECTRICAL ENGINEERING AND COMMUNICATION

DEPARTMENT OF FOREIGN LANGUAGES

DO ALTERNATIVE SEARCH ENGINES COMPARABLE TO GOOGLE EXIST?

EXISTUJÍ ALTERNATIVNÍ INTERNETOVÉ VYHLEDÁVAČE SROVNATELNÉ S GOOGLEM?

BACHELOR'S THESIS
BAKALÁŘSKÁ PRÁCE

AUTHOR   Otto Černovický
AUTOR PRÁCE

SUPERVISOR   Mgr. Bc. Dagmar Šťastná
VEDOUCÍ PRÁCE

BRNO 2022

1

Bachelor's Thesis

Bachelor's study field **English in Electrical Engineering and Informatics**

Department of Foreign Languages

Student: Otto Černovický *ID:* 211292

Year of study: 3 *Academic year:* 2021/22

TITLE OF THESIS:

Do alternative search engines comparable to Google exist?

INSTRUCTION:

Explore available alternative Internet search engines and focus on their technical aspects.

RECOMMENDED LITERATURE:

Dover, Danny a Dafforn, Erik. (2012). SEO: optimalizace pro vyhledávače profesionálně. Brno: Zoner Press. Encyklopedie webdesignera.

Hock, Randolph. (2013). The extreme searcher's internet handbook: a guide for the serious searcher. 4th ed. Medford, New Jersey: CyberAge Books.

Levene, Mark (2005). An introduction to search engines and web navigation. Hoboken, New Jersey: John Wiley & Sons.

Date of project specification: 10.2.2022 *Deadline for submission:* 31.5.2022

Supervisor: Mgr. Bc. Dagmar Šťastná

doc. PhDr. Milena Krhutová, Ph.D.
Subject Council chairman

Abstrakt

Tato bakalářská práce se věnuje internetovým vyhledávačům. Skoro každý v dnešní době nějaký používá, většina z nás ovšem netuší, jak tento proces funguje.

Cílem této práce je z jisté části pochopit, jak internetový prohlížeč funguje, stejně jako se seznámit s historií Googlu, Microsoftu a Yahoo!. Jelikož Google je nejpopulárnější z trojice, tak kromě historie i propagace a YouTube jsou zmíněny. Dalším důležitým bodem této semestrální práce je, jak internetový vyhledávač značky Google funguje. Pro lepší porozumění se první část tohoto bodu věnuje obecně vyhledávači a popisuje, jak k vyhledávání dochází. Druhá část se už věnuje konkrétním aktualizacím.

Vzhledem k obrovské rychlosti vývoje internetových vyhledávačů, jsem převážně čerpal z internetových a knižních zdrojů. Pro lepší reprezentaci vlivů určitých algoritmů, jsou poskytnuty fotky vyhledávání přímo z Googlu.

Klíčová slova

Google, Yahoo!, Microsoft Bing, vyhledávač, algoritmus

Abstract

This bachelor thesis deals with the search engine. Almost everyone uses one today, but most of us have no idea how this process works.

The aim of this thesis is to understand how the search engine works and het acquainted with the history of Google, Microsoft, and Yahoo!. As Google is the most popular of the three, history, advertising and YouTube are also mentioned. Another important point is how Google's search engine works. For a better understanding, the first part of this chapter describes how any search engine works. The second part depicts Google's significant updates.

Due to the tremendous speed of search engine development, Internet sources and books are main source to this thesis. To better represent the effect of certain algorithms, screenshots of searching queries are provided directly from Google.

Keywords

Google, Yahoo!, Microsoft Bing, search engine, algorithm

ČERNOVICKÝ, Otto. *Existují alternativní internetové vyhledávače srovnatelné s Googlem?* [online]. Brno, 2022 [cit. 2022-05-25]. Dostupné z: https://www.vutbr.cz/studenti/zav-prace/detail/142531. Bakalářská práce. Vysoké učení technické v Brně, Fakulta elektrotechniky a komunikačních technologií, Ústav jazyků. Vedoucí práce Mgr. Bc. Dagmar Šťastná.

Prohlášení

Prohlašuji, že svůj semestrální projekt na téma Existují alternativní internetové vyhledávače srovnatelné s Googlem? Jsem vypracoval pod vedením vedoucího semestrálního projektu a s použitím odborné literatury a dalších informačních zdrojů, které jsou všechny citovány v práci a uvedeny v seznamu literatury na konci práce.

     Jako autor uvedeného semestrálního projektu dále prohlašuji, že v souvislosti s vytvořením tohoto semestrálního projektu jsem neporušil autorská práva třetích osob, zejména jsem nezasáhl nedovoleným způsobem do cizích autorských práv osobností a/nebo majetkových a jsem si plně vědom následků porušení ustanovení § 11 a následujících zákona č. 121/2000 Sb., o právu autorském, o právech souvisejících s právem autorským a o změně některých zákonů (autorských zákon), ve znění pozdějších předpisů, včetně možných trestněprávních důsledků vyplývajících z ustanovení části druhé, hlavy VI. Díl 4 Trestního zákoníku č. 40/2009 Sb.

V Brně dne ………………   ……………………………
                                                                        (podpis autora)

List of contents

List of figures

Introduction

Over the last three decades, computer technology and web technology have been tremendously evolved from the computers that were able to carry out basic mathematical operation to tiny portable devices that users carry in their pockets. With the rise of computer users, there has been a significant rise of web users. According to Mark Levene, the Internet is a physical network connecting computers together. On the other hand, a virtual network is the Web linking together a massive amount of data and information (Levene, 2010). Numerous definitions of search engines can be found such as 'a program that searches for and identifies items in a database that corresponds to keywords or characters specified by the user, used especially for finding particular sites on the World Wide Web' (Oxford Languages, n.d.) or 'computer software used to search data (such as text or a database for specified information' (Merriam-Webster, n.d.). The search engine is a computer program with the intention of finding specific information in a large amount of data. This information can be further specified by the user. For instance if users are looking for image or web page content. Search engines are usually linked to the Internet, but they can be searching something completely different such as database. Every search engine works differently in detail but all of them follow the same basic routine consisting of three steps. The first step is *crawling*, which means scouting the Internet and looking for content of each URL. Crawling is followed by *indexing* that is processing and storing already found information and finally, *ranking*, which is based on several criteria creating a list from the most interesting to less interesting finding (Britney Muller & Mozz Staff, 2021).

Despite the fact that every search engine follows the same routine, they may vary in certain steps. Nowadays web page owners usually want to succeed and place their web page on as high position as possible. In order to reach and fulfill this goal, owners update their web pages to get a high score by the ranking algorithm. For instance, Google never revealed their ranking factors, resulting in the only possible option for the web pages' owner to guess them. On the flip side, Bing approached this in a very different way. Bing disclosed the whole list of ranking factors even with the value of each factor to the public eyes. Microsoft even took a step further by adding Bing Webmaster to the search engine with the intention to help owners to improve their ranking score. Yahoo! started as a directory of web links organized by users. Technically, it does not meet the criteria to be a proper search engine, but it represents

a rather strange and ineffective way (by today's standards) to navigate through millions of web pages and organize them.

With the rapid development of machine learning and artificial intelligence, interesting opportunities arise. Visual search is an interesting example of this progression. Search engines developed to the point, when users can simply create photography of objects and artificial intelligence recognize it and search it on the internet.

This bachelor thesis covers four principal chapters. Chapter 1 describes Google and the other two search engines Yahoo! and Microsoft Bing. It deals with Google's history including founders, data, and shifts on the board of companies and is followed by the Google's advertisement. Last but not least chapter 1 describes the history of Google Video, YouTube, and their development. Chapter 2 depicts content relevance, linking, popularity based measurements and Google's significant algorithms.

As chapter 2 focuses on Google, chapter 3 focuses on similarities and differences between Google, Bing, and Yahoo!. Chapter 3 reveals several differences between Google and Bing, for instance, social integrated results or image search results. The next subchapter follows the same routine and deals with the differences between Google and Yahoo!, especially it sheds light on Google and Yahoo! home page interfaces. Finally, chapter 3 explains Yahoo! instant search and its usages. The last chapter in this thesis talks about the possible space for improvement of some already working search engine features such as voice search and visual search.

1 Google and other search engines

1.1 Google

Google is an American company widely known for its search engine. Even though the search engine is a major reason for its rise, it is not the only thing Google is participating in. Google owns the most successful search engine, mobile hardware, cloud units, and other gadgets.

In the first part of my thesis, Google's history, advertising, and other important aspects of Google's rise, such as YouTube are mentioned. The chapter 1.1.1 focuses on history, it describes the company's founders and some other factors that played a role in its foundation. It also briefly covers and compare the number of queries in the beginning and after several years. The infrastructure and concept of Google's operation are mentioned, followed by a brief explanation. Lastly, significant changes concerning managers and other important figures were brought up. Since advertising is an inseparable part of Google's plan, following chapter 1.1.2 deals with Google's successful advertising strategies. The important change in ways of advertising and Google's reaction to it is described. Chapter 1.1.3 concerns in YouTube and its origin. All numbers such as years or amount of money used in this chapter are known or freely accessible for every internet user.

1.1.1 Google's history

Google was founded in 1998 by Sergey Brin and Larry Page as a subsidiary of the holding company Alphabet Inc. During 1998 Bring and Page started receiving financial means resulting in raising $1 million. After that they opened a shop named Google, which was a mistake because the intentional name was Googol. This is the mathematical term for number one followed by hundred of zeros. During 1999, Google was processing 500 000 queries per day (Hosch, 2020). During 2000, Google was used as a search engine for one the most popular sites Yahoo!. By the end of 2004, Google was used 200 million times per day and in the same year, Yahoo! left cooperation with Google. This growth did not stop and by the end of 2011, Google was managing 3 billion requests per day (Hosch, 2020). Google began operating as an

online search firm, but to this date it owns more than 50 Internet services. Google's broad product portfolio makes it one of the most influential high-tech companies along with Apple, IBM and Microsoft. Even though Google has many products on the market, its business core remained the same and that is its search engine. In 2016, almost all of Alphabet's revenue came from Google advertising based on users' search requests.

Brin and Page wanted to create a tool which would gather specific information from the Internet. The main difference between Google and other search engine is in the way the result is executed. Google's way is to consider not only several positive matches of search phrases but also consider a number of links each Web site has. The web page with more links leading to this site is supposed to be more valuable to the user. On the other hand, other search engines only considered a number of search phrases which appeared on them.

In order to handle this amount of data, Google has built 11 data centers around the world. These data centers are basically processors, hard drives and computers connected in a specific way. According to Hosch, (2020), heart of Google's operation is built around three parts of code: 'Google File System (GFS), Bigtable, and MapReduce. GFS handles the storage of data across several machines; Bigtable is company's database program; and MapReduce is used by Google to generate higher-level data.' (Hosch, 2020: 1).

Since Google has gone through extraordinary growth, investors' opinion was that Brin and Page needed professional manager. Based on investors' opinion, Eric Schmidt became a chairman and chief executive officer (CEO) of the company. Eric Schmidt, who has a doctorate in Computer Science, performed the same role in a company called Novell Inc. During this era, Brin became president of technology while Page served as the president of products. The trio ran the company as a triumvirate (association of three) and during 2011, Page shifted to the CEO role while Schmidt became an executive chairman. At the same time Brin adopted a position of a director for special projects. In August 2015, Google split into two ventures. The first venture embodied Internet search, advertising, YouTube, and remained under Google. On the other hand, Calico (research company), Nest became separate firms under Alphabet Inc. Following this course of action, Brin moved to a post of the president of Alphabet and Page became its CEO. Schmidt stood within the company as an executive chairman. Sundar Pichai, former senior vice president of product, altered to CEO of

Google. In 2018, Schmidt stepped down from the post of the executive chairman followed by Brin and Page who left their positions in 2019. Both founders remained members of Alphabet's board of directors.

1.1.2   Advertising

As already mentioned above, Google's primary income derives from advertising. During Google's development, advertising focus slowly shifted from newspapers and television to the Internet interface. An example of this phenomenon can be clearly seen during year 2000 newspaper's advertising peak was $64 billion (Hosch, 2020). Eleven years later, this number dropped to $20,7 billion. On the flip side, global online advertising increased from $6 billion in 2000 to more than $72 billion in 2011 (Hosch, 2020).

Google has not underestimated this shift. A large amount of money has been spent each year to attract different advertising companies. In favour of keeping following information organized and clear, Table 1. summarizes all of the acquisitions made by Google. For example, Applied Semantics was sold to Google for $102 million in 2003 (Hosch, 2020). This company developed a service that lets owners of Web sites run various types of advertisements. According to Hosch (2020), another $102 million purchase was executed in 2006, this time Google spent money on dMarc Broadcasting. In the same year, a deal offering $900 million was paid over three years to MySpace.com in return for rights to sell advertisements on their Web page (Hosch, 2020). To this date, the largest acquisition took place in 2007 when Google bought DoubleClick for $3,1 billion (Hosch, 2020). After three years, roughly when mobile advertising began to increase, Google purchased a mobile advertising network called AdMob. AdMob was worth $750 million at that time (Hosch, 2020). Google acquired $400 million company Admeld regarding online advertising in 2011. (Aarington, 2011). Admeld operated an advertising optimalization platform. During 2014, Google expended its acquisition portfolio with MDialog and Alental Technologies. MDialog specializes in bringing profit from advertising on wireless devices such as tablets and other gadgets.   Every acquisition mentioned above was part of Google's plan to expand from search engine business to advertising.

Year	Company	Amount of money in U.S. dollars	reason
2003	Applied Semantics	$102 million	Google acquired company
2006	dMarc Broadcasting	$102 million	Google acquired company
2006	MySpace.com	$900 million	Purchase of advertisement rights
2007	DoubleClick	$3,1 billion	Google acquired company
2010	AdMob	$750 million	Google acquired company
2011	Admeld	$400 million	Google acquired company
2014	Mdialog	The price has not been disclosed	Google acquired company
2014	Alental	The price has not been disclosed	Google acquired company

e

Table 1. Google's advertising acquisitions

Table 1 depicts Google's revenue from advertising. Each year's revenue is in billion of U.S. dollars. There is a visible yearly growth due to the increasing number of the internet users resulting in a broader advertising audience and also due to Google's dominating place on the market. Google is the most used search engine thus, its advertisement can be seen by most internet users.
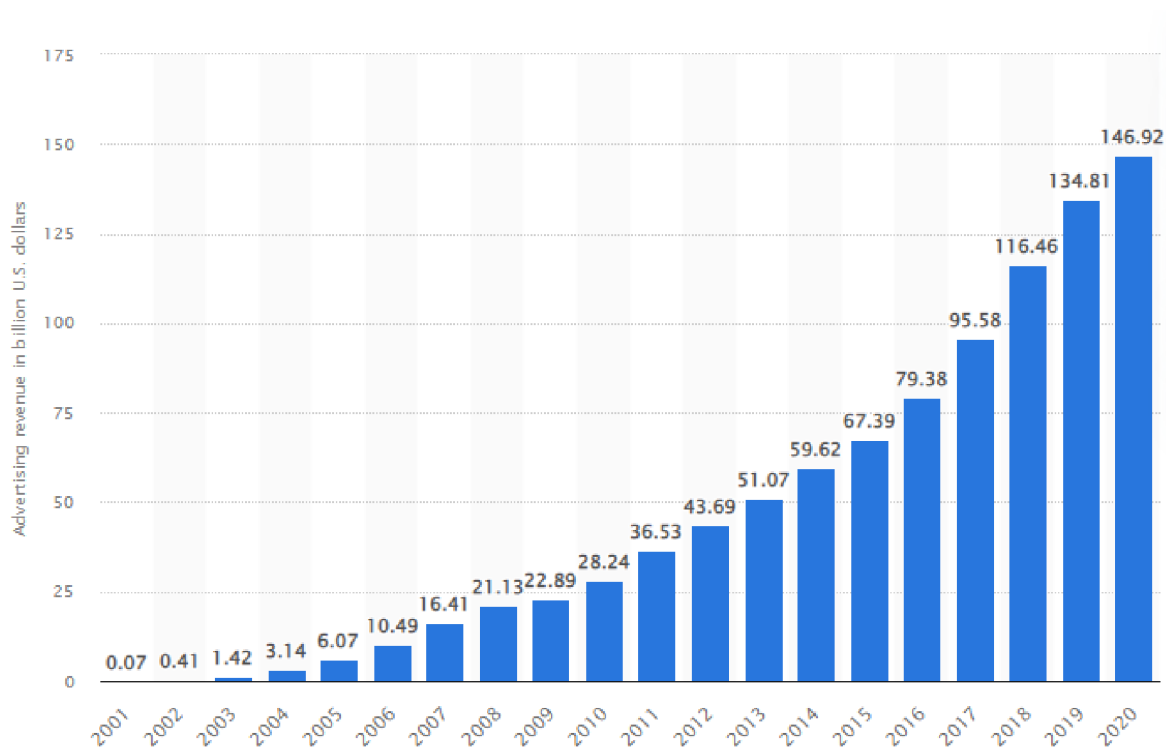

Figure 1. Google's advertising revenue

14

1.1.3 Google Video and YouTube

YouTube is the most visited web page consisting of video content. It mainly consists of shorter videos posted by its users. Content available on this Web page is free of charge, but it also incorporates a lot of space for advertising. According to TechABU, YouTube has approximately 2 billion users that have already registered with this platform. The biggest competitor to YouTube is Vimeo, which has 150 million users (TechABU, 2021).

Google's rapid expansion made an excellent foundation for dominating in Web services. An important part of the Web was sharing video content. Based on these facts, Google decided to launch Google Video in January 2005, which is the platform, where users were able to find television broadcasted content. After a few months later, Google added an option for posting videos that were created by users. Users who posted the video set the prices for each future download. In 2006, Google Video Store was introduced including premium content and other options and later that year, premium content was offered for free but with compulsory advertisements.

Regardless of its marketing advantages, Google was unable to overcome the sharp increase in the popularity of YouTube. The video content on YouTube was usually short videos gaining millions of views. Based on the fact that Google was not close to YouTube's popularity, Google decided to buy YouTube. This happened in 2006, when Google bought YouTube for $1,65 billion in stock. Brin and Pages' company did not merge Google Video and YouTube, they decided to continue to operate both web pages (Hosch, 2020). Finally, in 2012, Google shut down Google Video and removed all its content to YouTube. Despite YouTube's popularity and earnings, Google denied whatever this purchase has been profitable or not (Hosch, 2020).

1.2 Yahoo!

Yahoo! is the oldest of all three companies mentioned in this paper. Yet, once in its lifetime, it was the most successful and promising company. Some of Yahoo!'s

products are viewed as the foundation for most used applications today. Chapter 1.2.1 depicts Yahoo!'s rapid increase in popularity as well as its unexpected downfall.

1.2.1   History

Yahoo! Inc. was created in 1994 by Jerry Yang and David Filo. They started the company as 'Jerry and David's Guide to the World Wide Web' but as it got more popular, Yang and Filo realized its potential and tried to rename it Yahoo. Since there already was a company named Yahoo, the founders decided to add an exclamation mark. Yahoo! was an acronym for 'Yet Another Hierarchical Officious Oracle' (The Editors of Encyclopaedia Britannica, 2020). This event was followed by hiring an expert management team, including Tim Koogle and Jeffrey Mallett. Yahoo! started as a web directory, which was maintained by people called 'the surfers' and their job was to categorize web pages. In 1995 Yahoo! acquired companies such as Rocketmail and ClassicGames.com, which eventually turned into Yahoo! Mail and Yahoo! games. During 1998, Yahoo!, shifted from the web directory to web portal and at the same time, maps, shopping, weather, etc. were added. Yahoo! was the first web directory that gained the first place in other honourable mentions such as the first customizable web portal or the first portal offering localized directories for major cities. Because of its strategic purchases, Yahoo! was well ahead of the time. For example, Yahoo! Briefcase allowed users to store data on cloud servers long before Dropbox and Google Drive. Broadcast.com, later known as Yahoo TV, could be regarded as YouTube's predecessor. The same case with Instagram, Yahoo operated Flicker, which is widely viewed as an ancestor of Instagram. 2001 brought another milestone called Yahoo Music which was a platform intended for streaming music services. Yahoo! bought a start-up called Launchcast and later transformed it into Yahoo Music. Each of these significant acquisitions developed into successful investments.

Yahoo!'s era of wrong decisions started the option to license new search technology at that time worth $1 million. This new search technology will later be known as Google. In 2002, Yahoo!'s CEO Terry Semel made another offer to buy Google for $3 million. Page and Brin refused and expressed that they asked for $5 million (Tynan, 2018). Widely considered as the Yahoo!'s most famous missed opportunity is the offer to buy Facebook for $1,1 billion in 2006. After two years, Yahoo received an offer from Microsoft attempting to acquire Yahoo! for 44,6 billion (Tynan, 2018). Despite

previous unsuccessful efforts, Microsoft and Yahoo! came to an agreement consisting of Yahoo! using Microsoft's search engine Bing for Web site and handling premium advertisements for Microsoft. During this financial struggle, Yahoo!'s leaders decided to hire Marissa Mayer as the CEO. Marissa Mayer introduced several changes into Yahoo!'s working concept. Some of the remarkable changes that took place resulted in the introduction of the programme PB&J, whose intended task was to collect complaints, conversion of remote-working employees to in-office employees and many others. Whereas she significantly participated in the rise of Google, she failed help in Yahoo!'s transformation. The final offer was set in 2016. Verizon Communication offered $4,8 billion for the Yahoo!'s core assets. Unfortunately, Yahoo! faced a series of hyper attacks, that negatively affected more than one billion user accounts. This fact decreased the previous offer to $4,48 and Yahoo! became part of the subsidiary Oath in 2017 (Tynan, 2018).

1.3 Microsoft Bing

Microsoft Corporation is one of the most famous IT companies to this day. It is mainly known for its operating system known as Microsoft Windows. Operating systems are not the only a branch in which Microsoft is specialized, but it provides some other examples like consumer electronics, or personal computers. Microsoft Corporation also owns a search engine Microsoft Bing. Chapter 1.3.1 briefly describes the origin of Microsoft Bing and its consistent rise resulted in the eventual overcome of Yahoo!.

1.3.1 History

Microsoft Bing is a result of several search engines operated by Microsoft before Bing was introduced. MSN Search, Windows Live Search and Live Search could be considered as predecessors. Microsoft Bing offers a variety of search services such as web, video, image or map products. As a successor of the previous search engine, Bing was introduced in May 2009 by Microsoft's CEO, Steve Ballmer. In the same year, Yahoo! and Microsoft made an agreement and part of this agreement was that Yahoo! would use Bing to power search on its portal. Microsoft hoped to renovate search engines by adding a 'decision engine'. It basically displayed more information

than it was typical, resulting in a better base for a decision. The next improvement was the Explore Pane consisting of Quick Tabs, Related Searches and Search History. This panel was displayed on the left side. Yahoo! and Microsoft Bing's agreement was set to last ten years. Since it was 2009, Yahoo! was on fall so Bing powered Yahoo!. Next year, Marks Zuckerberg's web page Facebook was the second most visited site on the World Wide Web. Facebook's managers offered Microsoft to use Bing as a search engine for the searches that were made within Facebook. Despite the fact that Microsoft Bing's market share was low initially, its increase was more consistent and not as drastic as Yahoo! or Google's, Microsoft Bing managed to overcome Yahoo!'s market share. It is almost impossible to find information regarding further development of Microsoft Bing. To this day, Microsoft Bing is the second most used search engine.
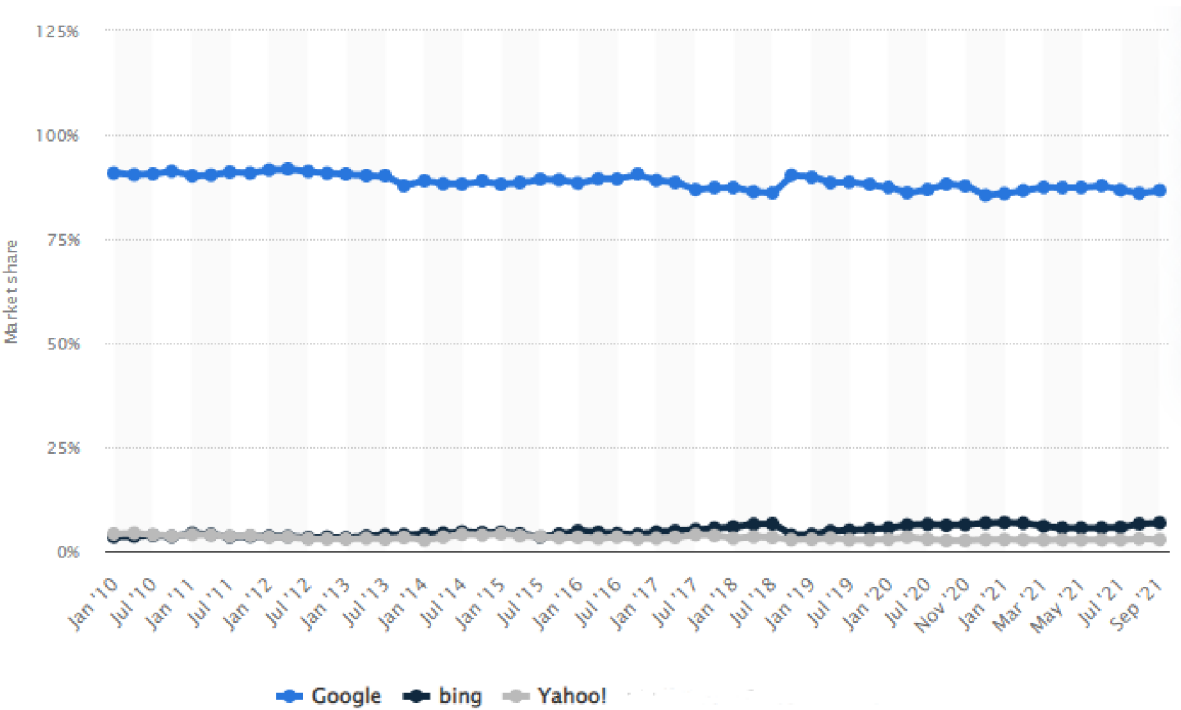
1.4 Global search market in figures

Figure 2. Worldwide desktop market share of Google, Bing, and Yahoo!

Figure 2 depicts a graph with three companies' shares of the global search market where it is clearly visible that Google's share is consistent at around 90%. An exception can be seen in July 2018. Google's share dropped to almost 85%, which was probably caused by the data breach. On the other hand, this event also affected Bing and Yahoo!'s shares. Bing gained an additional 1% of shares. In Yahoo!'s case, change was not that rapid. Since Yahoo!'s shares are continually decreasing, this event resulted in only a slight increase of 0,4%. From 2018 to this day, Google has maintained its share value colliding between 89% and 85%. According to Johanson, (2021), Google's market share was 85,86%, Bing's 6,84% and Yahoo!'s 2,76%.
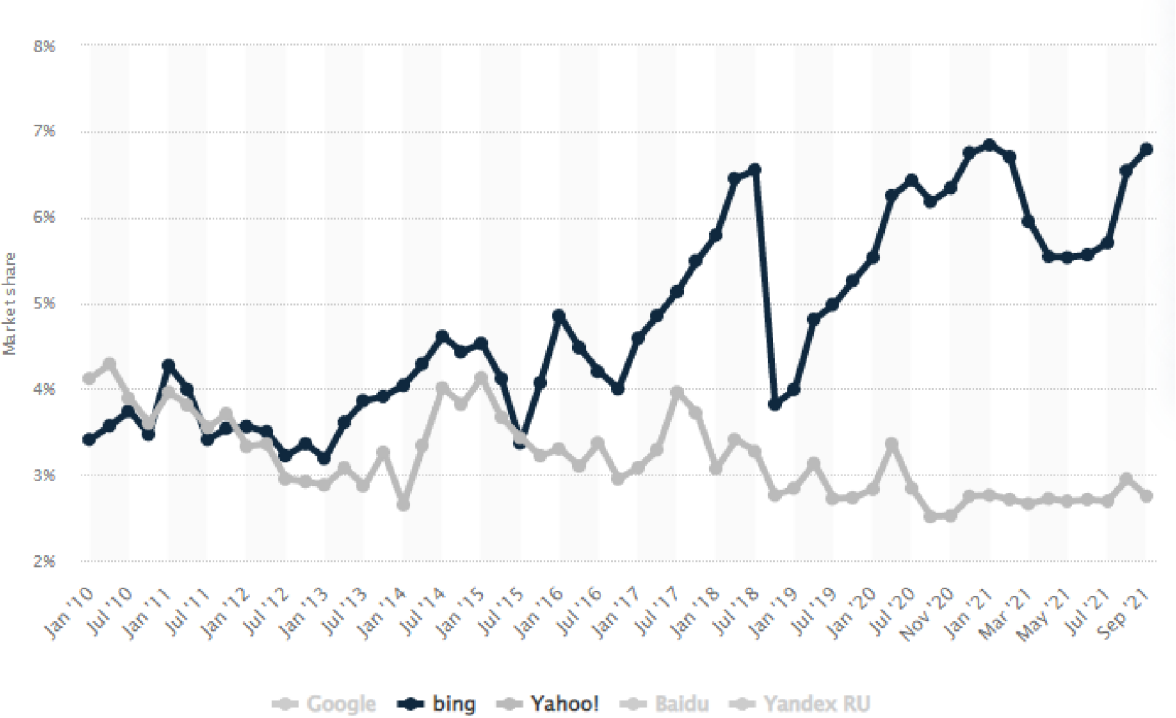
Figure 3. Worldwide desktop market share of Bing and Yahoo!

Yahoo! and Bing's market share is depicted in figure 3. Since Google's market share is not shown, this graph is more precise. During 2010, Yahoo! accounted for 4,2% of the global search market. On the other side, Bing's share represented 3,5%. During 2011 and 2012, Bing and Yahoo! competed repeatedly. From 2012 to 2019, Bing continued to gain users resulting in almost 7% of market share, while Yahoo!

oscillated between 3% and 4% until 2021. Since 2021 Yahoo! has maintained its 2,7% of market share.

2 Google search engine architecture

As it was already mentioned, Google is the most used search engine to this date due to delivering very accurate results and collecting enough information in the fastest time possible. In order to represent importance an example can be used, cutting down Google Maps' size, resulted in an increased number of users over 25% during the following weeks. Based on this fact, chapter 2 focuses on Google's search engine. It is crucial to take into consideration that all information about the search engine is not disclosed now, however, the following chapters try to describe some parts of the system behind simple googling. This process is very complex and contains dozens or maybe hundreds of smaller algorithms.

2.1 Content relevance

There is an almost uncountable amount of data on the Internet and to sort these amounts of data, content relevance has been introduced. Mark Levene wrote that, 'relevance is a relative concept and depends not only on the query but also on the user and the context in which a query is issued' (2005:94). Thus, the search engine takes into account other information such as geographical location, time of the day, and previous history of searches. Relevance can be divided into two groups; the first group is called the binary relevance assessment. This basically means whether the users are interested in searched pages or not. On the other hand, there is the nonbinary relevance assessment, and this group involves ranking each page in terms of how interesting and appealing this page is. The nonbinary relevance, unfortunately, does not work on the Internet because several reasons exist, and to mention the most important, they are namely the amount of data, users' reluctance to give feedback, or query sessions are usually short.

2.1.1 Processing Web Pages

The following paragraphs give an overview of how processing web pages functions. Before the search engine shows searches, all the web pages need to be *indexed*. Indexing basically means parsing the page into words. This algorithm focuses on, for

example, words, the string of numbers or pictures. On the flip side, some words occur on every internet page, and these words are mainly prepositions, definite and non-definite articles, conjunctions or short adverbs. These words are indexed, but this indexing is used under specific circumstances that may happen when the searcher forces them to be included, for example.

The next technique used to make indexing simpler is called *stemming*. This technique removes suffixes to expose the word's root known as the stem. As an example, may serve the word computer whose stem is 'comput' resulting in the same index word for other words such as computation, computers, computing (Levene, 2005). Even though this method decreases the number of index vocabulary, it increases the size of the posting list. Thus, partial stemming is usually used and removes only common suffixes such as -ed and -ing. Google did not support stemming at all until 2003. Since then, it is known that Google has used partial stemming, including singular and plural forms, although the full range of partial stemming has not been disclosed.

A similar feature to stemming is called *truncation*, which is the opposite action to a certain point. If the users search only part of the word, then the search engine finds results starting with this part of the word. Levene (2005) used "comp" as an example, in this case, the search engine will find web pages with computer, company and compare.

2.1.2   Caching queries

Another way to decrease the number of search engine computations is simply storing the precomputed result of specific queries. Logically, this is only used for queries that are searched very often. Levene (2005) used examples such as 'jennifer lopez' and 'harry potter'. Since these topics are searched every day, they need to be refreshed on a regular basis. A special type of side caching, which an individual made on the local device, can be observed. In order to implement this, a user would have to install software allowing local caching.

2.1.3 Phrase matching

Phrase matching is another mean to collect more relevant results for the user. This method compares the distance between keywords. As an example, Levene (2005) uses two words 'computer backgammon' as keywords. In this case, the search engine will find more relative match 'computer backgammon' than 'computer plays backgammon' (Levene, 2005). Since the phrase 'computer backgammon' has a shorter distance between keywords than 'computer plays backgammon' the search engine will deduce it as more relevant.

2.2 Linking

Larry Page and Sergey Brin introduced linking as another mean to judge relevance. Every web page has other links referring to them and web pages with more links are more relevant for users. Google's search engine considers relevance of these links as well. For example, if a user is searching for 'homemade waffles' a web page with two links from a similar topic would be more valuable than a web page with two links from completely different topic. The next criterion is that external links are considered more valuable than internal links. The *internal links* are links that lead into a different part of the same web page. According to Levene's opinion (2005), two types of links exist, the first one is referential, and the second is informational. The referential links usually help navigate the user. These links can have self-explanatory descriptions such as 'clicking this link will download PDF' on the other hand, the informational links concern web page's content. This means that if Page A has an informational link leading to Page B, it is likely that these pages have similar content (Levene, 2005).

2.2.1 PageRank

PageRank, to put into simpler words, is a number and this number represents the web page's popularity. Since Google did not disclose any details, it is a matter of guessing to which point this number affects the final result. However, PageRank does certainly affect the final-outcome. The PageRank reflects the probability of random users

arriving on a certain web page using random links. For PageRank's case, the user's choice of picking which links to follow is only affected by the number of accessible
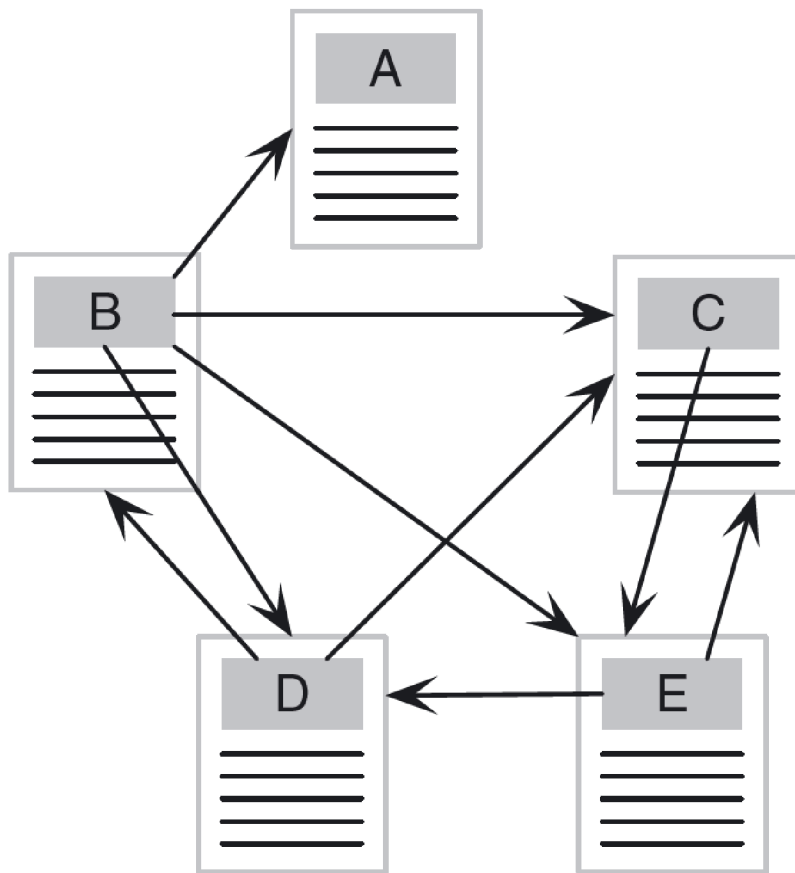

Figure 4. The example of the first error

links. This method involves two possible errors. The first error is called *dangling*. Levene's chart (2005) assumes that web pages are called A, B, C, D, E (see figure 3), and arrows connecting these pages are links that are possible to follow. If the users somehow connect to web page A, they do not have an option to continue the search. The solution to this problem is adding teleportation to this model. This feature will teleport the user to a randomly chosen web page.

A similar problem is called a *rank sink,* which is illustrated in figure 4. If the surfers arrive to web page A, they have only one option: to follow the link to web page B. On web page B, there is a similar situation resulting in moving to web page C. Web page C will ultimately return the users to web page A. This phenomenon will result in a false PageRank calculation since these three pages will be visited quite often. Nevertheless, to solve this issue, teleportation is needed. Even though users do not know what other aspects affect ranking, some successful individuals created illegal

ways to boost ranking values. This led to the birth of mass-market selling a huge number of links.
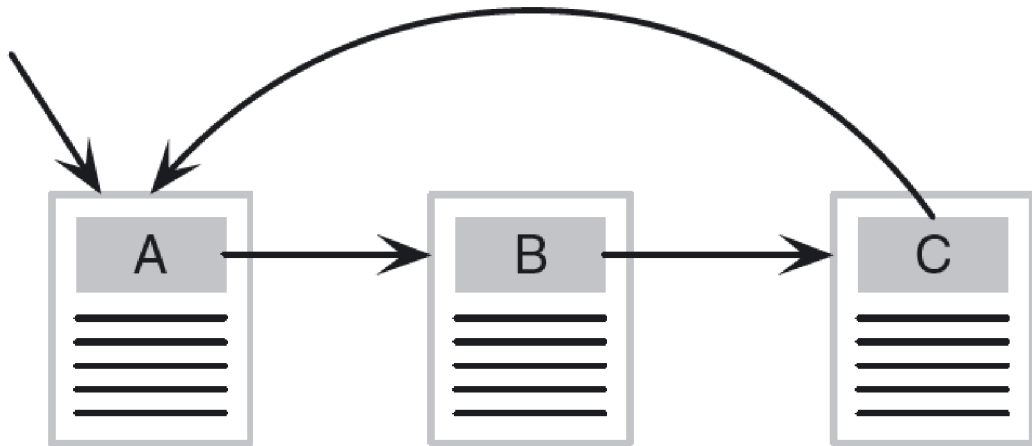
Figure 5. The example of the second error

Google decided to shut PageRank down in December 2013. It is currently unknown if there is any other algorithm that took PageRank's place, however, it is very likely that some of the successors does similar computations.

2.3   Popularity based measurement

Another possible measurement is performed by collecting data that concern the popularity of each web page. To measure popularity, the search engine collects several information, in this case the important information is the number of *clickthroughs,* A clickthrough is defined as an action when the user clicks on a link within the search engine's result page. Naturally, this method can be abused by simply repeatedly clicking on a link. To prevent this type of abusive actions, a specific type of threshold circumstance could be used. Factors such as web page's position on the search engine's list or time spent on the web page are also taken into account in the terms of the popularity measurements.

Another possible problem is the fact that the average user visits only one web page if it concerns something relatively simple, by doing so boosting popularity will occur. One of the possible solutions is to create a dependence between web pages' age and its popularity. Put that in simple words, the web page needs to obtain more users during its lifetime to stay the most popular web page in its category.

2.4　Google's significant algorithms

Even though Google is quite conservative about its search engine, it is proven that the three most important ranking factors are links and content, and RankBrain. Googles' employees did not share which is the most important. The rank of the mentioned factors could be as follows – links, content, and finally RankBrain. Content relevance and links were already described in chapter 2.1 on the flip side RankBrain is covered in chapter 2.4. chapter 2.4 deals with four important algorithms. The first subchapter 2.4.1 focuses on the Panda algorithm, including a thin content, duplicate content, and low-quality content. Basic working concepts and important differences are also mentioned. The following subchapter 2.4.2 interprets the Pinguin algorithm and the Hummingbird algorithm with its vital part RankBrain.

Before 2010, regular algorithm adjustments were made only once per month on average. Whenever search engines graded a web page as the most relevant, resulting in being the first web page on the final search list, it would stay in this order for the whole month. Nowadays, Google modifies its search engine algorithm 600 times per year (Haynes, 2014). Some of these modifications are small and some are more significant. Only the bigger and more significant changes get names. Since Google updates its algorithm very often, Google's search results change too.

2.4.1　Panda algorithm

The Panda algorithm was first launched in February 2011. Panda received its name after one of its creators, Navneet Panda. It focuses on the content of web pages. It intended to place high-quality content web pages in higher search results and also recognize or even punish sites with a low-quality or stolen content. When Panda was released, most of search engine experts, thought that links were also participating in its action. This claim turned out to be very unlikely, which gave birth to the idea that Panda is only about the content of web pages. In most cases, the Panda algorithm evaluates the whole web page as low-quality content but, there have been some cases in which only parts of the web page were rated this way. Despite the fact that no

Google employee has not disclosed specific parts behind the Panda algorithm, there is a list of questions regarding the web page's content. Some examples are: 'Would you trust the information presented in this article? Would you be comfortable giving your credit card information to this site? Does the article describe both sides of story? Would users complain when they see pages from this site?' (Singhal, 2011,1). These questions are not specific aspects of Panda's algorithm, but the question rather serve as a guidelines for web page owners since these questions can contribute to real life-user's experience.

The web page's content labelled as *thin content* is a web page that adds a small amount of information to the topic. Even though it may implicate that the web pages with a smaller number of words are the ones that are likely to have thin content, it does not have to be this case. Certain web pages with many words can contain thin content since some web pages have small diversity regarding their words.

Another possible problem which is likely to emerge is called *duplicate content*. As it already suggests, the first occasion may occur when the content or its majority is copied from another source. The second potential problem is the remarkable similarity of web pages, which often appears on online market web pages. In the eyes of the Panda algorithm, a web page offering the same product in different colours or sizes may look like duplicated content.

Last but not least, it is essential to make comments on *low-quality content*. This phenomenon goes hand in hand with thin content. The main difference lies in the fact that with thin content, the subject or topic is already described, so web pages that add very little to this information are likely to receive the thin content label. On the other side, according to Haynes (2014), low-quality content is usually based on someone's advice to keep updating their website. Users tend to write a huge amount of texts to keep their website updated, resulting in the content that is not connected to the topic or does not describe its depths enough.

Since Google refreshes the algorithm monthly, it may suggest that recovering from the Panda hit is not difficult (Haynes, 2014). Unfortunately, the opposite is true, and usually, it takes a month to do so. Some pieces of advice given to clients that want to recover are: removing doorway pages, writing and re-reading great content, creating mobile-friendly pages or going responsive and, a lot more (Glenn, 2014). Glenn also wrote in his article that 'one of my Panda audits yielded close to 20 pages of recommendations in Word' (2014,1).

2.4.2 Penguin algorithm

The Penguin algorithm was added in April 2012. The purpose of this update was to fight the unnatural number of links. As it has been already mentioned, web pages could unfairly boost their position in the ranking by cheating their number of links. Links are not the only aspect that the Pinguin algorithm checks, another example of other checked aspect by the Penguin algorithm can be web page spamming (Haynes, 2013).

Apart from checking links, The Penguin algorithm focuses on *keyword stuffing,* a self-explanatory phenomenon. If the author intentionally includes a large number of keywords to boost the number of relevant search phrases, this web page can get hit by the Penguin algorithm.

The second but the most important aspect in the eyes of the Penguin algorithm is the *link schemes.* This aspect simply describes everything unfair regarding links. One of the most popular ways to do so is by purchasing backlinks from low-quality or unrelated websites. This may seem pointless, but users observed that a certain number of unrelated links equals a smaller number of related links.

The solution to the Penguin hit is more straightforward than in Panda's case. However, the Penguin algorithm is not refreshing as often as Panda, it is reported to be over eight months (Haynes, 2014). To improve quality in the eyes of the Panda algorithm, web page's authors must identify and remove unnatural links pointing towards them.

2.4.3 Hummingbird algorithm

Hummingbird algorithm's name was inspired by the hummingbird, which is precise and fast. It was announced in September 2013, but it was already running for a month. Hummingbird is like an engine using different parts such as oil filters (in this case, Panda and Penguin) (Sullivan, 2013). Roughly a year before Google announced the introduction of Hummingbird algorithm, Google started to use *semantic search.* This simply means better understanding beyond the meaning of words by understanding

context. Hummingbird is just a tool to improve part of searching queries, that is, in this case long-tail queries. The most significant and visible change for users is, for example, searching query 'best place for chinese'. Since the Hummingbird algorithm focuses on semantics and meaning behind the words, results will not be the best place to live in China or other irrelevant information but the list of restaurants offering Chinese cuisine (see Fig. 6). Based on the fact that Google has access to the user's current location, the search engine will also display locations of restaurants relevant to the user's location. Another proof of semantic search and meaning behind the words is by using query 'suffle'. By simply using keywords in this query, the result could be a definition or maybe some images, however, the Hummingbird algorithm assumed that the user is searching for a recipe, so it offers links with recipes and also displays the table of nutrition (see Fig. 7.).
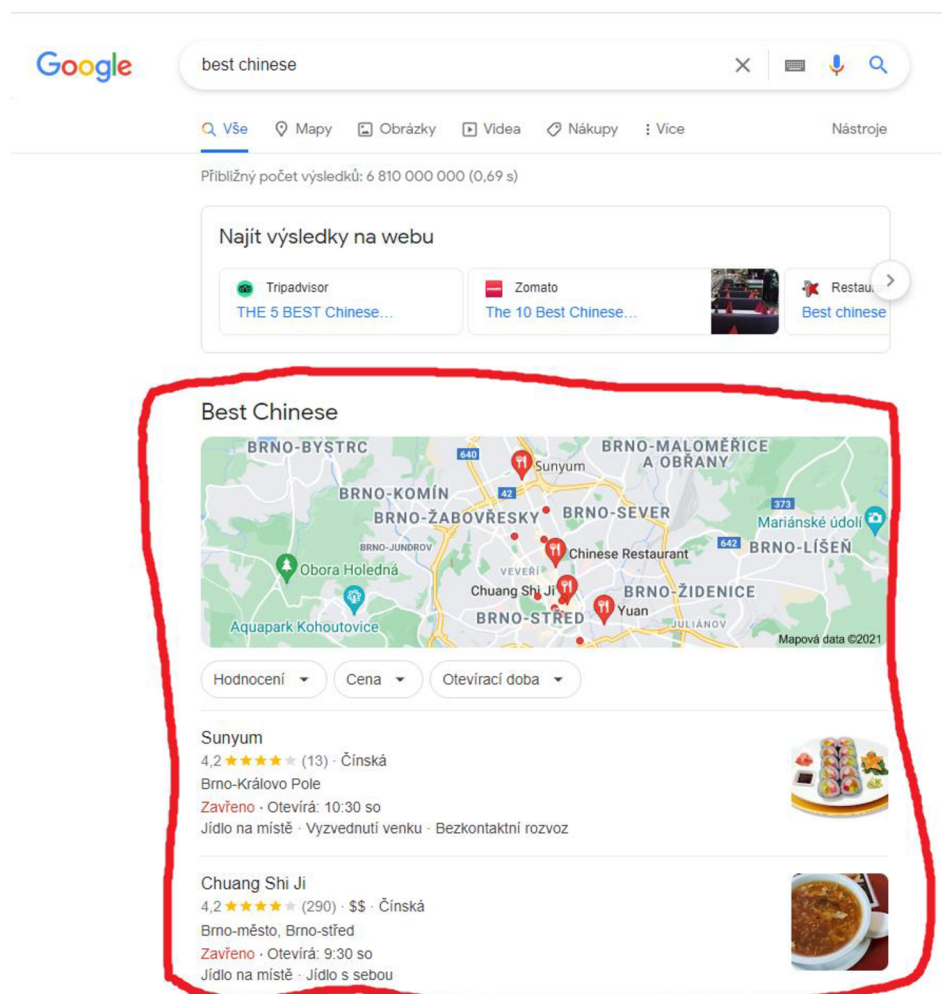

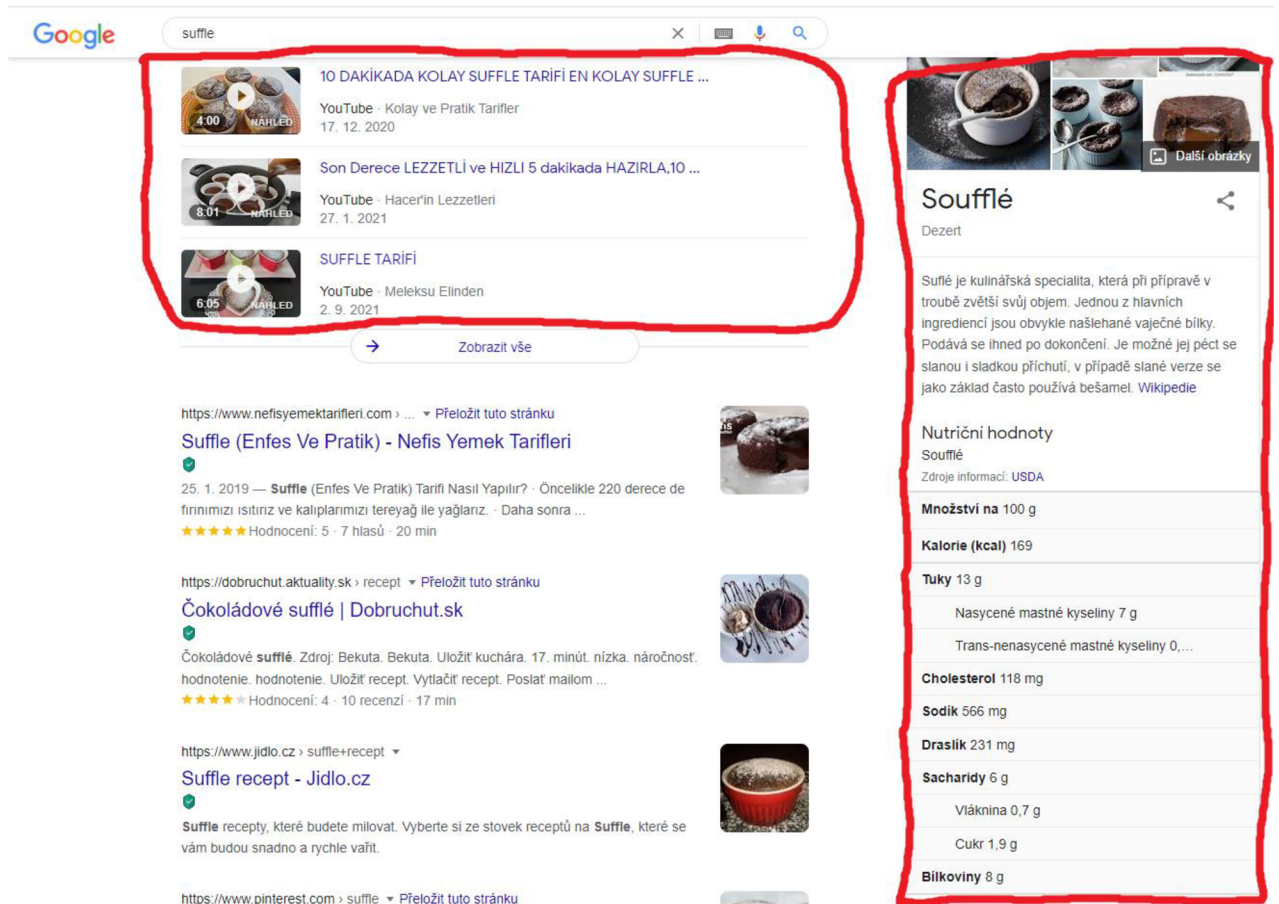*Figure 6.* The first example of the Hummingbird algorithm

Figure 7. The second example of the Hummingbird algorithm

For other examples of the Hummingbird algorithm see appendix section A.

2.4.4 RankBrain

The RankBrain is part of Google's algorithm, and it was introduced in early 2015. It is not entirely clear whether it is part of the Hummingbird algorithm or the Core algorithm. The RankBrain is a *machine learning* and *artificial intelligence* system that processes Google's searches (Sullivan, 2016). Machine learning is a process in which a computer teaches itself to do something. Artificial intelligence is a machine's ability to teach itself and build new connections based on the knowledge that it already possesses or on the knowledge being taught to the machine. As it has been already mentioned, the RankBrain is probably part of the Hummingbird algorithm, intending to deliver the most accurate results to query. As an example of his queries, Fishkin (2016) used 'best Netflix shows', 'What are good Netflix shows', 'best shows on Netflix' and 'What to watch on Netflix' and even though each query uses different language they simply mean the same thing. The searchers expect the same or very similar search results. Google uses different means to determine which results should

be displayed first, such as location or personification, but it also takes into account signals. The RankBrain, put in simple words, says which signals are more relevant in every search case. Fishkin (2016) used this list of signals: Keyword matching, Link diversity, Anchor text, Freshness. Out of these four signals, the freshness is most important since searchers do not want to get the best Netflix series, which is seven years old. It is not publicly known to which degree RankBrain impacts the final result, but this may not seem very important, to clarify this, the simple question 'How many signals are there?' has to be asked. Google's employees usually describe this by hundreds, however, Sullivan (2016) wrote in his article that Google confirmed having more than 200 major ranking signals, while the number of sub-signals may reach 10 000.

3 Microsoft Bing and Yahoo! Search engine architecture

This chapter focuses on the search engine of Microsoft Bing and Yahoo! more than on Google. From the outside, they may appear similar since they are all search engines, and they basically do the same task. Its users type in a particular word or phrase and the search engine finds appropriate websites, images, videos, etc. Based on this fact, the following chapter covers minor similarities and differences, which are unique to each company.

3.1 Differences between Google's and Microsoft Bing's search engine

Bing could be perceived as a better search engine regarding video and image search. The following chapters cover some possible reasons for Bing's popularity. Social integrated results and motivation behind this feature are explained. Google's and Bing's home pages are compared, finally, Microsoft rewards probably one of many Microsoft's ways to lure users is mentioned.

3.1.1 Social integrated results

In 2007 Microsoft paid $240 for a 1,6% stake in Facebook (Wortham, 2010). This acquisition laid the foundation for future possible cooperation. Around the year 2010 Microsoft introduced a new level of interconnection between its search engine Microsoft Bing and Facebook. Microsoft's intentions were simple in terms of connecting the opinions of users' family, friends, and their trusted circle since people usually tend to ask and discuss important topics with them. This new integration contained personalized search results. Bing shows search results that users' friends liked on Facebook. Since Microsoft Bing has the access to data of Facebook likes from users' friends, it changes search results for each user. Finally, Microsoft Bing now shows a similar product that users liked on Facebook. For instance, if users are looking for a camera Microsoft Bing shows them cameras that are popular among camera-oriented groups (Huc, 2011).

3.1.2   Interface

Google and Microsoft Bing's interfaces look alike. They are simple and easy to use, and content regarding they display almost identical results.
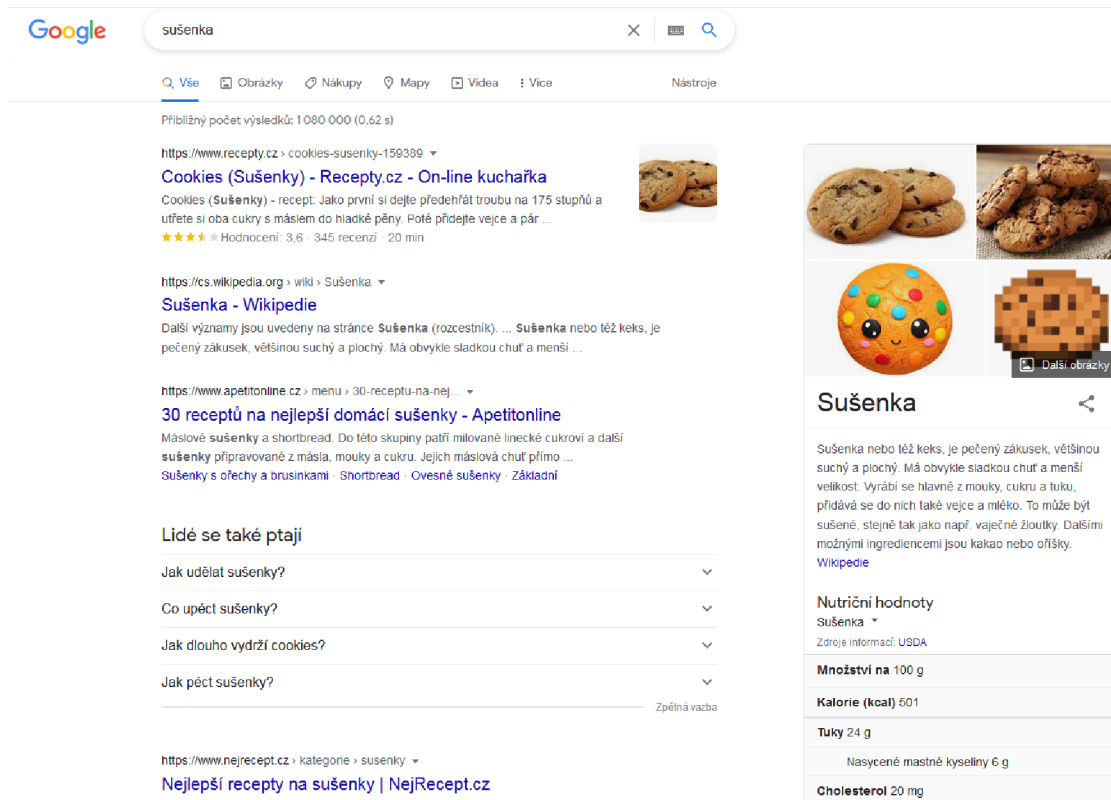

Figure 8. Google's search result of sušenka

As can be clearly seen in figure 8 the right side contains a picture followed by a definition with nutritional values. The search result part consists of receipt links with one link leading towards Wikipedia. In figure 9 the outlay is similar, the picture is followed by a short definition and nutritional values. For further examples please refer to appendix section B) Google and Bing's interface. Search engine results are in a different order but contain the same content. The only visible difference is in the number of results, since Google owns approximately 85% of the market share, and therefore the majority of users tend to use Google resulting in more web pages.
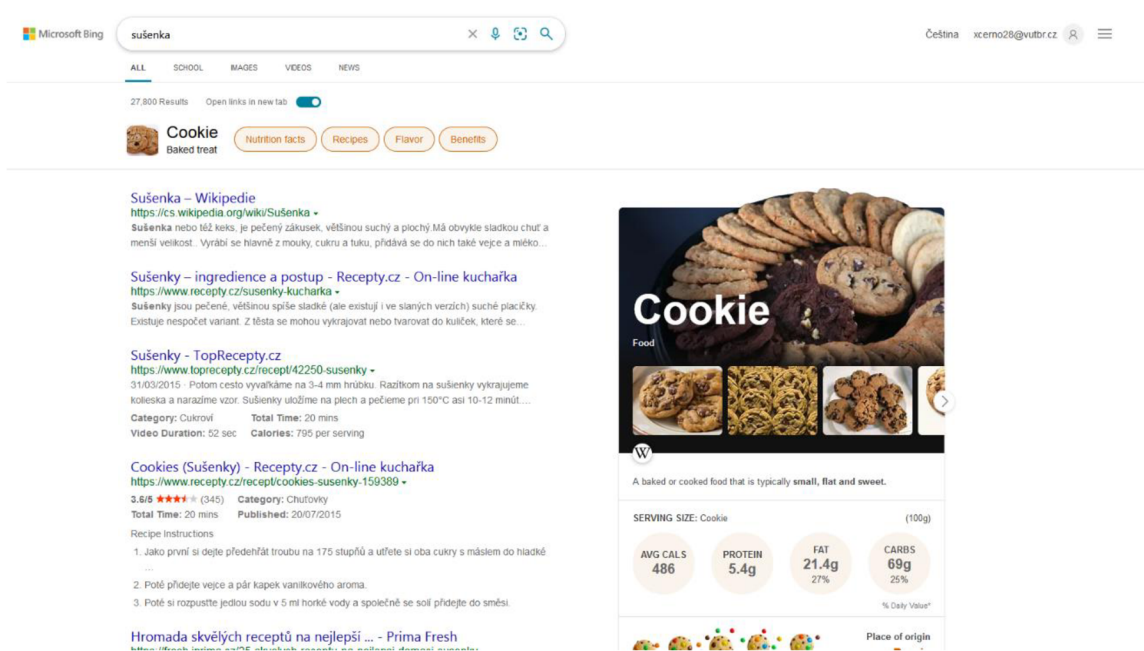
Figure 9. Microsoft Bing's search result of sušenka

The next important feature which Microsoft Bing offers is SEO (search engine optimization) reports. This feature allows authors and managers of a web page to change their web pages in order to place them in a better position. SEO reports are generated automatically every other week.

3.1.3   Image search results

According to several sources, Microsoft Bing is better at searching and displaying more relevant and higher-quality pictures.

Figure 100. Google's results of pizza
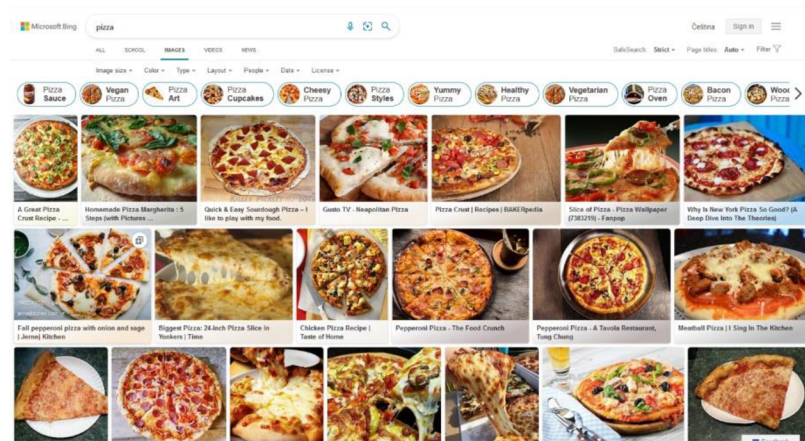
Figure 111. Microsoft Bing's results of pizza

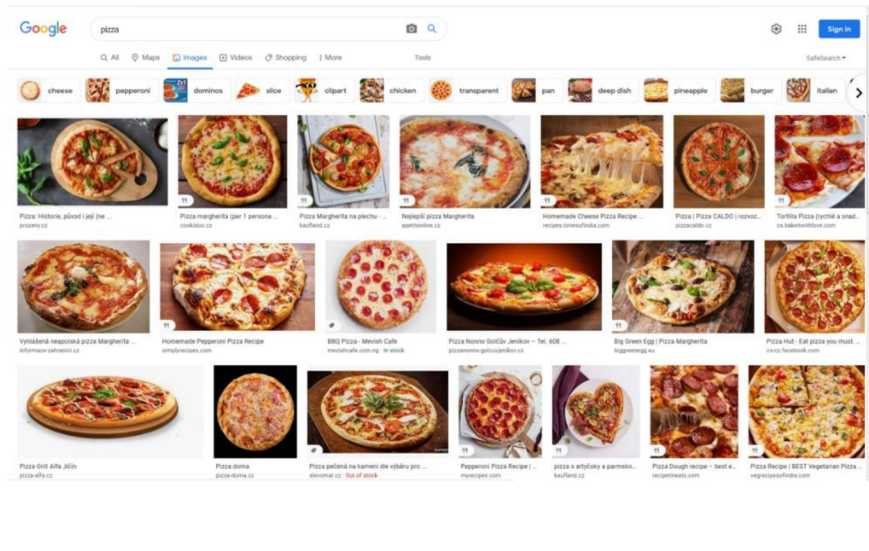Figure 10. and figure 11. are similar to each other. Both search engines created specific categories and displayed various numbers of pictures. Other examples can be seen in appendix, section C) Bing and Google image search similarities. Microsoft Bing offers a filter option that can be used by the user in order to do more specific searches.

Users may find results very similar on the other hand considering the number of signals Microsoft Bing collets show that they are not the same. Firstly, Bing's search engine has to decide whether it should display images or videos, considering the user did not specify this either in a search query or by picking a specific group. Sometimes searched query does not include both so the search engine will display only the

possible solution. On the flip side, some search results demand displaying both, some examples may be famous individuals such as Beyoncé (Barnard, 2020). Another factor that is taken into account is the number of clicks. The option that is more attractive, meaning the majority of users who visit it, will be displayed first (Barnard, 2020).

How does Microsoft Bing's search engine rank images? According to the searchers, relevance is the most important factor. There are other factors but not as important as relevance (Barnard, 2020). Some examples of minor factors are alt tags (alt texts describe what is in the picture), title tag (this is the text which appears when a user hovers over the picture), file name and finally content around the image.

To understand what is in the picture, Microsoft Bing and Google use a machine learning process. Around 2017, machine learning analyzed very specific images such as Tom Hanks, the Eiffel tower, roses, or German shepherd dogs with accurate tags. Since machine learning is exponentially evolving, it does not need well-tagged images and even the analyzing points can be very subtle. As an example, the author shows searching query 'san Francisco seen from Alcatraz', machine learning analyzed images to that degree that it knows which city is San Francisco and even so that it can identify pictures which were taken from Alcatraz Island see figure 12.
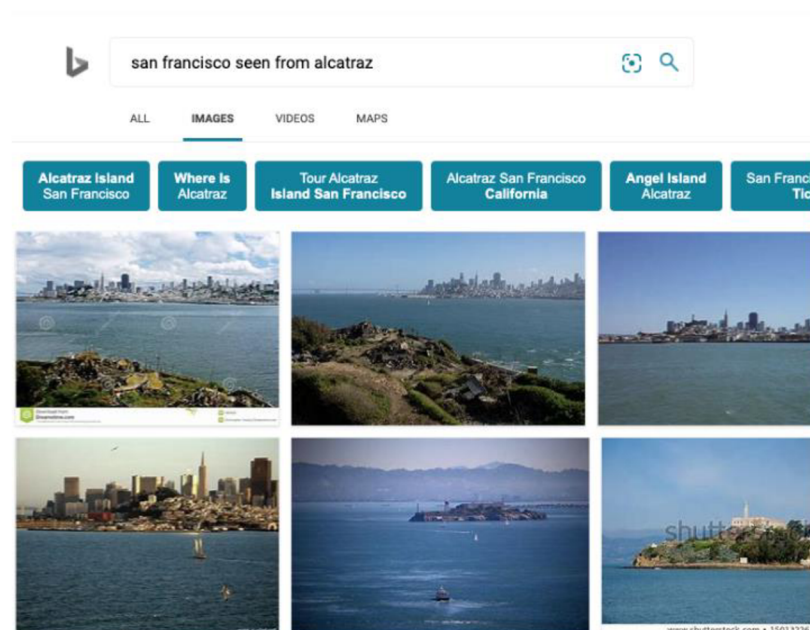

Figure 12. picture of San Francisco from Alcatraz Island

As it has been already mentioned, one of search engines' tasks is to identify what the searcher is looking for. For instance, if the searchers are using the query 'Prince Albert of Monaco' they do not know what he looks like. Microsoft Bing then understands the intent of searchers and will display images of Prince himself instead of Prince Albert of Monaco pictures with his wife as Google does (Merchant, 2013). This difference is shown in Figure 13 and in appendix, section D.

Figure 13. Comparison of Prince Albert of Monaco search results from
Microsoft Bing and Google

Computer vision technologies are another key aspect of understanding and ranking images. One of the machine learning algorithm's goals is to perceive certain images the same way as human brain does. The expression 'tallest peak' is used as an example when, the most searchers would expect images of different tallest mountains. Some potential difficulties may occur when the search engine does not understand correctly

which results in images of graphs, maps, or other mountains that are not the tallest ones.

Microsoft Bing's search engine captures the broader theme (Merchant, 2013). Once again, the example 'photos of Italy', which was employed in this case, means that searchers are probably trying to create an idea of what Italy looks like. Microsoft Bing's search engine identifies themes correctly and shows some typical places of Italy. On the other hand, Google tends to focus more on the words rather than the theme which resulted in a mixture of images containing typical places, maps and flags as is seen in figure 14. For other examples, please follow appendix, section E.

*Figure 14.*Comparison of Photos of Italy search results from Microsoft Bing and Google

Finally, the quality of images should be compared. The overall quality of an image is usually represented by two numbers. The first number corresponds to the number of pixels on the x-axis (number of columns), and the second number represents the number of pixels on the y-axis (number of rows). By multiplying these two numbers the final number of pixels in the image is calculated. Even though this is not the only way to determine image's quality, it certainly is the most used and simple one. According to Microsoft Bing's search engine majority of searchers desire to find higher quality images in the contrast to Google's strategy which does not consider the

quality of the image as important as other factors such as diversity, for example. As shown in figure 15. Furthermore, examples are available in the appendix, section F.

*Figure 15.*Comparison of Photo quality of search results from Microsoft Bing and Google

3.1.4   Video search results

As it has been already illustrated above, in some cases Microsoft Bing automatically displays video search results. News and entertainment are often displayed by video content (Barnard, 2020). Depending on relevance, the search engine decides on the platform the video is used from. Since YouTube is the most popular platform containing the most video content, it has an advantage to a certain degree. On the other hand, some topics demand different platforms, for instance, depicting news is used BBC (Barnard, 2020). Microsoft Bing does not prefer any type of platforms, despite the fact that some platforms have more suitable content for specific topics. YouTube is usually used for entertainment and DIY (do it yourself) purposes, on the flip coin BBC, Fox News are used for news. For the video to be displayed as the video answer to the searchers, there are several important factors. Some of these factors are popularity, the title of the video, and video quality (containing image and sound quality).

Microsoft Bing focuses on three main ranking factors: Quality, Authority, and Trust. Quality is already mentioned above regarding web pages, in the case of video content it does not differ, this basically contains accuracy, standards, etc. Authority is directly

affected by the number of links and organic clicks. Finally, trust is affected by the age of the subject in this case video, or whether it is useful to the audience (Merchant, 2020).

Lastly, the structure of Microsoft Bing videos tends to be more user-friendly. Videos can be filtered and categorized by different means as is typical for Microsoft Bing. The most noticeable for users is the way the videos are spaced out. In the case of Google, they are arranged in the same way as web pages are, see figure 16. If the users decide to click the video link, Google will redirect the users to their chosen link. Conversely, Microsoft Bing spreads out videos evenly with large *thumbnails* 'A video thumbnail is a still image that acts as the preview image for your video' (Knott, n.d., p 1) see figure 17. Another positively perceived fact is that clicking a link does not redirect the users to another web page. For other examples follow appendix, section G.
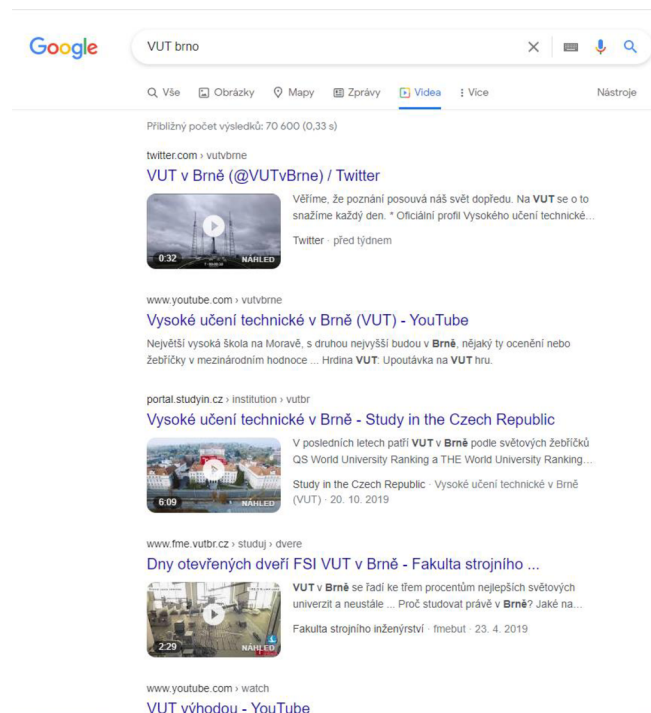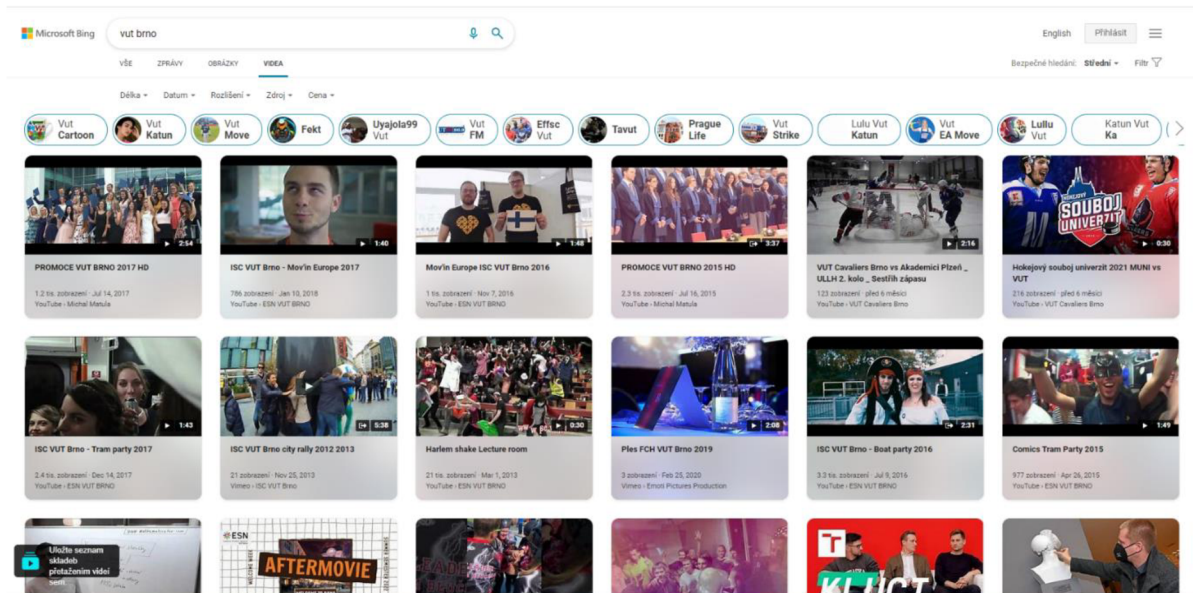

Figure 16. Google videos

Figure 17. Microsoft Bing videos

3.1.5 Microsoft Rewards

Microsoft decided to lure the users in a very odd way, that is by using Microsoft Rewards originally called Bing Rewards. Microsoft Rewards basically pay the users for their data. To use Microsoft Rewards, the users are required to create a Microsoft account and sign for Microsoft Rewards. From this time, the users may complete various tasks to obtain a certain number of points. These points can be later changed for different prizes such as gift cards or charity donations.

Potential Microsoft Rewards users are asked to complete different polls, quizzes, or even play games. Apart from these activities, the users are rewarded for using Bing and setting Microsoft Bing as their default web browser.

Microsoft Rewards, as it is advertised, is completely free to use, on the other hand, the users have to agree to share a lot of data. Ultimately, data is what Microsoft needs for the purpose of improving its search engine. Google owns the majority of the market share, which results in an enormous amount of data that provide a lot of feedback, thus resulting in improvement. On the flip side of the coin, Microsoft possess only fraction of data compered to Google, which means Microsoft obtains smaller amount of feedback ultimately resulting in difficulties regarding improving search engine.

Microsoft Bing and Google's interface look alike with the only major difference being the different order of search results themselves. This is probably caused by the

41

differences in measuring popularity and ranking each web page. Bing's integrated results gather reviews from Facebook friends and display them under search results. Even though, the idea was revolutionary the final outcome did not change the share of users as expected. The following difference is in Bing's image search results. According to users' opinions, Bing is simpler to use and offers images with higher quality and better relevance. Finally, Bing's video search results are considered more user-friendly. This is reflected in better video arrangement or in the possibility of not redirecting users to different web page whenever clicked on a video.

3.2   Differences between Google's and Yahoo!'s search engine

As it has been already covered in Yahoo!'s history, in 2009 Yahoo! and Microsoft announced the deal consisting of Microsoft Bings search results used in Yahoo!'s search results (Miller, 2015). Technically, since 2009 Yahoo! does not own any search engine but as the majority describes it, it is a portal. Despite this fact, there is a certain group of users who prefers to use Yahoo! rather than Google or Microsoft Bing.

3.2.1   Interface

Despite the fact, that Yahoo! uses results from Microsoft Bing, the way Yahoo! presents them is Yahoo's own conception. By comparing Google and Yahoo!'s home pages it is possible to determine a targeted group of users of each company. Google's home page is simple to navigate but to discover other functions such as mail, weather, and translator the users are required to have some basic knowledge of navigating through the web pages. On the other hand, Yahoo!'s home page seems to contain a lot of different components, such as, weather, news, and trending links as seen in figure 18. By choosing different companies' approach to home pages Yahoo! as a tool to search the internet tend to pick people older than 45 years old (Cow, n.d.).
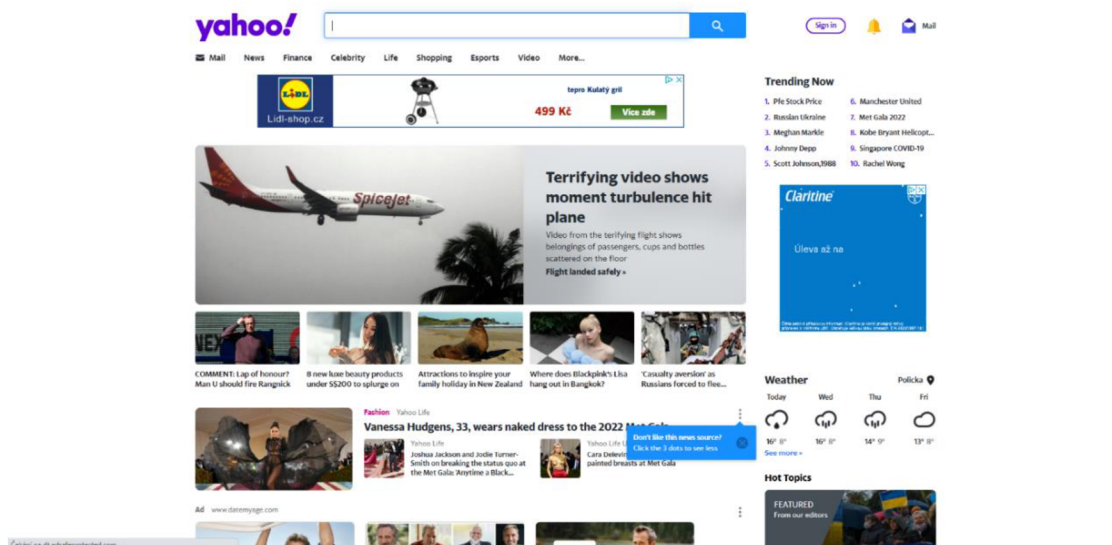

Figure 18. Yahoo!'s home page

43

3.2.2 Yahoo! instant search

As in other cases with Yahoo!, this company has created some useful projects but has decided to not use them to their full potential. One of these examples is Yahoo! instant search, whose name suggests its purpose was to instantly display search results to search queries. These results change after typing following letters. As I already mentioned, this feature was not launched on Yahoo!'s main page but on one of its smaller search engines called Alltheweb.com in 2005 (McGee, 2010). According to some ex-Yahoo! employees, Yahoo! was not able to take the risk and the infrastructure was not sufficient enough. Oliver Bayley presented its use in 2006, which is seen figure 19. The user typed 'jagu' into the search window, Alltheweb's search engine instantly displayed the best search results to 'jaguar', because it assumed that the user was looking for this query. The user could use a certain keyboard shortcut to instantly complete the word. In the case the search engine did not assume the search query correctly, the user may navigate through other potential options by using the mouse or keyboard arrows followed by displaying search results to each search query.
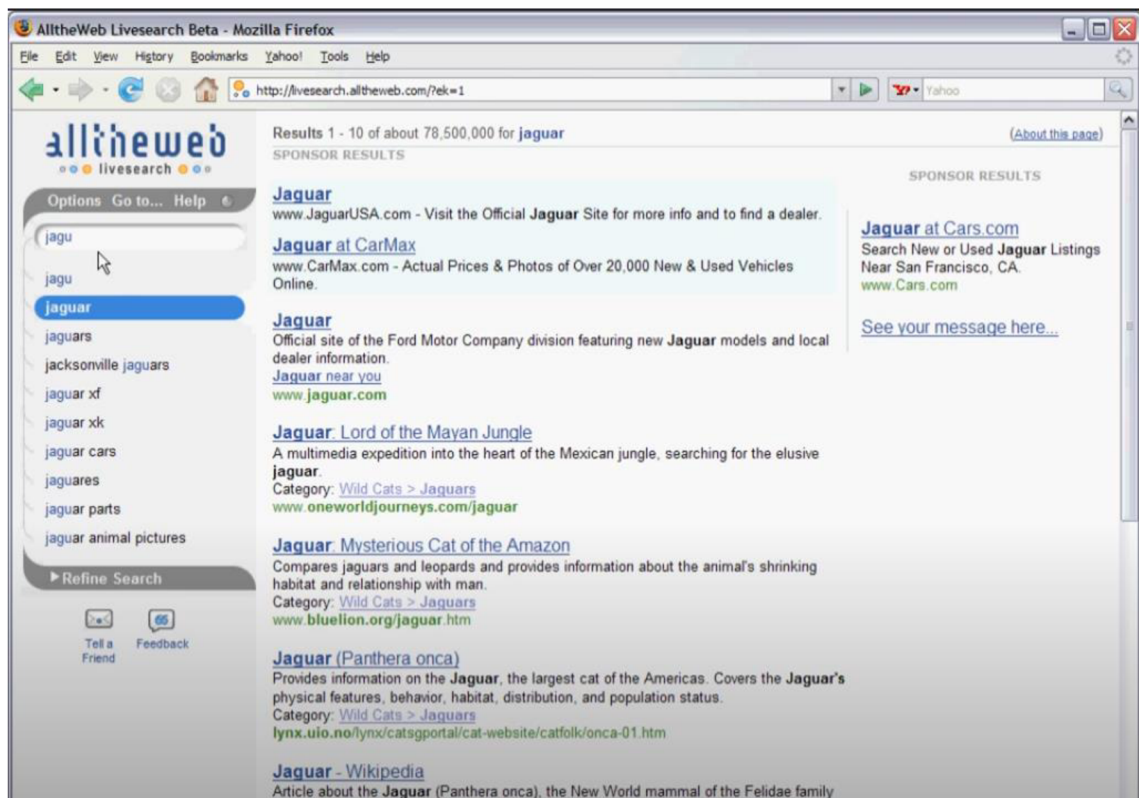

Figure 19. Yahoo! instant search

44

4   Future possibilities

Taking look at the fast development of the internet and search engine, it would be an understatement to say that predicting future trends is very difficult. Take Yahoo! as an example, where at one point at the time Yahoo! was considered one of the most successful IT companies and then followed by the sequence of unfortunate decisions resulting in its infamous fall. On the flip side of the coin, it is safe to say that some of the already functioning features will be improved to an almost perfect state. This chapter talks about voice and visual search, both already function to a certain point, but there is space for improvement.

4.1   Voice search

Voice search is basically pronouncing a search query instead of typing it in the search window. This feature is highly popular on mobile devices. Apple users may communicate with Siri, which is a built-in voice-controlled personal assistant (O'Boyle, 2021). Similar voice assistants from other companies are called Alexa, Cortana, or Bixby. These assistants do a lot more than simply type search queries instead of users they are also able to run or open certain applications or fulfill certain tasks such as starting the timer for 10 minutes. Regarding search engines, nowadays voice search works to a certain point, and that is the opportunity for the users to dictate certain search query and search it but everything else which follows must be done by the users themselves. This includes skimming through web page results or opening certain links. One future possibility may be improving this voice-controlled searching to the point when users do not need any contact with the computer itself but simply pronounce the desired action and the search engine does the task. From my perspective, one of the most difficult obstacles is computers' ability to understand spoken words. This is done by *NLP* (Natural language processing), which is a very complex algorithm combining computational linguistics, based modeling of human language, statistical, machine learning, and deep learning models (IBM, 2020).

4.2   Visual search

As voice search has been already described in the previous chapter, in this case visual search works very similarly. The difference is that the users do not pronounce search queries but rather upload a photograph or instantly take a photograph and the search engine finds corresponding web pages. In the visual search case, the usage is more oriented towards shopping since potential customers usually weigh visual information as more valuable than just word descriptions. In 2018, Snapchat released a feature allowing users to take a photo of a certain object or its barcode and then search for this product on the Amazon website (Toihiri, 2020). Once again, this feature is already working in a certain part of a search engine, but it is not used to its full potential. For visual search, search engine's artificial intelligence ability to recognize pictures is fundamental. Rationally indicating that with the improvement of artificial intelligence there will be significant growth of visual search.

Conclusion

The search engine is a very complicated and mysterious algorithm to a certain point. Even thought that the average Internet user uses a search engine on a regular basis, they usually do not know how it works.

Chapter 1 of the bachelor thesis describes the history of Google, Microsoft, and Yahoo!. Important milestones, probable reasons for Yahoo!'s descent, and crucial acquisitions regarding advertising were discussed.

Chapter 2 focuses on the more practical side of this topic. Firstly, fundamentals of indexing, crawling, and ranking are depicted, followed by part of the chapter 2 that deals with the Panda algorithm, the Penguin algorithm, the Hummingbird algorithm, and the RankBrain. The Panda algorithm is supposed to overlook web pages' content and, to a certain point, penalize those that consist of wrong content. On the other hand, the Penguin algorithm is regarding links. As the same case with the Panda, the Pinguin intends to overlook the correctness of links. Whether they were earned in legal way or not. The Panda and Penguin algorithms are limited only to certain parts of the search engine, the Hummingbird algorithm is similar to an engine, which uses different parts such as filters and pumps (in this case, the Panda and Penguin algorithms). Finally, the RankBrain is a unique feature including artificial intelligence to reach better search results for each user. However, the search on the subject of Google search engine was limited, as there were not many updated sources describing this topic.

Chapter 3's aim is to point out the differences between Google, Bing, and Yahoo!. The differences between Bing and Google are more notable in terms of, more specific images and video search, which is believed to be better or more accurate in Bing's case to a certain point. This chapter shows specific differences such as image quality, more accurate results, or slightly more relevant answers to search queries and represents this in screenshots taken from each search engine.

The second part of chapter 3 deals with Yahoo! and Google differences, since Yahoo! is powered by Bing to a certain point, the difference intertwines with the previous part of the chapter. Yahoo!'s instant search is discussed and presented too.

Finally, chapter 4 briefly introduces the improvement possibility of voice and visual search.

The thesis shows the way Microsoft, Google, and Yahoo! have developed. It also briefly explains the mechanism behind search engines and provides some examples of

specific algorithms included in Google's search engine. Even though search engines are still rapidly developing, they have shown a great deal of progress already. As several times stated above, this topic is rapidly evolving, which caused some difficulties regarding using sources, because some sources were already outdated.

To summarize this thesis, it is important to state that there is not the best and the worst search engine, they all are useful to a specific group of users, and it only makes sense to say, that some groups prefer to use Google, some of them prefer Bing and most certainly some users prefer Yahoo!.

Regardless of the opinion of users, Google owns the majority of the market share, which presents an opportunity for Google to further improve. Google handles more searches every day, resulting in more space for error which can be further transformed into improvement. On the other hand, Bing and Yahoo!, since Yahoo! uses Bing as their search engine nowadays, handle rapidly fewer searches daily, but still, this is a lot of searches every day. Based on this, Google will most likely be more relevant and more accurate regarding more difficult search queries. On the flip side, Bing is believed to handle image and video search better. When everything is taken into consideration, Google will probably stay on top of search engines unless Bing or some other search engine presents some revolutionary feature. Google is used by most people, which makes it better for daily searching, but if the users are keen on image search Bing, may be the better option for them. In conclusion, overall search results do not differ much. All the differences discussed in this thesis were not revolutionary and eye-catching but the analysis aims at details. Based on facts presented in this thesis, Bing can be classified as an alternative search engine to Google, since the majority of search results seems to be very similar, the only notable difference may occur when the users try to use very specific search queries.

Rozšířený abstrakt

Algoritmy známé jako internetové vyhledávače jsou jedním z nejvíce využívaných algoritmů v dnešní době vůbec. Koncovému uživateli poslouží několikrát denně při řešení malých či velkých problémů. Vzhledem k frekvenci používání této technologie je nepřekvapující, že tyto algoritmy jsou několikrát měsíčně upravovány a vylepšovány. Pro každodenní uživatele tyto změny nemusí být očividné a důležité, na druhou stranu pro majitele společností či firem tyto změny jsou velmi důležité. V dnešní době se bez internetové stránky rozroste velmi málo firem, společností a dalších korporací.

Cílem této bakalářské práce je zjistit, zda existuje nějaký internetový vyhledávač, který by mohl kvalitou konkurovat Googlu. Aby bylo možné takový závěr vyvodit, je důležité pochopit, jak takový vyhledávač funguje.

První kapitola se věnuje třem vyhledávačům, Microsoft Bing a Yahoo! společně s již zmíněným Googlem. Tyto vyhledávače jsou velmi pravděpodobně známé i uživatelům, kteří o toto téma jeví malý zájem. Tato práce se prvně věnuje Googlu a jeho historii. Google byl založen Sergeiem Brynem a Larrym Pagem v roce 1998. Nápad, který začal jako školní projekt, byl přes rozmezí několika desítek let přetvořen v jednu z nejúspěšnějších firem moderní doby. Jeden z důležitých faktorů úspěchu je jejich způsob prezentování reklam, reklamy jsou totiž nejvíce vydělávající část firmy Google. V neposlední řadě bylo zmíněno Google Video a YouTube, jelikož YouTube je jednou z nejvíce nebo možná tou nejpoužívanější platformou pro video obsah. Jelikož Google byl rychle rostoucí firmou, jednou z velmi podařených investic, která přispěla k dalšímu rozvoji firmy, byla právě koupě YouTube v roce 2006. Yahoo! bylo založeno v roce 1994 taktéž v té době studenty Jerrym Yangem a Davidem Filou. V rozmezí následujících 15 let bylo Yahoo! považováno za nejúspěšnější firmu. Bohužel Yahoo!, které v roce 2008 dostalo od Microsoftu nabídku za odkoupení společnosti v podobě 44,6 miliardy amerických dolarů, se rozhodlo nabídku odmítnout. I přes všechny snahy Yahoo! začalo ztrácet svoji pozici, a nakonec bylo v roce 2016 prodáno za 4,5 miliard amerických dolarů. V neposlední řadě Microsoft, který byl založen již v roce 1975 Billem Gatesem, představil svůj vyhledávač v roce 2009. Bing již od svého představení pomalu ale jistě nabírá své příznivce a po několika letech získal větší podíl uživatelů než Yahoo!. Jako poslední část první kapitoly jsou předvedené grafy podílu trhu všech tří vyhledávačů.

Druhá kapitola se věnuje pouze vyhledávači Googlu a tomu, jak funguje. Vzhledem k obrovskému množství informací na internetu, existuje systém, jehož účelem je vyhodnotit do jaké míry jsou stránky relevantní k vyhledávanému slovnímu spojení. Praktičtější příklad toho, jak funguje vyhledávač poskytla druhá část kapitoly, která se věnuje konkrétním algoritmům Googlu. Google panda je algoritmus, který kontroluje obsah těchto stránek. Jeho cílem je najít správný obsah, na druhou stranu, pokud algoritmus najde nekvalitní obsah, snaží se ho identifikovat a vlastníky stránky jistým způsobem potrestat. Dalším takovým příkladem je Google Penguin, který se vypořádává s počty odkazů mezi stránkami, funguje na podobném principu. V tomto případě jeho účelem je najít problémy v odkazech mezi stránkami a taktéž takové stránky potrestat. V roce 2013 byl představen Google Hummingbird, tento program využívá algoritmy, jako například Google Penguin a Panda a přiděluje jejich hodnocení celkovou váhu. Jako poslední byl zmíněn RankBrain, tato aktualizace představila a zahrnula do provozu umělou inteligenci a strojové učení. Cílem bylo celkově zlepšit přesnost vyhledávání.

Cílem třetí kapitoly je nalézt a popsat rozdíly a podobnosti mezi Bingem, Yahoo! a Googlem. Výsledky hledání na Googlu a Bingu si jsou strukturou velmi podobné. V obou případech se uživatel dostane k velmi podobným informacím, a dokonce jsou na podobné části stránek. Bing je některými uživateli považován za lepší vyhledávač obrázků. I přes to, že Google vyhledá podobné obrázky, Bing se zdá být více relevantní a prezentuje obrázky s lepší kvalitou. Dalším faktem je způsob vyhledávání videí, Bing se zdá být více uživatelsky přívětivý, protože infrastrukturou se podobá více YouTube. Každé video má svůj náhledový obrázek a při kliknutí na video, prohlížeč pouze zvětší obrázek a zde promítne video. Google k vyhledávání videí přistoupil jako k vyhledávání internetových stránek, tak že při kliknutí na video, je uživatel přesměrován na další stránku s videem. Yahoo! sice vlastnilo svůj internetový vyhledávač, ale mnozí ho nepovažovali za vyhledávač ale spíše za webový adresář. Toto se však změnilo v roce 2009 kdy se firma rozhodla používat výsledky hledání Bingu.

Poslední kapitola se věnuje budoucím možnostem, a to konkrétně hlasovému vyhledávání a vizuálnímu vyhledávání. Obe dvě vyhledávání již fungují ale myslím si, že je zde mnoho prostoru pro zlepšení.

Internetové vyhledávače se aktualizují velmi často, a proto je velmi obtížné odhadnout, zda Bing bude v budoucnosti schopen konkurovat Googlu, na druhou stranu, na základě analýzy provedené v této bakalářské práci, jsem došel k závěru, že rozdíly mezi

Googlem a Bingem nejsou tak patrné. Pro průměrného uživatele budou výsledky velmi podobné, v případě, že uživatel hledá velmi atypické slovní spojení pravděpodobněji nalezne lepší výsledky na Googlu. Podobně, pokud uživatel touží po průměrně lepší kvalitě obrázků lepší možností je používat Bing. Pro uživatele, který se rozhodl hledat videa bude příjemnější používat Bing.

List of references

Advertising revenue of Google from 2001 to 2020. (2020). In Statista. Retrieved from https://www.statista.com/statistics/266249/advertising-revenue-of-google/

Arrington, M. Google Acquires AdMeld For $400 Million. (2011). In TechCrunch. Retrieved from https://techcrunch.com/2011/06/09/google-acquires-admeld-for-400-million/?guccounter=1&guce_referrer=aHR0cHM6Ly9lbi53aWtpcGVkaWEub3JnLw&guce_referrer_sig=AQAAAEeniV7e9oQ2RSYFDvXxgOHN0V2gofufi4djVGjB5UYFJL_nS_4s-iWIUG-Qjh2F9Yx-B8Pq4MjGWrOQLQFhdzMMct9z-kCsTV4yjsbUufH9tw7EQ_C2HLNeWvNBZo2mxHQFB6ENNuSG5YbECGFmkutWxLc9CsS3k3yqvBphxsjR

Barnard, J. How Bing's Image & Video Algorithm Works. (2020) In SearchEngineJournal. Retrieved from https://www.searchenginejournal.com/how-bing-image-video-algorithm-works/364750/

Bing Search Engine: Complete Overview and History of Bing Search Engine. (2020). In Articls Series. Retrieved from https://www.articlss.com/2020/07/bing-search-engine-complete-overview.html

Britannica, T. Editors of Encyclopaedia. Yahoo!. (2020). In Encyclopedia Britannica. Retrieved from https://www.britannica.com/topic/Yahoo-Inc

Burton, W. Google vs Microsoft Bing: A Detailed Comparison of Two Search Engines. (2021) In SearchEngineJournal. Retrieved from https://www.searchenginejournal.com/google-vs-microsoft-bing/400855/#close

Cow, P. A. Google cs Yahoo: Comparing the Top Two Search Enginges. (n.d.). In purple cow. Retrieved from https://purplecowagency.com/google-vs-yahoo-comparing-the-top-two-search-engines/

Drake, S. Google nabs two startups for wireless tech, ad boost. (2014). In Silicon Valley Business Journal. Retrieved from https://www.bizjournals.com/sanjose/news/2014/06/19/google-nabs-two-startups-for-wireless-tech-ad.html

Fishkin, R. Optimalization for RankBrain… Should We Do It? (Is It Even Possible?). (2016). In Moz. Retrieved from https://moz.com/blog/optimizing-for-rankbrain-whiteboard-friday

Glenn, G. Panda 4.0 Analysis / Nuclear Option Rewarded, Phantom Victims Recover, and Industry Experts Rise. (2014). In CSQi. Retrieved from https://www.gsqi.com/marketing-blog/panda-4-analysis/

Google Ranking Indicators and the Death of PageRank. (2016). In Thoughtshift. Retrieved from https://www.thoughtshift.co.uk/google-ranking-indicators-and-the-death-of-pagerank/

Gregerson, E. Bing. (2020). In Encyclopedia Britannica. Retrieved from https://www.britannica.com/topic/Bing-search-engine

Haynes, M. Penguin is not just about bad links – Jon Mueller says there are other factors too. (2013). In Marie Haynes Consulting. Retrieved from https://www.mariehaynes.com/penguin-not-just-about-bad-links/

Haynes, M. Your Google Algorithm Cheat Sheet: Panda, Penguin, and Hummingbird. (2014). In Moz. Retrieved from https://moz.com/blog/google-algorithm-cheat-sheet-panda-penguin-hummingbird

Hosch, W. & Hall, M. Google. (2020). In Encyclopedia Britannica. Retrieved from https://www.britannica.com/topic/Google-Inc.

How Google's Search Engine Rally Works (A Peek Under The Hood). (n.d.). In Neilpatel. Retrieved from https://neilpatel.com/blog/how-google-search-engine-really-works/
Huc, M. Bing social integration with Facebook will help you to make better decision. (2011). In Pureinfotech. Retrieved from https://pureinfotech.com/bing-social-integration-with-facebook-will-help-you-to-make-better-decisions/

IBM, C, E. Natural Language Processing (NLP). (2020). In IBM. Retrieved from https://www.ibm.com/cloud/learn/natural-language-processing

Image SEO: How to optimize your alt text and title text. (2021). In yoast SEO for everyone. Retrieved from https://yoast.com/image-seo-alt-tag-and-title-tag-optimization/

Kalicube. Meenaz Merchant with Jason Barnard – How the Image and Video Algorithm Works at Bing. (2020). Retrieved from https://www.youtube.com/watch?v=kyPc-NqN9uY

Knott, R. What Are Video Thumbnails and Why Do They Matter? (n.d.) In TechSmith. Retrieved from https://www.techsmith.com/blog/what-are-video-thumbnails/#whatare

Levene, M. (2005). An introduction to search engines and web navigation. Hoboken, New Jersey: John Wiley & Sons. Available at: https://onlinelibrary.wiley.com/doi/book/10.1002/9780470874233

McGee, M. Yahoo Had Instant Search In 2005 & Droppped It; Bing Kind Of Has It Now. (2010). In SearchEngineLand. Retrieved from https://searchengineland.com/yahoo-had-instant-search-in-2005-and-dropped-it-50169

Merchant, M. A behind the Scenes Look at How Bing is Improving Image Search Quality. (2013). In Microsoft Bing Blogs. Retrieved from https://blogs.bing.com/search-quality-insights/2013/08/23/a-behind-the-scenes-look-at-how-bing-is-improving-image-search-quality

Miller, C. Google, Bing, Yahoo-What's the Difference?. (2015). In BoostBlog. Retrieved from https://www.boostability.com/content/google-bing-yahoo-whats-the-difference

O'Boyle, B. What is Siri and how does Siri work?. (2021). In Pocket-lint. Retrieved from https://www.pocket-lint.com/apps/news/apple/112346-what-is-siri-apple-s-personal-voice-assistant-explained

Search engine. (n.d.). In Merriam-Webster. Retrieved from https://www.merriam-webster.com/dictionary/search%20engine

Search engine. (n.d.). In Oxford Languages. Retrieved from https://www.google.com/search?q=definition+of+search+engine&oq=definition+of+search+e&aqs=chrome.1.69i57j0i19l7.6921j1j7&sourceid=chrome&ie=UTF-8

Singhal, A. More guidance on building high-quality sites. (2011). In Google Search Central Blog. Retrieved from https://developers.google.com/search/blog/2011/05/more-guidance-on-building-high-quality

Sullivan, D. FAQ: All About The New Google "Hummingbird" Algorithm. (2013). In Search Engine Land. Retrieved from https://searchengineland.com/google-hummingbird-172816

TechABU Top 10 YouTube Competitors and Alternatives [2022]. (2021). In TechABU. Retrieved from https://techabu.co/blog/youtube-competitors/

The Begginer's Guide to SEO. (2021). In Mozz. Retrieved from https://moz.com/beginners-guide-to-seo/how-search-engines-operate

Toihiri, N. The Future of Search with Artificial Intelligence. (2020). In Artefact. Retrieved from https://www.artefact.com/news/the-future-of-search-with-artificial-intelligence/

Tynan, D. The history of Yahoo, and how it went from phenom to has-been (2018). In FastCompany. Retrieved from https://www.fastcompany.com/40544277/the-glory-that-was-yahoo

Worldwide desktop market share of leading search engines from January 2010 to September 2021. (2021). In Statista. Retrieved from https://www.statista.com/statistics/216573/worldwide-market-share-of-search-engines/

Worrall, W. What Is Microsoft Rewards? Everything You Need to Know. (2021). In MakeUseOf. Retrieved from https://www.makeuseof.com/what-is-microsoft-rewards/

Yahoo!. (2008). In Silicon Valley Historical Association. Retrieved from https://www.siliconvalleyhistorical.org/yahoo-history