

**Česká zemědělská univerzita v Praze**

**Provozně ekonomická fakulta**

**Katedra informačního inženýrství**



**Diplomová práce**

**Automatizace publikovaných článků s Google Scholar**

**Bc. Jan Jiroušek**

**© 2020 ČZU v Praze**

## ZADÁNÍ DIPLOMOVÉ PRÁCE

Bc. Jan Jiroušek

Systémové inženýrství a informatika  
Informatika

Název práce

**Automatizace publikovaných článků s google scholar**

Název anglicky

**An Automatization of published articles for the google scholar**

---

### Cíle práce

Cílem práce je elektronicky propojit recenzovaný mezinárodní e-časopis Slovjani.info se systémem Google Scholar. Dílčí cíle práce jsou optimalizace stránek článků pomocí SEO a provedení testů na funkčnost webových crawlerů.

### Metodika

Práce bude obsahovat dvě části. Teoretickou a praktickou část. V teoretické části bude studie problematiky optimalizace zdrojového kódu pro crawlery, studie funkčnosti systému Google Scholar a dalších nástrojů a technik, které budou použité v praktické části. Praktická část bude obsahovat optimalizovací zdrojového kódu stránek publikovaných článků pro crawlery. Pro tuto optimalizaci budou využity technologie PHP, XML a HTML. Publikování článku na web bude spouštět skripty, které budou generují optimalizovaný zdrojový kód článku pro crawler. Tento kód bude splňovat veškeré formální požadavky na indexaci do systému Google Scholar.

## Doporučený rozsah práce

60 – 100 stran

## Klíčová slova

google scholar; XML; PHP; SEO; HTML

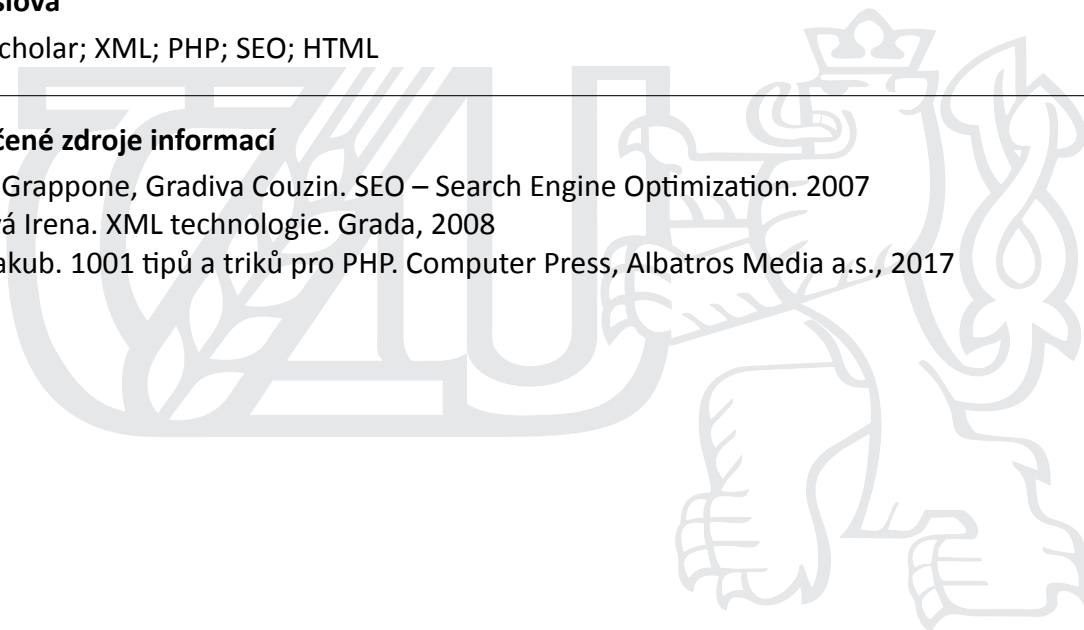
---

## Doporučené zdroje informací

Jennifer Grappone, Gradiva Couzin. SEO – Search Engine Optimization. 2007

Mlýnková Irena. XML technologie. Grada, 2008

Vrána, Jakub. 1001 tipů a triků pro PHP. Computer Press, Albatros Media a.s., 2017



---

## Předběžný termín obhajoby

2019/20 LS – PEF

## Vedoucí práce

doc. Ing. Vojtěch Merunka, Ph.D.

## Garantující pracoviště

Katedra informačního inženýrství

---

Elektronicky schváleno dne 11. 3. 2020

**Ing. Martin Pelikán, Ph.D.**

Vedoucí katedry

---

Elektronicky schváleno dne 11. 3. 2020

**Ing. Martin Pelikán, Ph.D.**

Děkan

V Praze dne 25. 03. 2020

### **Čestné prohlášení**

Prohlašuji, že svou diplomovou práci „Automatizace publikovaných článků s Google Scholar“ jsem vypracoval samostatně pod vedením vedoucího diplomové práce a s použitím odborné literatury a dalších informačních zdrojů, které jsou citovány v práci a uvedeny v seznamu použitých zdrojů na konci práce. Jako autor uvedené diplomové práce dále prohlašuji, že jsem v souvislosti s jejím vytvořením neporušil autorská práva třetích osob.

V Praze dne \_\_\_\_\_

### **Poděkování**

Rád bych touto cestou poděkoval svému vedoucímu této práce panu docentu Merunkovi, za jeho podporu a spolupráci. Dále za možnost rozvíjení mezinárodního e-časopisu Slovjani a za poskytnutí možnosti práce s vyhledávačem a nástroji Google. Dále také bych rád poděkoval své přítelkyni Šarlotě Volfové za podporu při psaní této práce.

# Automatizace publikovaných článků s Google Scholar

## Abstrakt

Tato diplomová práce se zabývá optimalizací zdrojového kódu webových stránek s vědeckými články pro vyhledávač Google Scholar. Cílem práce je propojení mezinárodně uznávaného e-časopisu Slovjani právě se systémem Google Scholar za využití optimalizace stránek pro vyhledávače.

V teoretické části je kladen důraz na vysvětlení všech povinností a náležitostí pro indexování vědeckého článku do Google Scholar.

Praktická část se zabývá vytvořením optimalizovaných stránek s články, otestováním validity a strukturovaných dat na stránce, a také propojením stránek s vyhledávačem Google a otestování indexace článků.

Výsledkem této práce jsou optimalizované stránky s články, které jsou dohledatelné skrze vyhledávače Google. Webové stránky této práce s články zatím nejsou indexované v Google Scholar, jelikož publikované články jsou indexovány do Google Scholar jinými systémy. Nově publikované články budou indexované do Google Scholar právě pomocí stránek, které jsou výsledkem této práce.

**Klíčová slova:** SEO; ASEO; Google Scholar; indexace článků; vědecké články; optimalizace zdrojového kódu

# **An Automatization of published articles for the Google Scholar**

## **Abstract**

This thesis deals with the optimization of the source code of web pages with scientific articles for Google Scholar. The aim of this work is to connect the internationally recognized e-magazine Slovjani with Google Scholar using search engine optimization.

In the theoretical part, the emphasis is on explaining all the obligations and requirements for indexing a scientific article into Google Scholar.

The practical part deals with creating optimized pages with articles, testing validity and structured data of pages, linking pages with Google search engine and testing indexing of articles.

The result of this work is optimized article pages which are traceable through Google search engines. There are no articles indexed in Google Scholar yet, because already published articles are indexed to Google Scholar by other systems. Newly published articles will already be indexed to Google Scholar using the pages resulting from this work.

**Keywords:** SEO; ASEO; Google Scholar; article indexing; scientific articles; source code optimization

## Obsah

<b>1</b>	<b>Úvod</b> .....	<b>13</b>
<b>2</b>	<b>Cíl práce</b> .....	<b>14</b>
<b>3</b>	<b>Metodika</b> .....	<b>15</b>
<b>I.</b>	<b>LITERÁRNÍ REŠERŠE</b> .....	<b>16</b>
<b>4</b>	<b>Google Scholar</b> .....	<b>17</b>
4.1	Hirschův index (H-index).....	17
4.2	Počet citací na stránku (CPP) .....	18
4.3	Výhody .....	18
4.4	Nevýhody .....	20
<b>5</b>	<b>Search engine optimization</b> .....	<b>21</b>
5.1	Vyhledávač.....	21
5.2	Zjednodušený princip fungování vyhledávače.....	21
5.3	Indexace.....	23
5.4	SEO nebo PPC.....	24
5.5	Druhá stránka vyhledávání Google .....	24
5.6	Vyhledávací algoritmy .....	25
5.6.1	On-page algorithms .....	26
5.6.2	Whole-site algorithms .....	26
5.6.3	Off-page algorithms .....	27
5.6.4	Additional algorithms.....	27
5.7	Faktory SEO .....	28
5.7.1	On-page faktory.....	28
5.7.2	Off-page faktory.....	30
<b>6</b>	<b>Academic search engine optimization</b> .....	<b>32</b>
6.1	Rozdíl ASEO a SEO.....	32
6.2	Hodnocení akademických vyhledávačů .....	34
<b>7</b>	<b>Hodnocení vyhledávače Google Scholar</b> .....	<b>36</b>



7.1	Relevance .....	36
7.2	Počet citací .....	36
7.3	Autor a název publikace .....	37
7.4	Ostatní faktory .....	37
<b>8</b>	<b>Indexace do Google Scholar .....</b>	<b>39</b>
8.1	Předpoklady pro indexaci článku do Google Scholar .....	39
8.1.1	Validní kód .....	39
8.1.2	Dostupnost pro indexovací roboty .....	39
8.1.3	Samostatnost jednotlivých článků .....	40
8.1.4	Vědecké meta-tagy .....	40
8.2	Reference .....	43
8.3	Microdata .....	44
<b>II.</b>	<b>PROJEKT .....</b>	<b>46</b>
<b>9</b>	<b>Příprava pro generování zdrojového kódu stránky .....</b>	<b>47</b>
9.1	Objekty .....	47
9.1.1	Třída Article .....	47
9.1.2	Třída Author .....	49
9.2	Načtení článku z databáze .....	50
<b>10</b>	<b>Generování zdrojového kódu stránky pro Google Scholar .....</b>	<b>52</b>
10.1	Seznam všech článků .....	52
10.2	Jednotlivé články .....	54
<b>11</b>	<b>Testování .....</b>	<b>61</b>
11.1	Ověření validity stránek .....	61
11.2	Ověření microdata .....	62
<b>12</b>	<b>Propojení do Google Scholar .....</b>	<b>64</b>
12.1	Nastavení Google Search Console .....	64
12.2	Odeslání žádosti o znovuinindexování stránek .....	67
12.3	Ověření indexace stránek Google .....	69

12.4	Registrace repositáře článků do Google Scholar.....	70
<b>13</b>	<b>Zhodnocení výsledků a doporučení .....</b>	<b>73</b>
<b>14</b>	<b>Závěr.....</b>	<b>75</b>
<b>15</b>	<b>Použitá literatura.....</b>	<b>76</b>
<b>III.</b>	<b>PŘÍLOHY .....</b>	<b>78</b>
<b>16</b>	<b>Přílohy .....</b>	<b>79</b>
16.1	Ukázka optimalizovaného kódu webové stránky článku .....	79
16.2	Ukázka celého výstupu z testování microdata .....	80
16.3	Ukázka výsledků vyhledání na Google .....	81
16.4	Ukázka pořadí výsledků s konkurencí.....	82
16.5	Ukázka emailu po registraci repositáře do Google Scholar .....	83

## Seznam obrázků

Obr. 1	Vyhledávač Google .....	19
Obr. 2	Vyhledávač Google Scholar.....	19
Obr. 3	Filtry vyhledávače Google Scholar.....	20
Obr. 4	Průměrný počet citací na umístění ve výsledcích vyhledávače [BEEL & GIPP & WILDE, 2009].....	37
Obr. 5	Výpis výsledku v Google Scholar.....	43
Obr. 6	Zdrojový kód referencí.....	44
Obr. 7	Ukázka zanořování microdata [SLÁDEK, 2010] .....	45
Obr. 8	Ukázka přístupu k vlastnostem objektu .....	47
Obr. 9	Zdrojový kód třídy Article .....	48
Obr. 10	Ukázka iniciace třídy Article .....	49
Obr. 11	Zdrojový kód třídy Author .....	49
Obr. 12	Ukázka načtení článku podle ID .....	51
Obr. 13	Meta informace pro seznam všech článků .....	52
Obr. 14	Generování seznamu článků .....	53
Obr. 15	Načtení článku podle ID.....	54

Obr. 16 Rozdělení klíčových slov podle delimiteru.....	54
Obr. 17 Vědecké meta-tagy článku .....	55
Obr. 18 Generování meta-tagů pro klíčová slova.....	55
Obr. 19 Generování meta-tagů pro autory.....	56
Obr. 20 Definování objektu ScholarlyArticle .....	56
Obr. 21 Generování micro-data.....	57
Obr. 22 Generování microdata - klíčová slova.....	58
Obr. 23 Generování microdata – autoři.....	59
Obr. 24 Generování microdata informace o publikaci .....	60
Obr. 25 Výsledek testu HTML validního kódu.....	61
Obr. 26 Výsledek testování microdata .....	62
Obr. 27 Upozornění ve výsledku testování microdata .....	63
Obr. 28 Meta-tag pro Google Search Console .....	65
Obr. 29 Google Search Console - zpracovávání dat .....	65
Obr. 30 Výsledek vyhledání Google před nastavením parametru id .....	66
Obr. 31 Google Search console - nastavení parametru URL .....	66
Obr. 32 Google Search Console - výsledek nastavení parametru URL .....	67
Obr. 33 Google Search Console - testování publikované URL adresy.....	67
Obr. 34 Google Search Console - výsledek testování publikované URL adresy.....	68
Obr. 35 Google Search Console - požádání o indexování stránek.....	68
Obr. 36 Výsledek vyhledání Google po nastavením parametru.....	69
Obr. 37 Ukázka výsledku vyhledání při zadání "Prohlášenje na paměť 100 let" .....	69
Obr. 38 Registrace repositáře - výběr druhu webové stránky .....	70
Obr. 39 Registrace repositáře - splnění pravidel registrace.....	71
Obr. 40 Registrace repositáře - vložení odkazu na seznam článků .....	71
Obr. 41 Registrace repositáře - příklady článků.....	72
Obr. 42 Ukázka optimalizovaného kódu webové stránky článku .....	79
Obr. 43 Ukázka celého výstupu testování microdata .....	80

Obr. 44 Ukázka výsledku vyhledání na Google .....	81
Obr. 45 Ukázka srovnání pořadí výsledků s konkurencí při hledání "Proglasenje na pamět 100 lět" .....	82
Obr. 46 Registrace repositáře - email po registraci .....	83

# 1 Úvod

Práce se zabývá tématem, jak správně optimalizovat zdrojový kód webové stránky pro indexovací roboty. Toto téma je v dnešní době základem a předpokladem pro úspěšně navštěvovaný web. Optimalizace zdrojového kódu webové stránky dokáže zvýšit návštěvnost a množství potencionálních zákazníků webu pouhým umístěním v horních příčkách výsledků vyhledávání, například ve vyhledávači Google. Právě díky této optimalizaci lze docílit indexace vědeckých článků do systému Google Scholar. Čím více budou vědecké články dohledatelné, tím více budou autoři článků citováni, čímž budou získávat prestiž a uznání.

Výběr téma mé diplomové práce úzce souvisí s mým vedoucím práce, panem docentem Merunkou, s nímž se nám společně podařilo formulovat a rozvinout toto téma za použití článků v mezinárodně uznávaném e-časopisu Slovjani.info, ve kterém pan docent Merunka publikuje. Téma mě zaujalo možností pracovat se světovým vyhledávačem Google a pochopením principu fungování a indexace stránek do vyhledávače.

## **2 Cíl práce**

Cílem této práce je propojení recenzovaného mezinárodního e-časopisu Slovjani.info se systémem Google Scholar. Propojení se zajistí optimalizací zdrojového kódu pro indexovací roboty, s čímž souvisí validní kód. Kód tedy musí být validní a navíc pro propojení s Google Scholar musí splňovat veškeré náležitosti a povinnosti pro indexování do systému Google Scholar. Google bude informován o existenci těchto stránek pro indexaci.

Dílčí cíle práce jsou: za prvé vytvoření optimalizovaných webových stránek článků pro indexovací roboty, které jim budou dostupné a za druhé ověření indexování těchto stránek vyhledávačem Google.

### **3 Metodika**

Teoretická část diplomové práce bude založena na analýze a kompilaci odborných zdrojů a publikací. Jedná se o vědecké práce z oblasti optimalizace zdrojového kódu pro webové stránky a stránky s vědeckými články, principu fungování vyhledávačů, přehledu a souhrnu faktorů při hodnocení stránek indexovacími roboty, využití strukturovaných dat pro webové stránky a specifikování předpokladů pro indexaci článků do Google Scholar.

V praktické části bude popsán způsob přípravy PHP objektů, do kterých budou uložena data článků a autorů z databáze. Dále budou tyto objekty využity pro generování optimalizovaného zdrojového kódu, jenž bude splňovat všechny povinnosti a náležitosti pro indexaci do vyhledávače Google Scholar. Po vytvoření optimalizovaných webových stránek s články, proběhne jejich otestování na validní kód a využití nástroje od Google pro testování strukturovaných dat. Po těchto úspěšných testech budou dále optimalizované stránky s články zaregistrovány do nástroje Google Search Console, který umožní indexování těchto stránek do vyhledávače. Po registraci těchto stránek bude proveden test indexování těchto stránek vyhledávačem. Pokud budou tyto stránky s články k nalezení přes vyhledávač Google, následuje registrování stránky se seznamem optimalizovaných stránek článků do repozitáře Google Scholar.

# **I. Literární rešerše**



## 4 Google Scholar

Google Scholar je akademický vyhledávací nástroj v současné době nejznámějšího vyhledávače společnosti Google. Google Scholar umožňuje vyhledávat mezi akademickými příspěvky na webu. Mezi tyto příspěvky patří vědecké časopisy, abstrakty, vzájemné hodnocení článků, diplomové práce, disertační práce, knihy nebo také technická hlášení univerzit, akademických institucí, vědeckých skupin z celého světa.

Google Scholar se stal celosvětovou bránou k přístupu k akademickým a vědeckým publikacím. Čím větší přístupnost vědeckých a akademických publikací, tím je větší přínosná hodnota pro akademiky a vědce. Ať už tím, že práce autorů příspěvků se dostanou k více čtenářům, tak i tím, že usnadní práci jiným vědcům, kteří pak mohou citovat potřebné informace pro jejich práce [DIJCK, 2010].

Co dělá Google Scholar tak silným nástrojem je jeho funkcionalita citačního indexu jednotlivých příspěvků. Pod každým nalezeným výsledkem ve výsledcích vyhledávače Google Scholar je uvedeno množství citací k danému příspěvku. Z tohoto čísla lze vyjádřit jakousi kvalitu příspěvku. [HARZING & ALAKANGAS, 2016] Podrobněji je tato část popsána v kapitole 6.2 Hodnocení akademických vyhledávačů.

Google Scholar si zaslouží uznání za nástroj, pro objevování vědeckých informací. Mnoho záznamů má informativní úryvky z celého textu pro všechny uživatele a mnoho z nich také nabízí souhrny článků pro kohokoli. Nástroj Google Scholar je volně přístupným ekvivalentem pro jiné známé podobné nástroje, které poskytují indexování a uložení příspěvků do databáze za peníze.

### 4.1 Hirschův index (H-index)

Hirschův index se také udává pod pojmem H-index. Přímá definice je: „A scientist has index  $h$  if  $h$  of his/her  $N_p$  papers have at least  $h$  citations each, and the other ( $N_p$

– h) papers have no more than h citations each“ [HARZING & VAN DER WALL, 2008].

Cílem tohoto indexu je změřit kumulativní dopad vědce podle toho, kolik citací obdržela jeho práce, kterou napsal. Hirsch tvrdí, že H-index je výhodnější než ostatní ukazatelé, jako je celkový počet citací, celkový počet příspěvků vědce nebo počet citací na stránku [Nozuri, 2005].

## 4.2 Počet citací na stránku (CPP)

Počet citací na stránku se počítá jako jednoduchý podíl celkového počtu citací za posledních pět let a počtu publikovaných článků vědce. Do podílu publikovaných článků se počítají pouze samotné publikované články. Nepatří zde náhledy článků, review článku, příspěvky v komentářích.

Výhoda Hirschova indexu spočívá v kombinaci počtu článků a stránek s jejich kvalitou (počet citací na tyto publikace) [Glänzel, 2006]. Akademik či vědec nemůže mít sám o sobě vysoký h-index, aniž by publikoval značené množství příspěvků. To stále ale nestačí. Tyto články musí být citované ostatními vědci či akademiky, aby se započítali do h-indexu. H-index je považován za výhodnější oproti celkovému počtu citací zejména v případě tzv: „one hit wonders“. Jedná se o autory nebo spoluautory příspěvku, který je vysoce citován a současně tito autoři nebo spoluautoři neprokázali jinou vědeckou aktivitu v rámci publikování příspěvků. Proto h-index zvýhodňuje akademiky či vědce, kteří publikují průběžně příspěvky či jiné publikace nad akademiky s minimální publikační aktivitou [Bornmann & Daniel, 2007].

## 4.3 Výhody

Velkou výhodou je, že Google Scholar prohledává zaindexované příspěvky z full text dokumentů nebo pouze z části těchto full textu dokumentů (ikdyž ukazuje pouze jeho úryvek). Dále také Google Scholar prohledává bibliografické data nebo abstrakty příspěvků. [Jacso, 2008]

Čtenáři knihoven, kteří se přihlásili k odběru digitálních archivů vydavatelů, jsou největšími příjemci služby Google Scholar, protože jediným vyhledáním ve vyhledávači Google Scholar vedou k digitálním full-text verzím článků a jejich příloh. Každý uživatel vyhledávače může vidět u nalezeného výsledku bibliografické data, abstrakt (pokud je k dispozici) a současně úryvek s nalezeným hledaným výrazem, který uživatel (čtenář) zadal do vyhledávače.

Další výhodou je nalezení podobných příspěvků na hledaný dotaz a následné srovnání výsledků vyhledávače podle relevance a jejich kvality (počtu citování daného příspěvku).

Google Scholar umožňuje zobrazení odkazu na PDF verzi celého článku v případě, že je článek dostupný zdarma. Pokud je článek dostupný pouze za peníze, Google Scholar odkazuje na stránku, s daným článkem.

Mezi další výhody patří jednoduchost ovládání vyhledávacího nástroje Google Scholar. Google Scholar vypadá totožně jako vyhledávač Google.

The image shows the classic multi-colored Google logo.

**Obr. 1 Vyhledávač Google**

The image shows the Google Scholar logo, with "Google" in its characteristic colors and "Scholar" in a grey, sans-serif font.

**Obr. 2 Vyhledávač Google Scholar**

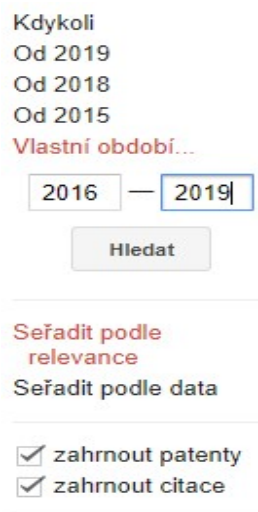
## 4.4 Nevýhody

Hlavní nevýhodu vidím v tom, že v Google Scholar nejsou indexované pouze vědecké či akademické příspěvky. Vždy je důležité si nejdříve přečíst z abstraktu či z úryvku informaci o čem nalezený příspěvek je.

Další nevýhodou je nalezení placených příspěvků či článků. Nalezený článek pak odkazuje na komerční stránku, která vyžaduje peněžní transakci za zpřístupnění nalezeného článku.

Google Scholar neumožňuje vyhledat příspěvky podle typu příspěvku nebo podle oblasti předmětu. Přidání těchto možností by umožnilo uživatelům vyhledat hledaný výraz například pouze v rámci diplomových prací z oblasti vývoje software nebo disertačních prací z oblasti IoT<sup>1</sup>.

Nevýhodou jsou dále nedostatečné filtry pro vyhledávání v Google Scholar. Aktuálně (Říjen 2019) Google Scholar umožňuje filtrovat výsledky pouze podle roku příspěvku.



Kdykoli  
Od 2019  
Od 2018  
Od 2015  
Vlastní období...  
2016 — 2019|  
Hledat

---

Seřadit podle  
relevance  
Seřadit podle data

---

zahrnout patenty  
 zahrnout citace

Obr. 3 Filtry vyhledávače Google Scholar

---

<sup>1</sup> Internet of Things

## **5 Search engine optimization**

Search engine optimization (dále jen SEO) je optimalizace zdrojového kódu webové stránky pro vyhledávače. Do této problematiky spadají postupy, techniky, které nám pomáhají dostat stránky na lepší pozice ve výsledcích vyhledávačů (Google, Bing, Seznam, Yahoo, ...).

Často řeším v práci otázku, co to vlastně SEO je a jak si ho nastavit. Takhle jednoduše to bohužel nelze pojmut, problém nezávisí pouze na těchto dvou otázkách. Abychom mohli se SEO optimalizací nakládat kvalitně, tak musíme znát pravidla a postupy. V první řadě musíme vědět, jak funguje vyhledávač, jak vyhledávač pracuje s naší stránkou a co je to indexace.

### **5.1 Vyhledávač**

Vyhledávač je jakýsi software různých společností, který hodnotí a řadí různé webové stránky podle jejich kvality. Tuto kvalitu určuje na základě vlastního algoritmu, který samozřejmě zná pouze daný vyhledávač. Avšak je důležité dodat, že algoritmy vyhledávačů jsou utajované.

Vyhledávač se snaží uživatelům (tazatelům) na vyhledávací dotazy zobrazovat jen kvalitní výsledky. Pokud by byl veřejně známý všem web masterům a SEO odborníkům, tak by nezobrazoval kvalitní výsledky, ale pouze kvalitně zpracované stránky, na základě tohoto algoritmu. Proto je tento algoritmus utajovaný. [SHARMA, 2015]

### **5.2 Zjednodušený princip fungování vyhledávače**

I přesto, že vyhledávací algoritmus není veřejně známý, jsou jistá pravidla, která se nemění, a které musíme dodržovat na webových stránkách. Tento vyhledávač má robota (crawlera), který navštěvuje naše stránky, indexuje je a tuto činnost provádí pravidelně.

Pravidelná návštěva tohoto indexovacího robota je nezbytnou součástí pro neustálou aktualitu ve výsledcích vyhledávače. Frekvence návštěvy webové stránky robotem je pouze na jeho rozhodnutí. Pokud crawler pozná, že se stránka nemění s postupem času nebo web je málo uživateli navštěvovaný, tak se na danou stránku robot vrátí po delší době. A to v řádech týdnů, měsíců nebo na základě vlastních časových intervalů pro opětovnou návštěvu webové stránky robotem. V opačném případě, pokud robot pozná, že se stránka aktualizuje často, nebo je zde vysoká návštěvnost uživateli, může indexovat stránku i několikrát za den.

Pod indexací si můžeme představit jakési uložení webové stránky do databáze vyhledávače, a to jak aktuální verze stránky, tak i všech předešlých verzí, které jsou již indexovány. Vyhledávač si pamatuje veškeré informace o všech webových stránkách, které indexuje. Po vzájemném propojení stránek a jejich prolinkování následuje v databázi hodnocení těchto stránek a přidělení plusových a minusových bodů. Tímto hodnocením vyhledávač uděluje stránce "rank" (=hodnocení).

Indexovací robot při své práci hledí na mnoho aspektů, které může hodnotit. Robot nehodnotí obsah textu na stránce, ale rozpozná, který text je pro něj důležitý v rámci nadpisů a různých HTML značek. Robot se zajímá o strukturu webové stránky. Hodnotí kód stránky a jeho strukturu kódu a HTML značky.

Dále zkoumá návštěvnost a chování návštěvníků stránky. V případě, že návštěvníci odchází ihned, co stránku navštíví, tak si crawler uvědomí, že na stránce je něco špatně, a tím sníží hodnocení stránky, aby se ukazovala pod stránkou, na které návštěvníci provádějí více akcí nebo na dané stránce vydrží déle. Pro pochopení je tedy nezbytné si pamatovat, že crawler neřeší význam obsahu stránky, on pouze čte kód a text stránky.

Následně seřadí všechny tyto stránky a připraví si jejich pozice ve výsledcích vyhledávání na daný dotaz, frázi nebo jiné klíčové slovo. V posledním kroku zobrazuje uživatelům (tazatelům) výsledky vyhledávání na hledaný výraz ve

výsledcích vyhledávání. Proto je zobrazení výsledků okamžité, jelikož vyhledávač je předpřipraven dopředu [DABAS & PATERIE, 2018].

Takhle zjednodušeně funguje vyhledávač Google. Ale každý vyhledávač funguje jinak, každý má svůj vlastní vyhledávací a indexovací algoritmus. Například v Česku je oblíbený vyhledávač Seznam. Jeho crawler funguje na dvou krocích. V prvním kroku vyše vyhledávač takového předrobota, který navštíví stránku a zjistí, jestli je stránka zajímavá a stojí za to ji indexovat. Pokud se předrobot rozhodne o indexování, pošle teprve pro indexovacího robota, který danou stránku indexuje. V opačném případě nechává stránku neindexovanou.

### 5.3 Indexace

Pod indexací si často lidé mylně představují skutečnost, že crawler navštíví webovou stránku, oskenuje si ji a uloží si ji do své databáze. Teoreticky, částečně i prakticky je to pravda. Ovšem neznamená to, že když robot navštíví stránku, podstránku, kategorii nebo produkt, tak že si ho automaticky uloží a hlavně, neznamená to, že stránku začne zobrazovat ve výsledcích vyhledávání.

Výše zmiňované rozhodnutí crawlera o smyslu indexování stránky je dále výchozí pro jeho další rozhodnutí. A to rozhodnutí zobrazovat tuto stránku ve výsledcích vyhledávání. Pokud se crawler rozhodne, že stránka není zajímavá, nebo že ji svým způsobem již zná, respektive je to obsah, který již někde viděl, tak se nerozhodne, aby stránku umístil na dobré pozici ve výsledcích vyhledávání.

Typickou chybu můžeme spatřit u eshopů. Chyba spočívá v tom, že eshop importuje zboží od dodavatele a současně kopíruje veškeré texty k produktu bez dalších úprav. V tu chvíli se stane, že pokud crawler navštíví tyto stránky, tak pozná, že text u produktu už někde viděl. Pro crawlera tato stránka nebude ničím novým a zajímavým a rozhodne se, že tyto stránky neupřednostní ve výsledcích vyhledávání na tento produkt. Proto je dobré u produktů vytvářet nové a vlastní originální texty. Jinak hrozí, že vyhledávač si bude o těchto stránkách myslet, že jsou nekvalitní a

může je označit za duplicitní. Poté jsou šance na dobré pozici ve výsledcích vyhledávače marné.

## **5.4 SEO nebo PPC**

PPC neboli Pay Per Click je platba za proklik. Inzerenti tu platí pouze za proklik ve výsledcích vyhledávačů nebo na sociálních sítích (nejčastěji na Facebooku) nebo kdekoli jinde na webu. Inzerenti tu platí pouze za skutečné návštěvníky jejich reklamy, nikoli za pouhé zobrazení reklamy [Portent, 2019].

Správně optimalizovaná stránka pro vyhledávač, respektive dobré SEO, má nesmírnou výhodu oproti PPC kampani v tom, že návštěvnost stránek se dlouhodobě zvyšuje, a to bez dalších finančních nákladů.

PPC kampaň je placená reklama. S tím také souvisí efektivita kampaně, která spočívá v tom, že pokud neplatíte, nemáte návštěvnost. Z dlouhodobého hlediska je SEO méně finančně nákladné. Neznamená to ale, že SEO je zdarma. Buď tomu musíme obětovat vlastní čas, nebo zaplatit SEO odborníka. Tento odborník se postará o optimalizaci.

Je obecně prokázáno, že pokud zadáme do vyhledávače dotaz, klikneme spíše na přirozený výsledek (neoznačen jako reklama), než na výsledek, který je označen jako reklamní. Z toho tedy jasně vyplývá, že první přirozené výsledky vyhledávače mají větší důvěryhodnost než zaplacené pozice ve výsledcích [HEITZMAN, 2019].

## **5.5 Druhá stránka vyhledávání Google**

Cílem SEO optimalizace je umístění webové stránky na prvním listě výsledků vyhledávače. A to na základě zadání klíčových slov. Pokud například bude stránka “až na druhé” stránce vyhledávače Google, je zde víceméně jistota, že stránku nikdo nenavštíví. V této souvislosti je trefné zmínit rčení, které se po pochopení problematiky zdá pravdivé. “Pokud nevíte kam schovat to, co nikdo nemá najít, dejte to na druhou stránku vyhledávání Google“.



To potvrzuje studie z roku 2015, která řešila, kolik lidí jde za první stránku vyhledávače Google, v níž bylo zjištěno, že druhou stránku vyhledávání Google navštíví pouze 0,66% návštěvníků. Nicméně podle této studie, ani umístění na první stránce není zárukou navštívení stránky. S tím souvisí zobrazení výsledků na stránce vyhledávače. Například ve vyhledávači google máme jasně viditelných pět nalezených stránek bez nutnosti posunutí po výsledcích na stránce. Šestý výsledek ve vyhledávání navštívilo v roce 2015 3.99% návštěvníků, zatímco první pozici navštívilo 42.30% návštěvníků [HALLUR, 2015].

Tudíž SEO je rozhodně nedílnou a nezbytnou součástí každého webu.

## 5.6 Vyhledávací algoritmy

Všechny části vyhledávače jsou důležité, ale vyhledávací algoritmus je ta hlavní věc, která stojí za správným a chtěným fungováním systému. Přesnější by bylo uvést, že vyhledávací algoritmus je základna celého systému vyhledávače, na níž je postaven zbytek systému. Jak vyhledávač funguje je dané jejím vyhledávacím algoritmem [LEDFORD, 2015].

V podstatě je vyhledávací algoritmus nástroj, který zodpovídá za vyřešení problému ve výpočtu všech možných odpovědí na hledaný výraz a následném vrácení optimálního řešení. Vyhledávací algoritmy v rámci vyhledávačů berou zadání (hledané slovo nebo fráze). Dále zkontrolují databázi a hledají, jestli mají na hledaný dotaz předpřipravené výsledky s URL adresami na stránky, kde se nachází relevance dotazu, popřípadě část obsahu stránky a zde je hledaný výraz nalezen.

Vyhledávací algoritmy vrací výsledky vyhledávání na hledaný dotaz na základě kvality stránek. Jak je tato kvalita stránek vypočítávána je tajemstvím každého algoritmu, ale existují zde jisté faktory, které tuto kvalitu mohou značně ovlivnit. Proč jsou tyto algoritmy hodnocení utajované? Odpověď nalezneme v kapitole „6.1 Vyhledávač“.

Vyhledávací algoritmy jsou řazeny do tří hlavních kategorií: on-page algorithms, off-page algorithms, whole-site algorithms. Každý z těchto algoritmů zkoumá jiné prvky na stránce, přesto, všechny tyto algoritmy jsou součástí jednoho velkého algoritmu [LEDFORD, 2015].

### **5.6.1 On-page algorithms**

Do této kategorie spadají algoritmy, které se zabývají především on-page faktory (více v kategorii 6.7.1 On-page faktory), které se zabývají myšlením uživatele a konkrétně, jestli stojí daná stránka za navštívení. S tím souvisí použití klíčových slov v rámci obsahu webových stránek. Tyto algoritmy také zkoumají korektnost a význam podobných slov klíčovým slovům [BANSAL & SHARMA, 2015].

On-page algoritmy dále zkoumají část obsahu, kterou uživatel na první pohled nevidí. Je zde řeč o meta značkách, které jsou převážně určené pro crawlery [LEDFORD, 2015]. On-page algoritmy dále procházejí faktory, které jsou popsány v kapitole 6.7.1 On-page faktory.

### **5.6.2 Whole-site algorithms**

Pokud On-page algoritmy řeší vztah významu klíčových slov k obsahu stránek, pak whole-site algoritmy řeší vztah jednotlivých stránek na webu. Například jestli jiné stránky mají stejnou obsahovou tematiku jako domovská stránka. Je to důležitý faktor z uživatelského pohledu, protože jestliže uživatel očekává tematiku webových stránek podle domovské stránky, nebude uživatel rád, pokud bude na ostatních stránkách jiná tematika [LEDFORD, 2015].

K zajištění, aby stránky na webu byly opravdu s tematikou, jakou mají být, je dobré použít stejnou architekturu všech stránek. Pokud máme v úmyslu mít na jedné stránce více tématik, je dobré je rozdělit do několika separátních webů.

Architektura stránek je to, jak je stránka použitelná pro návštěvníka na základě nadpisu stránky, který stanoví, o čem daná stránka je. Je to jeden z rozhodujících

faktorů, jak je stránka pro návštěvníka užitečná. Nadpis stránky je jeden z hlavních on-page faktorů na stránce.

### **5.6.3 Off-page algorithms**

Off-page algoritmy přinášejí na webovou stránku další dimenzi. Jedná se o algoritmy, které zkoumají odkazu na původní webovou stránku z jiných webových stránek. Pokud na webovou stránku odkazuje kvalitní webová stránka, vyhledávací algoritmus udělí vysoké skóre původní webové stránce [LEDFORD, 2015]. Více v kapitole 6.7.2 Off-page faktory.

### **5.6.4 Additional algorithms**

V rámci předchozích tří velkých kategoriích vyhledávacích algoritmů existují ještě další, avšak ne tak obsáhlé vyhledávací algoritmy, které stejně tak prochází webovou stránku a podílí se na udělení výsledného hodnocení. Mezi tyto algoritmy patří [LEDFORD, 2015]:

**List search:** List search algoritmus vyhledává mezi specifikovanými daty jediný klíč. Jedná se o jednoduchý lineární vyhledávací algoritmus, jehož výstupem je obvykle jeden element [TORCZON, 1997].

**Tree search:** Podle názvu algoritmu si můžeme představit postup vyhledávání. Postupuje se od kořene přes jednotlivé větve k listům. Tento algoritmus je užitečný pro vyhledávání na webových stránkách [TORCZON, 1997].

**SQL search:** Jednou nevýhodou u Tree search algoritmu je jeho architektura. Pokud se chceme dostat k nějaké větvi, musíme projít skrze jiné větve. To může vést k nechtěnému zvýšení složitosti vyhledávání. V případě SQL search se vyhledává bez přítomné architektury, tudíž můžeme vyhledat data bez procházení celé architektury [LEDFORD, 2015].

Toto jsou pouze základní vyhledávací algoritmy, které se používají ve vyhledávacích. Vyhledávače využívají celou řadu všech možných vyhledávacích algoritmů. Každý vyhledávací algoritmus se využívá pro jiné účely a data.

## 5.7 Faktory SEO

Faktorů, které ovlivňují umístění ve výsledcích vyhledávače je celá řada. Základním rozdělením je on-page faktory a off-page faktory. Jednoduše řečeno, on-page faktory se vztahují k samotné stránce. Off-page faktory se vztahují mimo samotnou stránku [Nagaraju, Dharma, Rajashekar, Mangayarkarasi, 2017].

### 5.7.1 On-page faktory

Spadají sem faktory, které se přímo týkají samotného obsahu, vlastností a kódu stránky. Mezi on-page faktory patří:

Titulek stránky – na samotný titulek kladou vyhledávače velký důraz. A proto značka title patří mezi nejdůležitější značky na celé stránce. Základním požadavkem na titulek je jeho stručnost a výstižnost informace o obsahu dané stránky. Titulek stránky by měl být pro každou stránku unikátní [GOOGLE.COM, 2020]. Viz ukázka kódu níže:

```
<title>Titulek stránky</title>
```

Meta description – tato značka se používá pro stručný popis dané stránky. V rámci této značky by měla být krátká informace o obsahu stránky. Pokud použijeme pro popis stránky vhodná klíčová slova, můžeme zjevně ovlivnit pozici stránek ve výsledcích vyhledávače [GOOGLE.COM, 2020]. Viz ukázka kódu níže:

```
<meta name="description" content="Popisek stránky">
```

Strukturovaná URL adresa – indexovací roboti si libují na strukturované URL adrese. Patří to k dalším úspěchům v horních příčkách vyhledávání. URL adresa by měla být co nejjednodušší a pokud možno srozumitelná. Pro adresu se můžou používat vhodná slova, která odpovídají obsahu stránky. Pomlčka v URL adrese

slouží pro oddělování slov k umožnění větší přehlednosti a výstižnosti odkazu [GOOGLE.COM, 2020].

Ukázka vhodné URL adresy:

`www.domena.cz/kontakt-na-opravare`

Ukázka nevhodné URL adresy:

`www.domena.cz/index.php?id=4684864...`

Strukturované adresy lze dosáhnout pomocí konfiguračního souboru *.htaccess*, který se nachází v kořenovém adresáři webového úložiště. Pokud bych chtěl ukázat překladové pravidlo pro uvedený příklad s URL adresou s kontaktem na opraváře, pravidlo by vypadalo následovně:

```
RewriteRule ^kontakt-na-opravare$ index.php?id=4684864...  
[L,QSA]
```

Parametr QSA uvádí, že se atributy za znakem „?“ mají přenášet taky. Parametr L udává, aby server neaplikoval již žádná překladová pravidla, jakmile přeloží současné pravidlo.

Nadpisy a zvýrazněný text – dobře strukturovaná stránka je taktéž základem úspěchu. K tomu nám slouží nadpisy, které mají odpovídající velikost oproti ostatnímu textu na stránce. Pro nadpisy se používají tagy <h1> až <h6>. S tím souvisí logická hierarchie, protože s velikostí nadpisu se zvyšuje nebo snižuje jeho důležitost na stránce. Pokud se na stránce zvolí špatná hierarchie nebo pořadí těchto nadpisů, může to mít za následek zpomalení vyhledávacího robota a horší indexaci stránky. Například tag <h1> by měl být na stránce pouze jednou [Nagaraju, Dharma, Rajashekar, Mangayarkarasi, 2017].

Pro zvýrazněný text nám slouží tagy <strong> nebo <b>. Je dobré mít na stránce takto zvýrazněné alespoň jedno klíčové slovo.

Popisy obrázků – Indexovací roboti prochází veškerý obsah na stránce, nicméně u obrázků netuší, jaký mají obsah (alespoň zatím, v budoucnu se možná indexovací roboti naučí poznat, co je vlastně na obrázcích). Proto se dává u obrázků popisek,

který by měl stručně vystihnout, co je na obrázku znázorněno. Popisku obrázku se dokáže pomoci atributu alt [GOOGLE.COM, 2020].

```

```

Navigace – je součástí každého webu. A to pro návštěvníky webu a samotné roboty. Navigace pomáhá s orientací na webu. Opět tu platí, že každá položka v navigaci by měla být stručná a věcná.

Klíčová slova – jedná se o ty nejdůležitější slova na stránce, která stojí za to být zvýrazněna. Jsou to ta nejdůležitější a nejvýznamnější slova, která reprezentují obsah webové stránky. O klíčových slovech by se dala napsat samotná kapitola, o jejich analýze, nástrojích a tak dále. Nicméně tato práce je zaměřena na SEO, tudíž stačí ke klíčovým slovům jen pár slov.

Odkazy – každý odkaz by měl obsahovat výstižný text k danému odkazu. Dobře zvolený popis odkazu pomůže crawlerovi napovědět něco o dané stránce odkazu. Opět zde platí stručné a přesné vystižení odkazované stránky. Odkazy můžeme rozdělit na tři kategorie:

1. Interní – odkazovaná stránka je v rámci jednoho webu
2. Externí – odkazuje na jiný web z našeho webu
3. Zpětné – odkazuje z jiného webu na náš (off-page faktor)

### **5.7.2 Off-page faktory**

Off-page faktory v kombinaci s on-page faktory ovlivňují umístění ve výsledcích vyhledávače. Rozdíl mezi nimi je v tom, že off-page faktory se nenachází na daných stránkách. Ovlivnit je lze hůře než u on-page faktorů. Jedná se o zpětné odkazy nebo například zmínky na různých sociálních sítích.

Off-page faktory mají díky své horší ovlivnitelnosti větší váhu než u on-page faktorů. Pokud jsou off-page faktory lepší, tím lépe je pak vyhledávače upřednostňují a umisťují v horních pozicích vyhledávačů.

Zpětné odkazy – jedná se o hypertextové odkazy, které odkazují na danou stránku z jiných stránek. Platí zde jednoduché pravidlo, čím více je odkazů na danou stránku, tím má lepší hodnocení u vyhledávačů. S tím souvisí i kvalita odkazu a kvalita samotné stránky, ze které je na danou stránku odkazováno. Jak je to s kvalitou zpětných odkazů? Patří sem například:

- Text odkazu – jedná se o část odkazu, na kterou lze kliknout
- Příbuznost stránek – kvalitní odkaz je ze stránky, která se zabývá podobnou tematikou jako odkazovaný web
- Stáří odkazu – roboti berou v potaz i stáří odkazů, novější a věcnější odkazy znamenají lepší výsledky indexace

Off-page faktorům určitě stojí věnovat větší úsilí a čas. Je důležité také dodržovat pravidla a nepodléhat praktikám, které by časem mohly dané stránky penalizovat. V poslední době se roboti bojují s praktikou tzv. linkfarmy. Jedná se o umělé weby, které posílají obrovské množství zpětných odkazů na vybrané weby.

To by bylo pro on-page a off-page faktory. Jsou zde vyjmenované pouze ty základní, tudíž známé a v přístupných odborných publikacích dohledatelné. Jinak indexovací roboti využívají něco přes 200 různých faktorů, avšak převážná drtina těchto faktorů je utajovaná [Nagaraju, Dharma, Rajashekar, Mangayarkarasi, 2017].

## **6 Academic search engine optimization**

SEO pro klasické weby je popsáno v předchozích kapitolách. SEO pro odborné články můžeme nazvat Academic search engine optimization (dále jen ASEO). U klasického SEO se optimalizuje webová stránka, také i vědecké články se dají optimalizovat. Respektive optimalizují se odborné články pro vyhledávač Google Scholar.

Stejně jako u normálního vyhledávače Google, tak stejně u Google Scholaru se výsledky zobrazují odshora. Nahoře ve výsledcích jsou výsledky s vysokým hodnocením, podobně jako u klasických webů. Výsledky v horních pozicích Google Scholar nám vypovídají víceméně o kvalitě publikace a hlavně jsou i více citovány. Zobrazení na vyšších pozicích můžeme ovlivnit optimalizací.

Každý vědec, odborník, respektive autor vědecké a odborné literatury by si měl uvědomit, že jejich článek nebo publikace bude indexována různými vědeckými databázemi jako je například Google Scholar nebo digitální knihovna IEEE Xplore, PubMed (tato databáze je zaměřená na biomedicínské odborné články) nebo SciPLore [BEEL & GIPP & WILDE, 2009]. Tyto systémy zvyšují jejich šance na rozšíření publikovaného článku dále do vědecké a odborné komunity. Samozřejmě, nejde tu jen pouze o rozšíření článku dále do komunity, ale taktéž o zobrazení mezi výsledky v horních příčkách výsledku vyhledávání. Protože stejně jako u klasického vyhledávače například Google, většina dotazujících navštíví první nebo druhý výsledek vyhledávání.

### **6.1 Rozdíl ASEO a SEO**

Podle definice Academic search engine optimization je vytvoření, publikace a modifikace odborné literatury způsobem, který usnadňuje akademickým vyhledávačům jak procházet stránky s těmito články, tak je indexovat [CHAKRAVARTY & RANDHAWA, 2006].

Rozdíl ASEO a SEO spatřujeme víceméně ve čtyřech významných ohledech.



První rozdíl je v tom, že v rámci SEO je na trhu gigant nad vyhledávači a to Google [CHITU, 2009]. Tudíž optimalizovat webovou stránku stačí pouze pro crawlery Google. U odborné literatury toto nestačí. U akademických vyhledávačů neexistuje žádná nadvláda, jako u klasických internetových vyhledávačů. Tudíž autoři odborné literatury by měli optimalizovat své publikace pro několik akademických vyhledávačů. Pokud samozřejmě každý akademický vyhledávač má různě odlišný způsob procházení a indexaci publikace, může pak být optimalizace komplikovaná [CHAKRAVARTY & RANDHAWA, 2006].

Druhý rozdíl je ten, že u klasického SEO se webmasteři obvykle nemusí starat, zda jejich webová stránka bude indexovaná. Pokud na jejich stránku existuje link z již indexované stránky, tak crawler tuto stránku zaručeně navštíví a indexuje ji. Situace je jiná u ASEO, kde pouhá malá část této publikované odborné publikace je dostupná na webu a současně dostupná pro akademické vyhledávače jako je například CiteSeer. Většina akademických článků je uložena v databázích nakladatelů. Tyto databáze jsou součástí takzvaného “academic invisible web” [LEWANDOVSKI & MAYR, 2006] a jak akademické tak i normální vyhledávače zde nemají přístup. Pár akademických vyhledávačů jako je například Google Scholar nebo Scirus spolupracují s nakladateli odborných článků. Nicméně tato spolupráce nepokryje všechny publikované odborné články.

Dalším rozdílem je upravitelnost indexovaného obsahu. Webmasteři v případě SEO mohou jednoduše svoji webovou stránku aktualizovat, pozměnit, upravit tak, aby co nejvíce zvýšili hodnocení u indexovacích robotů. To je nesmírná výhoda oproti ASEO. Autoři publikovaných článků si musí uvědomit v rámci optimalizace, že jakmile bude článek publikován a následně indexován, tak už jednoduše upravit nelze. Z tohoto hlediska autoři musí pečlivě dbát na optimalizaci svých publikací [CHAKRAVARTY & RANDHAWA, 2006].

Do posledního rozdílu spadá samotná indexace. V rámci webových robotů se indexuje drtivá většina textu na webové stránce. Oproti tomu u odborných publikací

se u některých akademických vyhledávačů indexuje pouze nadpis publikace a její abstrakt [CHAKRAVARTY & RANDHAWA, 2006]. To znamená, že pokud se autor rozhodne pro dané akademické vyhledávače, stačí mu se soustředit pouze na nadpis a abstrakt. V opačném případě se musí soustředit na celý text publikace.

## 6.2 Hodnocení akademických vyhledávačů

Principiálně způsob vyhledávání všech akademických vyhledávačů je totožný. Tazatel (uživatel vyhledávače) zadá hledaný výraz do vyhledávače, který očekává v nadpisu publikace, abstraktu, jménu autora nebo samotného textu publikace. Tudíž všechny publikace, které mají shodu s hledaným výrazem, jsou vypsány v určitém pořadí ve výsledcích vyhledávače.

Různé akademické vyhledávače využívají různých hodnotících algoritmů, které určují, v jakém pořadí se zobrazují publikace ve výsledcích vyhledávače. Některé akademické vyhledávače umožňují přidat tazatelem vlastní kritérium, podle kterého má výsledky upřednostnit. Do těchto možností patří například datum publikování článku, relevance shody nadpisu s hledaným výrazem, počet citací, které se vztahují k dané publikaci nebo například jméno nebo reputace autora publikace. Některé hodnotící algoritmy všechny tyto faktory sjednotí do jednoho (častější případ vyhledávačů) a poté tazatel nemá šanci si upravit výsledky vyhledávání podle jeho chtěného faktoru [BEEL & GIPP & WILDE, 2009].

Relevance publikace se určuje podle množství, přesněji tedy podle toho, kolikrát je hledaný výraz v publikaci a na jaké části této publikace se vyskytuje. Jednoduše řečeno, kolikrát se hledaný výraz vyskytuje v publikaci a čím více jsou důležitá pole výskytu (nadpis publikace, abstrakt) tím lepší má pak hodnocení. Mezi takováto pole, které akademické vyhledávače hodnotí, patří například: Nadpis, jméno autora/autorů, abstrakt, klíčová slova, samotný text článku, tabulky, anotace, popis, jméno souboru nebo také URI (jednotný identifikátor zdroje) [CHOOCHAIWATTANA, 2017].

U elektronických souborů, jako je například formát PDF, je důležité, aby daný soubor obsahoval metainformace. Protože pokud indexovací robot narazí na tento soubor, tak neví a nedokáže rozpoznat, že se jedná o odbornou literaturu. Tuto informaci mu musíme předat, a to právě pomocí metainformací.

Dále je také důležité si uvědomit, že pokud máme v publikovaném článku tabulky, tak budou indexovány pouze, pokud se jedná o skutečný text nebo vektorovou grafiku. Rastrovou grafiku vyhledávače nejsou schopny indexovat [CHOOCHAIWATTANA, 2017].

## **7 Hodnocení vyhledávače Google Scholar**

Akademický vyhledávač Google Scholar patří k těm vyhledávačům, které kombinují několik hlavních faktorů do jednoho hodnotícího algoritmu. Mezi nejdůležitější faktory hodnocení patří relevance, počet citací, jméno autora nebo autorů a název publikace [BEEL & GIPP & WILDE, 2009].

### **7.1 Relevance**

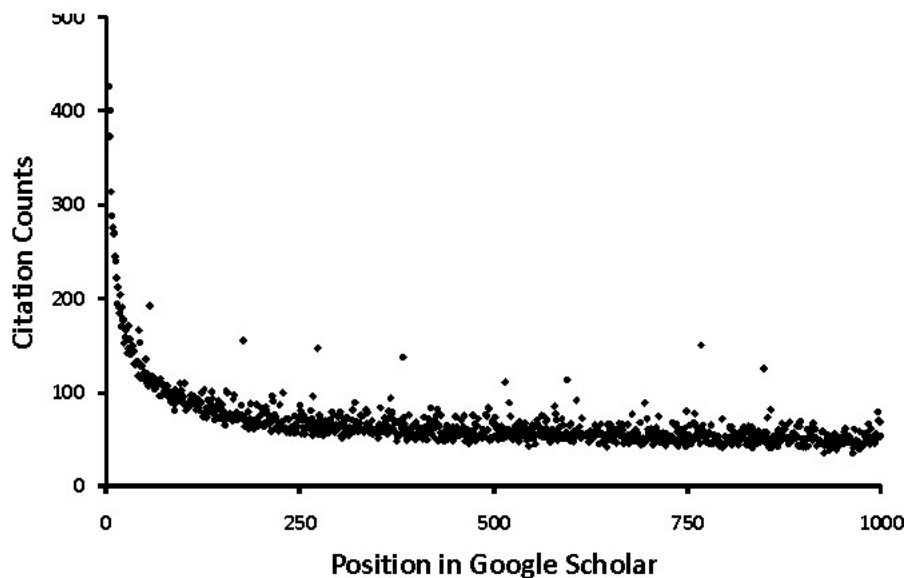
Akademický vyhledávač Google Scholar se převážně soustředí na shodu názvu dokumentu s hledaným výrazem. Publikace, které obsahují hledaný výraz v nadpisu publikace, jsou častěji umístěné v horních příčkách ve výsledcích vyhledávače [MIKKI, 2009].

I přesto, že Google Scholar indexuje celý text publikace, tak se zdá, že množství nalezených shod v dokumentu nemá takový význam na celkové hodnocení pro umístění ve výsledcích vyhledávače [MIKKI, 2009]. Tento jev můžeme vidět na obrázku Obr. 4 Průměrný počet citací na umístění ve výsledcích vyhledávače [BEEL & GIPP & WILDE, 2009].

Google Scholar patří mezi akademické vyhledávače, které indexují i vektorovou grafiku. [BEEL & GIPP & WILDE, 2009]

### **7.2 Počet citací**

Počet citací hraje taktéž velkou roli v umístění publikací ve výsledcích vyhledávače. Více na obrázku Obr. 4 Průměrný počet citací na umístění ve výsledcích vyhledávače [BEEL & GIPP & WILDE, 2009]. Kde je zobrazen průměrný počet citací na příčce v umístění ve výsledcích vyhledávače. Vyhledávač Google Scholar upřednostňuje články, jenž jsou vysoce citovány. Značí to jakousi kvalitu citovaných publikací.



Obr. 4 Průměrný počet citací na umístění ve výsledcích vyhledávače [BEEL & GIPP & WILDE, 2009]

### 7.3 Autor a název publikace

Pokud hledaný výraz v akademickém vyhledávači obsahuje jméno autora či jméno publikace, pak vyhledávač nalezené publikace s daným autorem upřednostní na pozicích výsledků vyhledávání. Jméno autora a samotný název publikace má velkou váhu na hodnocení vyhledávačem [BHUBI, 2019].

### 7.4 Ostatní faktory

Akademický vyhledávač Google Scholar ve standartním vyhledávání nehledí na datum publikace. Přesto Google Scholar nabízí možnosti vyhledání publikací a článků od časového údaje, který si uživatel zvolí.

Dále Google Scholar hledí na reputaci autora publikací. Patří sem skutečnost, jak moc jsou citovány všechny jeho publikace oproti ostatním. V tom případě může vyhledávač upřednostnit publikaci s menším počtem citací oproti publikaci s větším

počtem citací. A to na základě faktu, že autor publikace s menším počtem citací je autorem, který má celkově citovanější publikace [BHUBI, 2019].

## **8 Indexace do Google Scholar**

Google Scholar využívá automatizovaný software nazývaný parsery. Tyto parsery slouží k identifikaci bibliografických dat o publikacích, odkazů a citacích mezi nimi. Nesprávná identifikace bibliografických dat vede ke špatné indexaci webové stránky a tím i publikaci na ní odkázané. Některé dokumenty nemusí být indexovány vůbec, některé mohou být indexovány se špatnými autory nebo se špatným názvem. Na základě špatné indexace poté dochází k horšímu hodnocení akademickým vyhledávačem a neumístění se v horních příčkách výsledků vyhledávání. A to z důvodu špatně indexovaných bibliografických dat, které se neshodují s referencemi na ně z jiných příspěvků či publikací. Pro eliminaci těchto možných problémů stačí poskytnout bibliografické informace správným způsobem, které mohou automatizované parsery přečíst a poté správně indexovat [DIJCK, 2010].

### **8.1 Předpoklady pro indexaci článku do Google Scholar**

Jakmile je napsán odborný či vědecký článek autorem, dochází k dalšímu kroku, a to je online publikace. Pokud autor má v úmyslu indexování jeho publikací do Google Scholar, musí být splněny následující předpoklady [Google Scholar, 2019].

#### **8.1.1 Validní kód**

Validní HTML kód je základním a nejdůležitějším předpokladem pro indexaci článku do Google Scholar. A nejen do Google Scholar, validní HTML kód je vždy předpoklad úspěchu webové stránky [Google Scholar, 2019].

#### **8.1.2 Dostupnost pro indexovací roboty**

Prvním a základním předpokladem indexování do Google Scholar je samotný fakt, že webová stránka, kde jsou články publikovány je dostupná pro indexovací roboty. Dále se předpokládá, že je robot na dané webové stránce schopen články nalézt.

Architektura webu by měla zajistit, že na každou stránku je možné se dostat z domovské stránky pomocí interních odkaz [Google Scholar, 2019].

### **8.1.3 Samostatnost jednotlivých článků**

Dále se do těchto předpokladů řadí umístění každého článku a jeho abstraktu na samotnou webovou stránku a samotného PDF dokumentu. V současné době (září 2019) totiž indexovací roboti nejsou schopni efektivně indexovat několik abstraktů z jedné webové stránky nebo z jednoho PDF dokumentu. Stejně tak roboti neumí rozdělit několik příspěvků či článků z jedné webové stránky nebo jednoho PDF dokumentu do několika samotných článků. Každý příspěvek či publikace by měla mít tedy vlastní stránku a PDF dokument [Google Scholar, 2019].

### **8.1.4 Vědecké meta-tagy**

Meta-tagy (nebo také meta značka) jsou klíčové parametry pro správnou indexaci bibliografických dat. Citace (jak jsem psal v přechozích kapitolách) jsou vlivným hodnotícím faktorem, proto musí Google Scholar rozumět odkazům a citacím mezi články. Toho se dá docílit pomocí odborných meta-tagů [HBUBI, 2019].

Stejně jako u klasického SEO, odborné články a publikace vyžadují pro správnou indexaci celou řadu specifických meta-tagů. Právě díky těmto specifickým (můžeme říct odborným) meta-tagům jsou indexovací roboti schopni správně extrahovat bibliografická data [Google Scholar, 2019].

Existuje celá řada formátů, které se dají použít pro jeden údaj. Záleží pouze na tom, jaký redakční systém daný web provozuje. Google Scholar podporuje následující formáty meta-tagů z několika redakčních systémů (následující zmíněné meta tagy v různých formátech se týkají pouze názvu publikace) [MONPERRUS, 2008]:

- Highwire Press (`citation_title`)
- Eprints (`eprints.title`)



- BE Press (`bepress_citation_title`)
- PRISM (`prism.title`)
- Dublin Core (`DC.title`)

Pro jeden článek lze použít více formátů pro dané meta-tagy. Každý odborný vyhledávač podporuje jiné formáty a na základě většího množství zadaných formátů pro meta-tagy u daného článku si vybere, které podporuje, a které jsou pro jeho systém indexování optimální [BHUBI, 2019].

Mezi tyto základní vědecké meta-tagy, sloužící pro dobrou bibliografickou indexaci, patří meta-tagy níže vypsané. Ty jsou zároveň u vědeckého článku nebo publikace považovány za povinné (v závorce jsou napsány příklady zápisu těchto meta-tagů):

Značka nadpisu\* (`citation_title`, `DC.title`) – musí obsahovat název článku nebo příspěvku. Tento tag se nesmí používat pro název časopisu nebo knihy, ve které byl příspěvek publikován [BHUBI, 2019].

Značka autora\* (`citation_author`, `DC.creator`) – je nutno, aby obsahoval pouze jméno nebo více jmen autorů, kteří se podíleli na tvorbě daného článku. Nesmí se zde uvádět jména autorů webové stránky ani jiné příspěvovatele než autory článku. U každého článku musí být minimálně jeden autor. Každý tag autora odpovídá jednomu autorovi. Pokud článek napsalo více autorů, každý autor bude mít svůj tag autora, a to v řádcích pod sebou. Formát zápisu jména může být například: “Jiroušek, Jan” nebo “Jan Jiroušek”. Tituly a certifikáty by se neměly uvádět do tohoto pole.

Značka data publikace\* (`citation_publication_date`, `DC.issued`) – do této značky patří datum publikování článku. Jedná se o datum, které se bude objevovat v citacích na daný článek z jiných článků. Formát zápisu tohoto data je buď “2019/09/02”, nebo jednoduše “2019”. Oba dva formáty jsou pro parsery v pořádku.

---

\* Tato značka je pro zavedení do Google Scholar povinná

Nesmí se plést se značkou data uložením článku na web (`citation_online_date`) [MONPERRUS, 2008].

Značka odkazu na PDF dokument článku.\* (`citation_pdf_url`, `DC.identifier`) – v odkazovaném PDF dokumentu je veškerý obsah článku nebo publikace. Jedná se o absolutní URL adresu, na které je existující PDF dokument. Považuju tuto značku za povinnou, jelikož Google Scholar v jedné své příručce pro správné indexování zmiňuje, že meta-tagy, konkrétně název článku, datum publikace a autor jsou bez nalinkovaného PDF dokumentu sami o sobě zbytečné [BHUBI, 2019].

Pro časopisné, konferenční články a publikace se musí zadat doplňující bibliografická data v následujících tazích: `citation_journal_title` nebo `citation_conference_title`, `citation_issn`, `citation_isbn`, `citation_volume`, `citation_issue`, `citation_firstpage`, `citation_lastpage` [BHUBI, 2019].

Pro závěrečné, disertační práce a technické práce se musí uvést taktéž další bibliografická data. A to v následujících značkách: `citation_dissertation_institution`, `citation_technical_report_institution` nebo `DC.publisher` pro název instituce a `citation_technical_report_number` pro číslo technické zprávy. Tyto značky jsou důležité pro citaci na danou práci z jiných článků a publikací [BHUBI, 2019].

Všechny značky, které byly zmíněné, jsou oficiálně podporované systémem Google Scholar. Existuje celá řada dalších meta značek, které nejsou plně funkční a oficiálně podporované Google Scholar. Tedy jejich efekt na indexaci je zpochybnitelný.

Mezi neoficiálně podporované značky patří: `citation_fulltext_html_url` (pro odkaz na plný text článku umístěný na webu); `citation_abstract_html_url` (pro odkaz na abstrakt článku umístěný na

webu); `citation_language` (pro upřesnění jazyka publikovaného článku); `citation_patent_number`; `citation_keywords`; `citation_patent_country`. Určitě by se našly další neoficiální značky, do této práce jsem ale vypsál pouze zajímavé a srozumitelné z názvu značky [MONPERRUS, 2008].

## 8.2 Reference

Akademický vyhledávač Google Scholar poskytuje u nalezených výsledků na daný dotaz zobrazení počtu citací daného článku (viz Obr. 5 Výpis výsledku v Google Scholar). Jak je popsáno výše ohledně hodnocení výsledků vyhledávače, tak číslo množství citací článků uvádí kvalitu článku. A to pouze za předpokladu, že lidé citují pouze kvalitní články a texty [BHUBI, 2019].

[Analysing Google rankings through search engine optimization data](#)

[S Furnell](#), MP Evans - Internet research, 2007 - emerald.com

Purpose—The purpose of this paper is to identify the most popular techniques used to rank a web page highly in Google. Design/methodology/approach—The paper presents the results of a study into 50 highly optimized web pages that were created as part of a Search Engine ...

☆ 99 Počet citací tohoto článku: 182 Související články Všechny verze (počet: 6) Web of Science: 56 98

### Obr. 5 Výpis výsledku v Google Scholar

Aby indexovací roboti Google Scholar mohly spočítat počet citací daného článku nebo textu musí být u článku, který z původního textu čerpá, umístěna citace. Indexovací roboti umí přečíst citaci z PDF dokument v seznamu citací.

Další možností je v případě elektronických článků (umístěných na webu) použití speciálních meta značek. Aby indexovací robot správně přečetl citaci přes meta značku, tak daná značka musí mít obsahovat následující parametry: `name` a `content`. Každý z těchto parametrů má dále za povinnost splňovat určitou syntaxi a klíčová slova. Parametr `name` musí obsahovat `citation_reference`. Tím indexovací robot pozná, že se jedná o citaci na jiný článek nebo text. Složitější je parametr `content`. Do tohoto parametru se uvádějí stejné značky, které se používají pro bibliografické údaje, jenž jsem popsal výše. Tyto značky jsou psány do jednoho

parametru, proto musí být hodnoty (značky) oddělené středníkem. Značky, které se dají využít pro citaci, jsou následující [BHUBI, 2019]:

- `citation_title` (název citovaného článku)
- `citation_author` (jméno citovaného autora, pokud je autorů více, každý autor se uvádí zvlášť)
- `citation_journal_title` (název například časopisu)
- `citation_volume` (může se zde uvést například číslo výtisku)
- `citation_pages` (zde se uvádí rozsah citovaných stránek)
- `citation_date` (datum citace)

Pro ukázkou je použit zdrojový kód článku na webu Nature.com. V článku jsou použity bibliografické meta značky a současně jsou zde využity meta značky určené pro citace. (viz Obr. 6 Zdrojový kód referencí)

```
<meta name="citation_reference" content="citation_journal_title=Nature; citation_author=S. Ide;
citation_volume=573; citation_publication_date=2019; citation_pages=112-116; citation_doi=10.1038/s41586-019-1508-
5; citation_id=CR1"/>

<meta name="citation_reference" content="citation_journal_title=Science; citation_author=W. L. Ellsworth, G. C.
Beroza; citation_volume=268; citation_publication_date=1995; citation_pages=851-855;
citation_doi=10.1126/science.268.5212.851; citation_id=CR2"/>

<meta name="citation_reference" content="citation_journal_title=Bull. Seism. Soc. Am.; citation_author=R.
Abercrombie, J. Mori; citation_volume=84; citation_publication_date=1994; citation_pages=725-734;
citation_id=CR3"/>

<meta name="citation_reference" content="citation_journal_title=Geophys. Res. Lett.; citation_author=J. Mori,
H. Kanamori; citation_volume=23; citation_publication_date=1996; citation_pages=2437-2440;
citation_doi=10.1029/96GL02491; citation_id=CR4"/>

<meta name="citation_reference" content="citation_journal_title=Nature; citation_author=E. L. Olson, R. M.
Allen; citation_volume=438; citation_publication_date=2005; citation_pages=212-215;
citation_doi=10.1038/nature04214; citation_id=CR5"/>
```

Obr. 6 Zdrojový kód referencí

### 8.3 Microdata

Microdata jsou technologií, která má za úkol označovat obsah webové stránky tak, aby byl strojově čitelný a zpracovatelný. Tím se dosáhne lepší dostupnosti informací pro prohlížeče nebo vyhledávače. Microdata svým způsobem fungují jako popis informací na webové stránce [SLÁDEK, 2010].

Pro použití microdata se využívají online dostupné slovníky. Každý slovník obsahuje jasně definované jména vlastností. Například jméno slovníku může být „Článek“, přičemž vlastnosti mohou být „Název“, „Autor“ nebo „Datum“. U microdata se vyskytuje i dědičnost jednotlivých vlastností. Například vlastnost článku „Autor“ může dědit další vlastnosti ze slovníku „Osoba“. Celá řada slovníků a popis jejich vlastností jsou dohledatelné například na stránce [www.schema.org](http://www.schema.org) [YANOLJA, 2020].

U microdata se používá zanořování. Například zmíněný slovník „Osoba“ se použije jako kořenový element. A uvnitř tohoto elementu se použijí například elementy jako `<h1>` nebo `<p>` pro vypsání jednotlivých vlastností.

```
<section itemscope itemtype="http://data-vocabulary.org/Person">
  <h1 itemprop="name">Jan Sládek</h1>
  <p></p>
  <p><a itemprop="url" href="http://honzasladek.com">osobní stránky</a></p>
</section>
```

**Obr. 7 Ukázka zanořování microdata [SLÁDEK, 2010]**

V ukázce jsou použity tři základní atributy pro použití microdata. A to: `itemscope`, `itemtype` a `itemprop`. `Itemscope` se používá pro označení celku, pro daný celek budou použity vlastnosti, které patří celku. `Itemtype` definuje odkaz na slovník, který je použit. Posledním důležitým atributem je `itemprop`. `Itemprop` definuje název vlastnosti, která spadá do slovníku [SLÁDEK, 2010].

Právě microdata jsou využita v praktické části této práce, kde se využívá slovník „ScholarlyArticle“ pro jednotlivé články.

## **II. Projekt**

## 9 Příprava pro generování zdrojového kódu stránky

Přípravou pro generování zdrojového kódu webové stránky se rozumí veškerý php kód, který stojí za načtením článků z databáze do dále zmíněných objektů.

Pro tuto práci je využito čistého PHP jazyka bez jakéhokoli dodatečného frameworku.

### 9.1 Objekty

V praxi se při vývoji webových stránek a aplikací více a častěji využívají objekty pro jednoduchou správu, načtení a uložení dat z databáze. V práci se využívají dva hlavní objekty pro čtení dat z databáze a následnému generování optimalizovaného zdrojového kódu pro indexovací roboty. Jedná se o třídu `Article` a třídu `Author`.

#### 9.1.1 Třída `Article`

Objekt `Article` představuje samotný článek. Je to stavební kámen celé této práce, který je využíván po celý průchod aplikací. Tento objekt se nachází v umístění `“/scholar/article.class.php”`.

Objekt `Article` obsahuje jednu proměnou `data`. V té jsou uloženy veškeré informace o článku, které jsou načtené z databáze do jednoho asociativního pole. Díky tomu, že se jedná o asociativní pole, je přístup k jednotlivým vlastnostem jednoduchý (viz Obr. 8 Ukázka přístupu k vlastnostem objektu).

```
$article->nazvanje  
$article->leto
```

Obr. 8 Ukázka přístupu k vlastnostem objektu

Objekt `Article` obsahuje konstruktor, `get` a `set` metodu, metodu na načtení autorů a metodu pro přidání instance objektu `Author` do vlastností objektu článku. Aby se dosáhlo uložení dat do asociativního pole a k jejich jednoduchému přístupu

jako na předchozí ukázce přístupu k vlastnostem, musí být použit speciální getter a setter (viz obrázek Obr. 8 Ukázka přístupu k vlastnostem objektu).

Metoda `load_authors()` volá funkci, která načte z databáze autory daného článku a vytvoří ke každému autorovi instanci třídy `Author` a tu uloží do asociativního pole pod klíčem `authors`.

Metoda `add_author()` obsahuje jeden parametr, který představuje instanci třídy `Author`. Tato metoda vloží zadanou instanci do pole. Zmíněné pole se nachází v asociativním poli instance třídy `Article` pod klíčem `authors`.

```
class Article {
    var $data;

    function __construct($input) {
        if (is_array($input)) {
            $this->data = $input;
        }
    }

    public function __set($property, $value) {
        return $this->data[$property] = $value;
    }

    public function __get($property) {
        if (isset($this->data[$property])) {
            return $this->data[$property];
        } else {
            return null;
        }
    }

    public function load_authors() {
        DB_get_authors_article($this);
    }

    public function add_author($author) {
        array_push($this->data['authors'], $author);
    }
}
```

Obr. 9 Zdrojový kód třídy `Article`



Iniciace třídy `Article` je díky zadanému konstruktoru a daným metodám `get` a `set` velice jednoduchá. Vypadá následovně:

```
while ($record = mysqli_fetch_assoc($query)) {  
    $article = new Article($record);  
    array_push($articles, $article);  
}
```

Obr. 10 Ukázka iniciace třídy `Article`

### 9.1.2 Třída `Author`

Objekt `Author` představuje data z databáze, které představují autory článků. Tato třída obsahuje pouze konstruktor, `get` a `set` metodu. K vlastnostem instance třídy `Author` se přistupuje stejně jako v případě třídy `Article`. Data jsou opět uložena v asociativním poli pro jednoduchý přístup.

```
class Author {  
    var $data;  
  
    function __construct($input) {  
        if (is_array($input)) {  
            $this->data = $input;  
        }  
    }  
  
    public function __set($property, $value) {  
        return $this->_data[$property] = $value;  
    }  
  
    public function __get($property) {  
        if (isset($this->data[$property])) {  
            return $this->data[$property];  
        } else {  
            return null;  
        }  
    }  
}
```

Obr. 11 Zdrojový kód třídy `Author`

## 9.2 Načtení článku z databáze

Načítání článků z databáze vypadá jako každé jiné čtení dat z databáze pomocí SELECT. Všechny funkce, které souvisí s jakoukoli prací s daty v databázi, se nachází v souboru v “/scholar/db\_client.php”. Zde jsou umístěny funkce pro načtení všech článků (DB\_get\_articles()), načtení článků podle ID (DB\_get\_article(\$id)), načtení článků podle roku (DB\_get\_articles\_year(\$year)), načtení všech autorů (DB\_get\_authors()), načtení autorů podle ID (DB\_get\_author(\$id)), načtení autorů podle článku (DB\_get\_authors\_article(\$article)). Více funkcí pro spolupráci s databází není potřeba, jelikož se jedná pouze o čtecí aplikaci. Aplikace nemanipuluje s daty v databázi, pouze je čte.

```

function DB_get_article($id) {
    $dbConnect = db_connect();

    $sql = "SELECT
            id,
            leto,
            izdanje, nazvanje,
            title, abstrakt,
            summary, kljucna,
            keywords,
            stranica_od, stranica_do
        FROM
            clanok
        WHERE
            clanok.id='$id'";

    $query = MySQLi_Query($dbConnect,$sql);
    if (!$query) {
        return "Error:" . mysqli_error($dbConnect);
    }
    $article = new Article(mysqli_fetch_assoc($query));
    $article->load_authors();

    mysqli_close($dbConnect);
    return $article;
}

```

**Obr. 12 Ukázka načtení článku podle ID**

Ostatní funkce pro čtení dat z databáze, které jsou vypsány v předchozím odstavci, vypadají podobně, jen se mění SQL dotaz a nachází se zde jiné parametry funkcí.

## 10 Generování zdrojového kódu stránky pro Google Scholar

Jakmile jsou načtena potřebná data a následně uložena v objektech, poté aplikace pokračuje dál generováním optimalizovaného zdrojového kódu pro indexovací roboty.

Všechny možné náležitosti optimalizovaného kódu pro indexování článků do Google Scholar jsou popsány v teoretické části této práce.

Generování zdrojového kódu můžeme rozdělit na dvě části. V první části se jedná o stránku, na níž jsou vypsány všechny články, které fungují jako odkazy přímo na jednotlivé články. Zde můžeme mluvit o splnění předpokladu indexování. Samotný předpoklad říká o existenci seznamu všech článků, které odkazují na jednotlivé články.

Druhá část představuje samostatný článek, který obsahuje všechny data a informace od roku vydání, jméno autora a jeho instituce, či ISSN.

### 10.1 Seznam všech článků

Na stránce se seznamem všech článků neexistují striktní požadavky na jejich indexaci. Musíme pouze dodržet splnění následujících předpokladů, které jsou pro indexaci nezbytné. Jedná se o validní HTML a meta informaci o názvu publikační asociace a ISSN.

```
<meta name="citation_journal_title" content="Slovjani.info">  
<meta name="citation_journal_issn" content="ISSN 2570-7108 (print), ISSN 2570-7116 (on-line)">
```

Obr. 13 Meta informace pro seznam všech článků

Níže popisují generování seznamu článků. Jedná se o jednoduchý výpis HTML elementů a vlastností článků.

```

<body>
  <h1>List of articles</h1>
  <?
  if (isset($_GET['year'])) {
    $articles = DB_get_articles_year($_GET['year']);
  } else {
    $articles = DB_get_articles();
  }

  $url = (isset($_SERVER['HTTPS']) && $_SERVER['HTTPS'] === 'on'
    ? "https"
    : "http")
    . "://$_SERVER[HTTP_HOST]/scholar/article.php";
  foreach($articles as $article) {
    ?>
    <div class="article_con">
      <div class="article_title">
        <a href="<?echo $url . "?id=" . $article->id?>">
          <?echo $article->nazvanje?>
        </a>
      </div>
    </div>
    <?
  }
  ?>
</body>

```

**Obr. 14** Generování seznamu článků

V prvním kroku se zkontroluje, jestli je v URL požadavek na články podle roku. Články podle roku vyžaduje nástroj pro napojení repositáře pro indexaci článků do Google Scholar (více v kapitole Registrace repositáře článků do Google Scholar).

Podle toho, jestli chceme nebo nechceme články podle roku, se volá jiná metoda pro načtení článků z databáze. V dalším kroku se zjistí, zda-li se jedná o HTTPS nebo HTTP protokol. Na základě toho se mění URL adresa pro použití jako odkaz na jednotlivé články.

Následuje cyklus, který prochází všechny načtené články z databáze. Pro každý článek se vytvoří odkaz, který využije URL vytvořené v přechozím kroku a přidá se

parametr GET s klíčem id a hodnotou ID článku. Jako text odkazu je použit název daného článku.

## 10.2 Jednotlivé články

Generování zdrojového článku je daleko obsáhlejší než generování seznamu všech článků. Samotný článek již obsahuje všechny náležitosti pro indexaci a všechny dostupné bibliografické údaje. Toto generování kódu připadá souboru “/scholar/article.php”.

V prvním kroku se načte článek podle zadaného ID v URL adrese.

```
$article = DB_get_article($_GET['id']);
```

**Obr. 15 Načtení článku podle ID**

V proměnné article se nachází všechna dostupná data k danému článku včetně autorů. Názvy sloupců v databázi jsou uloženy v novoslovanštině nebo v angličtině. Jedná se o samotné ID, leto (rok), izdanje (číslo vydání časopisu), nazvanje (název článku), title (anglický překlad názvu článku), abstrakt, kljucna (klíčová slova; jsou oddělená středníkem), keywords (anglický překlad klíčových slov), stranica\_od (strana začátku článku v publikaci), stranica\_do (strana konce článku v publikaci). Všechny tyto informace budou zpracovány pro vědecké meta-tagy vztahující se k článku. Každá instance článku obsahuje pole autorů. Jedná se o instance objektů Author. Ty obsahují následující data z databáze. ID, kratko\_ime (zkrácený zápis jména, např.: Merunka, V.), polno\_ime (celé nezkrácené jméno), titul, kontakt (e-mail autora), institucija (název instituce, kde autor působí), orcid.

Dalším krokem je rozdělení klíčových slov podle delimiteru. Jako oddělovač jednotlivých klíčových slov je použit středník. Rozdělení na jednotlivá slova probíhá následujícím způsobem:

```
$keywords = explode(";", $article->kljucna);
```

**Obr. 16 Rozdělení klíčových slov podle delimiteru**

Jakmile jsou načteny data z databáze a rozdělena klíčová slova, následuje generování potřebných meta-tagů. Jedná se o meta-tagy, které jsou popsány v kapitole 9.1.3 Vědecké meta-tagy. Jako titulek stránky je použit název článku (nazvanje). Jedná se o následující meta-tagy: název periodické publikace, ISSN periodické publikace, název článku, datum, rok publikování, abstrakt článku, popis (je použit abstrakt), číslo stránky, na které článek začíná a končí a odkaz na PDF dokument, ve kterém se článek nachází. Zmíněné meta-tagy se generují následujícím způsobem:

```
<meta name="citation_journal_title" content="Slovjani.info">
<meta name="citation_journal_issn" content="ISSN 2570-7108 (print), ISSN 2570-7116 (on-line)">
<meta name="citation_title" content="<? echo $article->nazvanje ?>">
<meta name="citation_publication_date" content="<? echo $article->leto ?>">
<meta name="citation_publication_year" content="<? echo $article->leto ?>">
<meta name="citation_abstract" content="<? echo $article->abstrakt ?>">
<meta name="description" content="<? echo $article->abstrakt ?>">
<meta name="citation_firstpage" content="<? echo $article->stranica_od ?>">
<meta name="citation_lastpage" content="<? echo $article->stranica_do ?>">
<meta name="citation_pdf_url" content="<? echo "{\$url}/archive/{\$article->leto}-{\$article->izdanje}.pdf" ?>">

<title><?echo $article->nazvanje?></title>
```

**Obr. 17 Vědecké meta-tagy článku**

Mezi další klíčové meta-tagy patří také klíčová slova. Pro ty platí jednoduché pravidlo o samostatném meta-tagu pro každé klíčové slovo s názvem tagu: citation\_keywords. Pro toto pravidlo bylo vytvořeno v předchozích krocích pole s jednotlivými klíčovými slovy. Generování těchto tagů s klíčovými slovy probíhá takto:

```
foreach($keywords as $keyword) {
    ?>
    <meta name="citation_keywords" content="<? echo $keyword ?>">
    <?
}
```

**Obr. 18 Generování meta-tagů pro klíčová slova**

Pro generování seznamu klíčových slov je použit cyklus foreach, který projde všechny prvky v poli a pro každý prvek vypíše meta-tag se stejným jménem, ale rozdílným contentem.

Dalšími povinnými meta-tagy jsou značky připadající autorům. Toto pravidlo je zmíněno v kapitole Vědecké meta-tagy. Každý autor nebo spoluautor článku má svůj samostatný meta-tag, stejně tak jako v případě klíčových slov. V mém případě se pro každého autora generuje jméno autora, instituce autora a e-mail autora. Generování těchto meta-tagů probíhá následujícím způsobem:

```
foreach($article->authors as $author) {  
    ?>  
    <meta name="citation_author" content="<? echo $author->polno_ime ?>">  
    <meta name="citation_author_institution" content="<? echo $author->institucija ?>">  
    <meta name="citation_author_email" content="<? echo $author->kontakt ?>">  
    <?>  
}
```

#### Obr. 19 Generování meta-tagů pro autory

Tímto jsem vyčerpal téma ke generování meta-tagů. Následuje generování samotného obsahu stránky. Jelikož tyto stránky jsou určeny především pro strojové zpracování dat a nikoli pro oči návštěvníka stránky, není obsah stránky článku nijak nastylován nebo jinak doplněn dodatečnou grafickou úpravou. K výše uvedenému je třeba doplnit, že stránka bude určena především pro strojové zpracování dat, a proto jsou použita tzv. microdata. Díky těmto microdatům bude dosaženo lepšího strojového zpracování a lepšího přístupu prohlížečům a vyhledávačům k informacím článků. Více v kapitole Microdata.

Pro každý článek je použit slovník ScholarlyArticle<sup>2</sup>, který definuje jednotlivé vlastnosti článku. Definice takového objektu vypadá následovně:

```
<article itemscope itemtype="http://schema.org/ScholarlyArticle">  
...  
...  
</article>
```

#### Obr. 20 Definování objektu ScholarlyArticle

Jednotlivé vlastnosti, které se použijí pro obsah objektu ScholarlyArticle jsou víceméně totožné jako u generování meta-tagů pro článek. Mezi tyto vlastnosti patří,

---

<sup>2</sup> Dostupný na adrese <http://schema.org/ScholarlyArticle>



v závorkách je popsáno, co je pro danou vlastnost použito za informaci: name (název článku), headline<sup>3</sup> (název článku), datePublished (rok publikování), image<sup>3</sup> (prázdná hodnota), abstrakt, jednotlivá klíčové slova, Publisher (objekt typu Organization, který definuje publikační organizaci), jednotliví autoři (objekty typu Person), isPartOf (objekt typu PublicationIssue, který definuje stránku článku a rok publikace). Vlastnost image je podle Google a definice strukturovaných dat povinná. Jelikož každý článek nemá vlastní obrázek, je použito logo stránky slovjani.info pro články. U vlastnosti Publisher jsou povinné dvě informace, a to název organizace a odkaz na logo organizace. Logo je další typ objektu ImageObject.

Tyto vlastnosti se generují následujícím způsobem:

```
<h1 itemprop="name"><? echo $article->nazvanje ?></h1>
<span itemprop="headline"><?echo $article->nazvanje?></span>
<span itemprop="datePublished"><?echo $article->leto?></span>
<span itemprop="image">http://slovjani.info/images/logo-left.gif</span>
<span itemprop="mainEntityOfPage">http://slovjani.info/</span>
<dl>
<dt>Publisher</dt>
<dd>
  <span itemprop="publisher" itemscope itemtype="http://schema.org/Organization">
    <span itemprop="name">Slovenská unie z. s.</span>
    <span itemscope itemprop="logo" itemtype="https://schema.org/ImageObject">
      <span itemprop="url" >http://slovjani.info/images/loga/su-mini.jpg</span>
    </span>
  </span>
</dd>
</dl>
```

**Obr. 21 Generování micro-data**

Klíčová slova mají být vypsána samostatně, toho se dosáhne například cyklem foreach:

---

<sup>3</sup> Tato vlastnost je definována googlem jako povinná. Více v kapitole Ověření správnosti microdata

```

<dt>Keywords</dt>
<dd>
  <ul>
    <?
      foreach($keywords as $keyword) {
        ?>
        <li itemprop="keywords"><? echo $keyword ?></li>
        <?
      }
    ?>
  </ul>
</dd>

```

**Obr. 22 Generování microdata - klíčová slova**

Dále jsou zmíněni autoři a spoluautoři článku. Každý autor je popsán slovníkem Person<sup>4</sup>. Autoři jsou popsáni jednotlivě v samostatných objektech. Oproti meta-tagům autorů je zde navíc ještě orcid. Jedná se o alfanumerický kód, který je jedinečný pro vědecké a akademické autory. Tato informace je popsána vlastností „sameAs“ v elementu link. Generování microdata pro autory je zobrazeno na následujícím obrázku.

---

<sup>4</sup> Dostupný na adrese: <http://schema.org/Person>

```

<dt>Authors</dt>
<?
foreach($article->authors as $author) {
  ?>
  <dd itemscope itemtype="http://schema.org/Person" itemprop="author">
    <dl>
      <dt><? echo $author->polno_ime ?></dt>
      <dd>
        <span itemprop="name"><? echo $author->polno_ime ?></span>
        <?
        if ($author->institucija != "") {
          ?>
          <span itemprop="affiliation"><? echo $author->institucija ?></span>
          <?}&?>
          <link itemprop="sameAs" href="http://orcid.org/<? echo $author->orcid ?>">
          <?
          if ($author->orcid != "") { ?>
          <a href="http://orcid.org/<? echo $author->orcid ?>" itemprop="url">
          <?} ?>
          <h3 itemprop="givenName"><? echo $author->kratko_ime ?></h3>
          <?
          if ($author->orcid != "") { ?>
          </a>
          <?} ?>
        </dd>
      </dl>
    </dd>
  </dd>
<?
}
?>

```

**Obr. 23** Generování microdata – autoři

Zbývá generovat doplňující informace o článku. Tyto informace jsou popsány slovníkem `PublicationIssue`<sup>5</sup>. Do tohoto slovníku spadají vlastnosti názvu, ISSN, E-ISSN, číslo vydání, rok vydání, číslo stránky začátku a konce článku v publikaci. Tyto informace jsou generovány níže:

---

<sup>5</sup> Dostupný na <http://schema.org/PublicationIssue>

```

<div itemprop="isPartOf" itemscope itemtype="http://schema.org/PublicationIssue">
  <div itemscope itemtype="http://schema.org/Periodical">
    <h1 itemprop="name">Slovjani.info</h1>
    <ul>
      <li>ISSN: <span itemprop="issn">2570-7108</span></li>
      <li>E-ISSN: <span itemprop="issn">2570-7116</span></li>
    </ul>
  </div>
  <span itemprop="issueNumber"><? echo $article->izdanje ?></span>
  <span itemprop="pageStart"><? echo $article->stranica_od ?></span>
  <span itemprop="pageEnd"><?echo $article->stranica_do ?></span>
  <span itemprop="DatePublished"><?echo $article->leto ?></span>
  <span itemprop="publisher">Slovjani.info</span>
</div>

```

**Obr. 24 Generování microdata informace o publikaci**

Celková ukázka celého výsledku optimalizovaného kódu webové stránky jednotlivého článku je umístěna v příloze Obr. 42 Ukázka optimalizovaného kódu webové stránky článku

Struktura kódu a jednotlivé vlastnosti článku jsou strukturované a optimalizované pro indexovací roboty.

## 11 Testování

Předpokladem Google pro indexování vědeckého článku do systému Google Scholar je splnění určitých předpokladů validního obsahu stránek jednotlivých článků [Google Scholar, 2019].

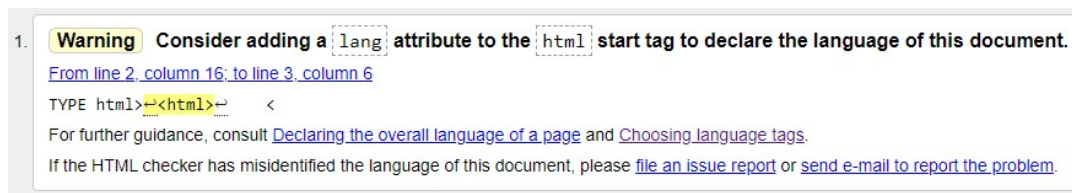
Indexovací roboti pouze indexují validní a kvalitní obsah. Kvalitního obsahu je dosaženo optimalizací zdrojového kódu stránky článku. Viz příloha: Obr. 42 Ukázka optimalizovaného kódu webové stránky článku

Pro ukázkou testování validity stránky i microdat byl použit článek pana Merunky s názvem: „Proglasenje na pamět 100 lět od osnovanja Češkoslovačskoj republiky“.

### 11.1 Ověření validity stránek

Jako úplně základní předpoklad je validní HTML kód webové stránky. K otestování validního zdrojového HTML kódu byl využit komunitní nástroj dostupný na adrese: <https://validator.w3.org/>.

Jako vstup do validátoru byl použit zdrojový kód stránky článku, který je dostupný klávesovou zkratkou CTRL + U. Výsledek je zobrazen na následujícím obrázku. Testovaný HTML kód stránky je validní. Validátor nenašel žádné kritické chyby až na jedno upozornění. Upozornění se týká chybějící informace o jazyku stránky. Jelikož novoslovanština nemá zatím oficiálně definovanou světově uznávanou zkratku, není tato informace o jazyku definovaná v kódu stránky. Důležité je, že na stránce nejsou žádné chyby v kódu. Pokud by v kódu byly nalezeny chyby, indexovací roboti by v nejlepším případě pouze penalizovaly kód, v nejhorším případě by článek vůbec neindexovali.



Obr. 25 Výsledek testu HTML validního kódu

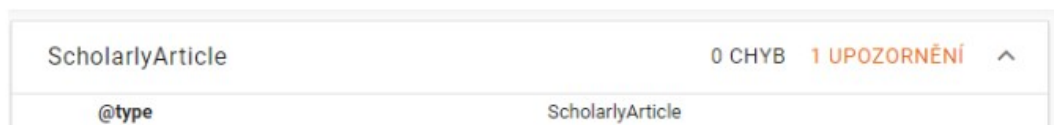
Tímto otestováním bylo zjištěno, že stránka je validní pro indexovací roboty. Dále je potřeba mít obsah kvalitní. Tomu se věnuje další test.

## 11.2 Ověření microdata

Microdata definují strukturu a informace o datech na stránce. Pokud budou všechny informace o článku jednoznačně definované svým obsahem, dosáhne se kvalitního obsahu stránky. Google má vlastní nástroj pro testování strukturovaného obsahu na stránce. Tento nástroj je dostupný na adrese: <https://search.google.com/structured-data/testing-tool/u/0/>.

Do validačního nástroje se vloží zdrojový kód obsahu, který chceme zkontrolovat. V tomto případě byl použit stejný zdrojový kód jako v případě testování validity.

Výsledkem testování strukturovaných dat na stránce byl zjištěn výsledek na následujícím obrázku:

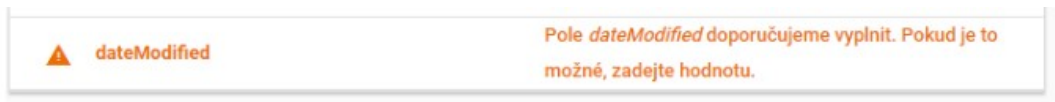


Obr. 26 Výsledek testování microdata

Výstupem testování je získání vypsaní samotného objektu, který je definován pomocí microdata. Jedná se o objekt `ScholarlyArticle`, přičemž jeho vlastnosti jsou vypsané jednotlivě po řádcích. Jelikož výstup testování je obsahově rozsáhlý, tak je jeho plná verze zobrazena v příloze: Ukázka celého výstupu z testování microdata.

Validační nástroj správně přečetl a zařadil jednotlivé vlastnosti různých objektů zařazených do `ScholarlyArticle`, jako je `author`, `publisher` nebo `isPartOf`.

Výsledek testování také upozorňuje na jedno doporučení. Jedná se o vlastnost `dateModified`, která se skriptem negeneruje díky skutečnosti, že je tato informace v současné době, pro moji práci, nedostupná nebo neexistuje.



**Obr. 27 Upozornění ve výsledku testování microdata**

Jakmile jsou stránky článků připraveny a obsah těchto stránek je validní a je zkontrolován, přechází se na další krok práce. A to informování vyhledávače Google o existenci těchto stránek, aby jednotlivé články mohly být indexovány do Google Scholar.

## 12 Propojení do Google Scholar

V případě, že máme optimalizované stránky pro webové crawlery a současně jsou tyto stránky validní s kvalitním obsahem, následuje zažádání Google o indexaci těchto stránek. Nejdůležitějším předpokladem zaindexování Google Scholar je samotná přístupnost webových crawlerů na dané stránky. To znamená, že stránky, které jsou výstupem této práce, jsou dostupné pro crawlery.

Mezi způsoby požádání Google o indexování stránky patří například registrace stránek do nástroje Google Search Console a sdílení stránek v příspěvku na sociální síti Twitter (webový crawleri procházejí a indexují obsah nových příspěvků včetně odkazů, které jsou v příspěvku). Tím se dosáhne toho, že webový crawler navštíví daný odkaz, projde jeho obsah a případně ho zaindexuje. Nevýhodou tohoto způsobu (použití odkazu v příspěvku na Twitteru) je skutečnost, že nemáme žádnou kontrolu nebo další nástroje po případném zaindexování stránky [STEWART, 2019].

### 12.1 Nastavení Google Search Console

Pro potřeby této práce je použita první zmíněná metoda. A to využití nástroje Google Search Console. Tento způsob byl vybrán z možnosti náhledu přehledu o indexování stránek, o případných chybách na stránkách, informacích o prokliknutí na články, informacích o počtu zobrazení ve výsledcích vyhledání, možností vyžádání znovuinindexování obsahu při případné změně obsahu. To vše tento nástroj umožňuje.

Registrace stránek do nástroje Google Search Console je poměrně jednoduchá. Google nabízí několik způsobů registrace. Jedná se o způsoby přidání služby s předponou adresy (například <https://www.example.com/test>). Další způsob je služba typu domény (například <https://www.example.com>). Posledním způsobem je přidání služby hostované Googlem, jedná se o blogy na stránkách Google. Pro tuto práci byla vybrána první metoda. Konkrétně byla přidána webová stránka <http://janjir.slovjani.info/scholar/articles.php>.



Po odeslání požadavku Google Search Console nabídne možnosti verifikace pro ověření, že jsem autor stránek. Vybral jsem optimální možnost - způsob ověření vložení speciální meta značky do hlavičky stránky. Tato značka je zobrazena na následujícím obrázku.

```
<meta name="google-site-verification" content="-_Rp_RiQRx20h[redacted]Jx4PVEi4aMjp6nv8xzm8" />
```

**Obr. 28 Meta-tag pro Google Search Console**

Po odeslání požadované stránky jako požadavek přidání nové služby do Google Search Console se ukáže ve všech dostupných nástrojích (přehled, výkon, kontrola adresy URL, pokrytí, rychlost) následující hláška (viz Obr. 29 Google Search Console - zpracovávání dat).



Data se zpracovávají, zkuste to znovu přibližně za den

**Obr. 29 Google Search Console - zpracovávání dat**

Tato hláška se zobrazovala zhruba tři týdny než se projeví nějaké výsledky zpracování dat. Po těchto výsledcích se zobrazuje jedna platná stránka, na které není nalezena žádná chyba. Od chvíle zobrazení prvních výsledků v Google Search Console po registraci stránky byla dohledatelná přes Google vyhledávač jedna stránka. Jedná se o stránku, která byla zadána při registračním formuláři (stránka <http://janjir.slovjani.info/scholar/articles.php>). Jednotlivé články dohledatelné nejsou. Tento problém se zařídí v dalším kroku, a to nastavení parametrizace URL. Před nastavením parametru id v URL adrese (pokud je zadán do vyhledávače Google tento dotaz): „site:janjir.slovjani.info“, Google najde pouze jednu domovskou stránku se články. Výsledek vyhledání je zobrazen na dalším obrázku.

## Articles - SLOVJANI.info

Vuk kako naš sovrěmenik: vplyv Vuka Stefanovića Karadžića i Matice srbskoj na subjekty historije srbskoj kulture · Sveta Ljudmila 1100 lět: tradicija - narodne ...

### Obr. 30 Výsledek vyhledání Google před nastavením parametru id

Pro specifikování jednotlivých článků se využívá parametr id v URL adrese. Například <http://janjir.slovjani.info/scholar/articles.php?id=151>. Pro registraci parametru id je využit nástroj pro parametrizaci URL v Google Search Console. Nastavení parametru id pro URL adresu je zobrazeno na následujícím obrázku.

Parametr: id

Mění tento parametr obsah stránky viditelný uživateli?

Ano, mění, reorganizuje nebo zužuje obsah stránky ▾

Jak tento parametr ovlivňuje obsah stránky?

Specifikuje ▾

Upřesňuje, o čem stránka pojednává – například téma, publikum, číslo položky apod.

Které adresy URL s tímto parametrem má Googlebot procházet?

- Nechat rozhodnutí na prohlížeči Googlebot <sup>?</sup> (Výchozí)
- Každou adresu URL <sup>?</sup> (obsah stránky se pro každou hodnotu změní)
- Pouze adresy URL s hodnotou  <sup>?</sup> (může skrýt obsah před prohlížečem Googlebot)
- Žádné adresy URL <sup>?</sup> (může skrýt obsah před prohlížečem Googlebot, přepíše nastavení jiných parametrů)

► Zobrazit příklady adres URL

**Uložit**

Zrušit

### Obr. 31 Google Search console - nastavení parametru URL

Pro nastavení parametru URL se využívá jednoduchý formulář, ve kterém se nastavuje: jestli parametr mění obsah stránky viditelný uživateli, jak parametr ovlivňuje obsah stránky, a které adresy URL má Googlebot procházet. V mém případě se jedná o možnost “Ano, mění, reorganizuje nebo zužuje obsah stránky”, dále možnost, že parametr specifikuje obsah stránky a možnost volby procházení všech stránek Googlebotem. Po nastavení parametru id se ukazuje v nástroji následující výsledek nastavení parametru.

## Parametry adres URL

Pomozte Googlu procházet váš web efektivněji. Sdělte nám, jak máme zacházet s parametry ve vašich adresách URL.  
[Další informace](#)

Stáhnout tuto tabulku	Přidat parametr	Zobrazit	Řádky: 25	1-1 z 1	<	>
Parametr	Monitorované adresy URL	Nakonfigurováno	Účinek	Procházení		
id	-	11. 2. 2020	Specifikuje	Každou adresu URL	Upravit	Vymazat

1-1 z 1 < >

Obr. 32 Google Search Console - výsledek nastavení parametru URL

## 12.2 Odeslání žádosti o znovuindexování stránek

Jakmile je nastaven URL parametr pro procházení článků, následuje požádání Google o znovuindexování stránek včetně využití parametru id v URL adrese.

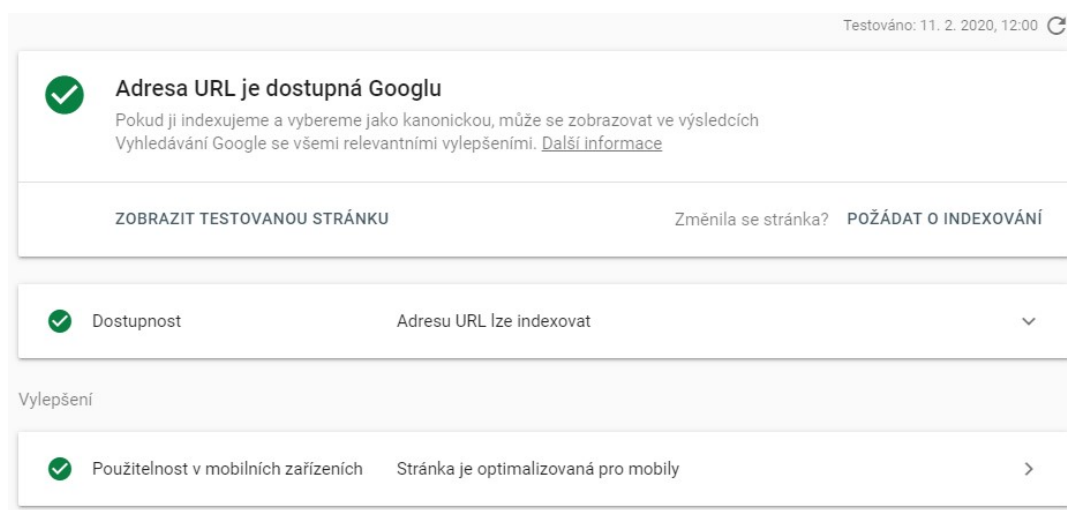
Pro odeslání žádosti o opětovnou indexaci se využívá nástroj „Kontrola adresy URL“. Tento nástroj umožňuje otestování publikované URL adresy nebo otestování stránek při využití mobilních zařízení.

Předtím než je zaslán požadavek o indexaci, tak je proveden test pomocí tlačítka „Otestovat publikovanou adresu URL“. Po zvolení tohoto testu se ukáže následující dialogové okno.



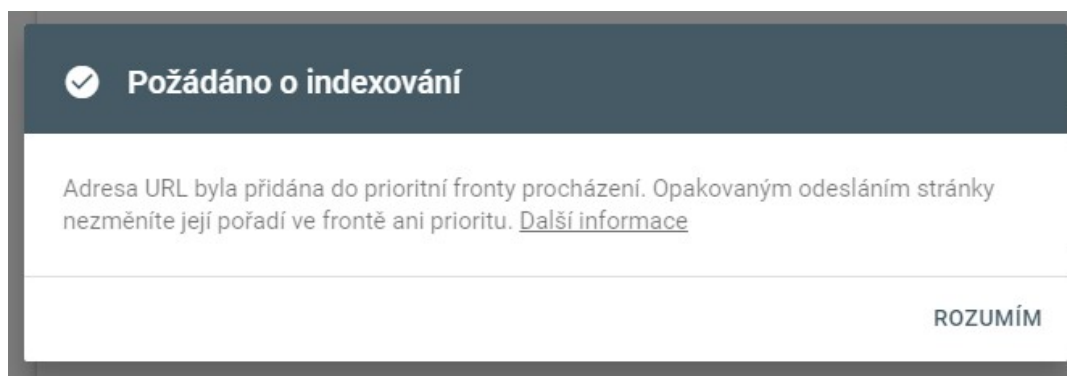
Obr. 33 Google Search Console - testování publikované URL adresy

Jako výsledek testování je zobrazena informace o dostupnosti URL adresy Google a zobrazení jednotlivých možností pro vylepšení stránek pro mobilní zařízení. V mém případě všechny testy proběhly v pořádku. Výsledek testování je zobrazen na dalším obrázku.



**Obr. 34 Google Search Console - výsledek testování publikované URL adresy**

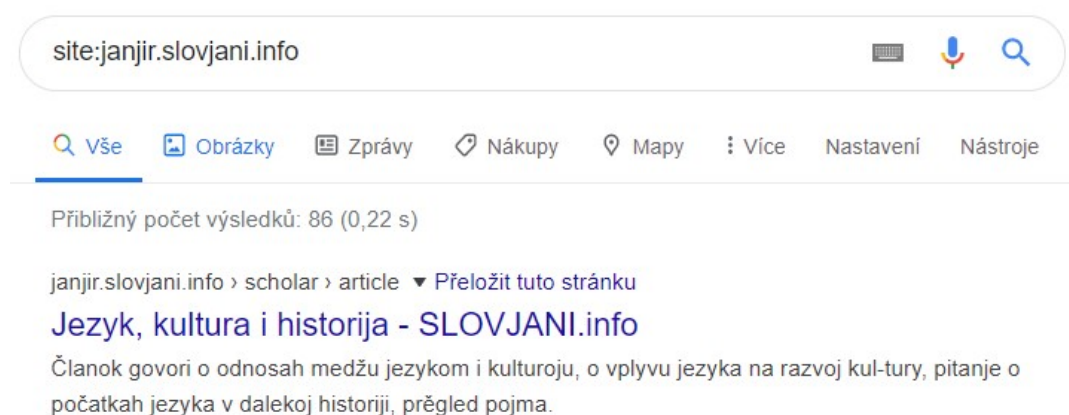
Testování proběhlo v pořádku ve všech bodech, tudíž mohu přejít na požádání o indexování stránek, aby jednotlivé články byly dohledatelné Googlem. Pro zažádání o opětovnou indexaci se zvolí možnost z předchozího obrázku: „Požádat o indexování“. Po zvolení této možnosti se zobrazí následující dialog o informování o požadavku o indexování.



**Obr. 35 Google Search Console - požádání o indexování stránek**

## 12.3 Ověření indexace stránek Google

Zhruba po půl hodině po odeslání požadavku o indexaci se projevil výsledek indexace jednotlivých článků. Pokud teď zadám do vyhledávače Google následující dotaz „“, Google nalezne 86 stránek, které představují jednotlivé články. Výsledek vyhledání je nalezen na následujícím obrázku, tento výsledek je krácen pouze na jeden článek. Plná verze výsledku je zobrazena v příloze: Ukázka výsledků vyhledání na Google.



**Obr. 36 Výsledek vyhledání Google po nastavením parametru**

Teď mám jistotu, že Google ví o všech člancích, které jsou indexovány. Jako další test můžu vyhledat Googlem například frázi „Proglasenje na pamět 100 lět“. Jedná se o část nadpisu jednoho článku. Výsledek vyhledání je ukázán na dalším obrázku.



**Obr. 37 Ukázka výsledku vyhledání při zadání "Proglasenje na pamět 100 lět"**

Dále dobrým výsledkem je skutečnost, že moje stránky s články mají kvalitnější obsah než jiné systémy pro databázi článků, jako je například ceool. Moje stránky, které jsou výsledkem této práce, se umísťují na prvních příčkách vyhledání. To

dokazuje, že zdrojový kód stránek se články je kvalitní a má vypovídací schopnost o člancích. Ukázkou pořadí výsledků oproti konkurenčním stranám je zobrazena v příloze: Obr. 45 Ukázkou srovnání pořadí výsledků s konkurencí při hledání "Prohlášení na paměť 100 let".

## 12.4 Registrace repositáře článků do Google Scholar

Jakmile je ověřeno, že jednotlivé stránky s články jsou validní, jsou dostupné pro indexovací roboty a je otestováno indexování stránek může se přejít k dalšímu kroku a to uložení repositáře do Google Scholar. Tímto přidáním repositáře se dosáhne maximálního pokrytí všech možností pro indexování článků do vyhledávače Google Scholar. Indexovací roboti umí indexovat články i bez uloženého repositáře, pokud přidám repositář, tak se jedná o takovou pojistku o zaručeném indexování článků do Google Scholar.

Pro uložení repositáře se využívá další nástroj od Google, jedná se o nástroj Google Scholar Inclusions<sup>6</sup>.

Na začátku registrace repositáře se vybírá druh webové stránky, na které se nachází repositář.



Type of website:	DSpace repository	EPrints repository	Other repository
	Open Journal Systems OJS website	Other journal website	Personal publications

Obr. 38 Registrace repositáře - výběr druhu webové stránky

Po zvolení druhu stránky se pokračuje na další povinná pole formuláře pro registraci repositáře. V prvním kroku se musí zaškrtnout všechny checkboxy, které nám specifikují jednotlivá pravidla pro registraci. Všechna tato pravidla musí být

<sup>6</sup> Dostupný na adrese.

[https://partnerdash.google.com/partnerdash/d/scholarinclusions?visit\\_id=637177949682818379-1686071171&rd=1#p:id=new](https://partnerdash.google.com/partnerdash/d/scholarinclusions?visit_id=637177949682818379-1686071171&rd=1#p:id=new)

splněna. První pravidlo říká, že každý článek je na separátní adrese, další pravidlo říká o existenci autorem napsaném abstraktu nebo samotný text celého článku. Následuje pravidlo o přístupu k článku, každý článek musí být přístupný bez nutnosti se přihlásit. Pravidlo o robots.txt souboru (pokud existuje), který povoluje robotům přístup na všechny články a poslední pravidlo o tom, že se jedná o repositář rozdílný od digitální knihovny nebo Eprints. Tyto pravidla jsou zobrazena na následujícím obrázku:



Google Scholar inclusion checklist:

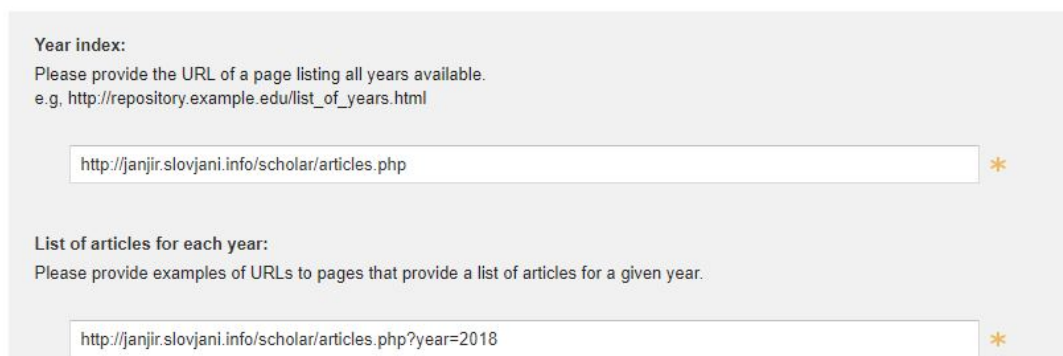
Please check all the boxes.

- My website has a separate URL for each article, as well as a browse interface for all article URLs.
- Each article URL contains either the complete author-written abstract or the entire text of the paper.
- All article URLs can be read by any user without a login or payment.
- The robots.txt file on my website, if my website has one, allows Googlebot to crawl all article URLs as well as the browse interface.
- My inclusion request is for a repository type other than DSpace or Eprints.

Detailed technical guidelines can be found at <https://scholar.google.com/intl/en/scholar/inclusion.html>.

#### Obr. 39 Registrace repositáře - splnění pravidel registrace

Dalším krokem v registračním formuláři je samotný odkaz na celkový seznam článků a také na filtrování článků podle roku. Toto filtrování je popsáno v kapitole Seznam všech článků.



Year index:

Please provide the URL of a page listing all years available.  
e.g. [http://repository.example.edu/list\\_of\\_years.html](http://repository.example.edu/list_of_years.html)

\*

List of articles for each year:

Please provide examples of URLs to pages that provide a list of articles for a given year.

\*

#### Obr. 40 Registrace repositáře - vložení odkazu na seznam článků

Posledním krokem registračního formuláře je vložení odkazů jako příklady na jednotlivé články. Tento krok je ukázán na dalším obrázku.

**Article examples:**

Please provide examples of full-text article URLs. Please pick some recently added articles as well as some articles that were added a while ago.



**Obr. 41 Registrace repositáře - příklady článků**

Po odeslání registračního formuláře dorazí e-mail o informování úspěšné registrace repositáře, jenž informuje o přijetí registrace. Ukázka e-mailu je v příloze Obr. 46 Registrace repositáře - email po registraci.



### **13 Zhodnocení výsledků a doporučení**

Jako na způsob propojení do systému Google Scholar jsem se primárně soustředil na optimalizaci zdrojového kódu stránek s články pro indexovací roboty, a to z důvodu, že indexace do Google Scholar kromě ručního vložení článku vkládá články automaticky, a to právě pomocí indexovacích robotů.

Výsledkem této práce jsou optimalizované webové stránky s jednotlivými články pro indexovací roboty, které splňují veškeré povinnosti a náležitosti pro indexaci do Google Scholar. Tyto stránky jsou zaregistrovány v systému Google Search Console, který umožňuje požádání o opětovnou indexaci článků, ale hlavním důvodem zvolení tohoto nástroje je svým způsobem informování Google o existenci těchto stránek a jejich následné indexaci. Tímto nástrojem bylo dosaženo, že Google ví o těchto stránkách a dokáže je procházet a indexovat. Jako pojistka indexování stránek s články bylo pro jistotu využito nástroje, který umožňuje registrovat repositář článků. Tudíž Google ví o těchto stránkách dvojím způsobem, první je nástroj Google Search Console a druhý je uložení repositáře článků.

V současné chvíli samotné články k nalezení v Google Scholar nejsou z důvodu, že jiné repositáře s články jsou již na Google dlouho (například ceeol.com). Drtivá většina článků pro které byly touto prací vytvořeny optimalizované stránky jsou již zaindexovány v Google Scholar právě zmiňovaným systémem ceeol.com. Jak se tato optimalizace stránek s články projeví, a který zdroj bude využit se ukáže při dalším vydání časopisu slovjani.

Oproti konkurenčnímu ceeol.com mají webové stránky s články z mé práce kvalitnější obsah, byť vypadá na první pohled stejně. Toho bylo dosaženo využitím micro-data. V současné chvíli pro indexování článků do Google Scholar nemůžu pomocí optimalizovaných stránek více udělat.

Do budoucna bych doporučil autorům článků, kteří publikují na stránce slovjani.info, aby řádně vyplňovali všechny možné jim dostupné bibliografické

informace, protože na těchto informacích závisí optimalizovaný kód webové stránky článku, jež využívají tyto informace a data.

## 14 Závěr

Výsledkem této práce jsou optimalizované stránky s důrazem na splnění všech předpokladů a povinností pro indexaci článků do vyhledávače Google Scholar. Optimalizované stránky s články se podařilo indexovat a dohledat přes vyhledávač Google. Pro indexování článků do Google Scholar se musí počkat do dalšího vydání e-časopisu Slovjani.info, pro který je tato práce určena, jelikož jsou dříve publikované články e-časopisu Slovjani.info již indexované v Google Scholar jinými systémy.

Úplné výsledky této diplomové práce se projeví časem, jelikož optimalizace zdrojového kódu je vždy během na dlouhou trať. Touto prací budou muset optimalizované stránky pokořit konkurenční systémy zabývající se podobnou tematikou. Optimalizované stránky, jakožto výsledek této práce, obsahují kvalitnější obsah webových stránek s články než konkurenční systémy. Na rozdíl od konkurenčním systémů mé stránky obsahují strukturovaná data o člancích.

Přínosem této práce do budoucna bude automatická indexace jednotlivých článků publikovaných v mezinárodním e-časopisu Slovjani.info do vyhledávače Google Scholar.

## 15 Použitá literatura

- ALEX CHITU. Google's Market Share in Your Country. [online], 2009.  
<http://googlesystem.blogspot.com/-2009/03/googles-market-share-in-your-country.html>
- BANSAL, Meenakshi; SHARMA, Deepak. Improving Webpage Visibility in Search Engines by Enhancing Keyword Density Using Improved On-Page Optimization Technique. *International Journal of Computer Science and Information Technologies*, 2015, 6.6: 5347-5352.
- BEEL, Jöran; GIPP, Bela; WILDE, Erik. Academic Search Engine Optimization (aseo) Optimizing Scholarly Literature for Google Scholar & Co. *Journal of scholarly publishing*, 2009, 41.2: 176-190.
- BHUBI, A. (2019). [online]. Distilled.net. Google Scholar: Indexation & Ranking.  
<https://www.distilled.net/resources/google-scholar-indexation-ranking/>.
- BORNMANN, Lutz; DANIEL, Hans-Dieter. What do we know about the h index?. *Journal of the American Society for Information Science and technology*, 2007, 58.9: 1381-1385.
- Dabas, Bharat & Pateria, Rajive. (2018). SEARCH ENGINE : AN OVERVIEW.
- GLÄNZEL, Wolfgang. On the opportunities and limitations of the H-index. *Science focus*, 2006.
- GUPTA, Vishal. A keyword searching algorithm for search engines. In: 2007 Innovations in Information Technologies (IIT). IEEE, 2007. p. 203-207.
- HALLUR, Akshay. How many people go beyond the 1st page of Google? Quora.com [online]. 2015 [cit. 2019-11-10]. Dostupné z: <https://www.quora.com/How-many-people-go-beyond-the-1st-page-of-Google>
- HARZING, Anne-Wil K.; VAN DER WAL, Ron. Google Scholar as a new source for citation analysis. *Ethics in science and environmental politics*, 2008, 8.1: 61-73.
- HARZING, Anne-Wil; ALAKANGAS, Satu. Google Scholar, Scopus and the Web of Science: a longitudinal and cross-disciplinary comparison. *Scientometrics*, 2016, 106.2: 787-804.
- HEITZMAN, Adam. ORGANIC VS. PAID SEARCH: (66 ASTONISHING) STATISTICS FOR 2020. Highervisibility [online]. 2019-07-02 [cit. 2019-11-05]. Dostupné z: <https://www.highervisibility.com/blog/organic-vs-paid-search-statistics/>
- CHAKRAVARTY, Rupak; RANDHAWA, Sukhwinder. Academic search engines: librarian's friend, researcher's delight. 2006.
- CHOOCHAIWATTANA, Worasit. An Academic Search Engine for Personalized Rankings. *Journal of Reviews on Global Economics*, 2017, 6: 350-354.
- Inclusion Guidelines for Webmasters [online] Google Scholar,  
<https://scholar.google.com/intl/en/scholar/inclusion.html#indexing>
- JACSÓ, Péter. Google scholar revisited. *Online information review*, 2008, 32.1: 102-114.
- LEDFORD, Jerri L. Search engine optimization bible. John Wiley & Sons, 2015.
- LEWANDOWSKI, Dirk; MAYR, Philipp. Exploring the academic invisible web. *Library hi tech*, 2006, 24.4: 529-539.
- MIKKI, Susanne. Google scholar compared to web of science. A literature review. *Nordic Journal of Information Literacy in Higher Education-NORIL*, 2009, 1.1.

- MONPERRUS, M. (2008). [online]. Monperrus.net. Accurate bibliographic metadata and google scholar.  
<https://www.monperrus.net/martin/accurate+bibliographic+metadata+and+google+scholar>
- NORUZI, Alireza. Google Scholar: The new generation of citation indexes. *Libri*, 2005, 55.4: 170-180.
- ORDUNA-MALEA, ENRIQUE & AYLLÓN, JUAN & MARTÍN-MARTÍN, ALBERTO & DELGADO LÓPEZ-CÓZAR, EMILIO. (2014). About the size of Google Scholar: playing the numbers.
- SHARMA, Dushyant, et al. A Brief Review on Search Engine Optimization. In: 2019 9th International Conference on Cloud Computing, Data Science & Engineering (Confluence). IEEE, 2019. p. 687-692. PARK, Lemuel S. (2015). Search engine optimization for category specific search results. U.S. Patent No 9,116,994.
- SLÁDEK, Jan. Webdesignérův průvodce po HTML5: Microdata. Zdroják.cz [online]. 20.7.2010 [cit. 2020-01-15]. Dostupné z: <https://www.zdrojak.cz/clanky/webdesigneruv-pruvodce-po-html5-microdata/>
- STEWART, Ryan. 8 Ways to Get Your Pages Indexed by Google, Quickly. Webris [online]. 2019-11-28 [cit. 2020-02-13]. Dostupné z: <https://webris.org/google-index/>
- TORCZON, Virginia. On the convergence of pattern search algorithms. *SIAM Journal on optimization*, 1997, 7.1: 1-25.
- Urlagunta, Nagaraju & Teja, Dharma & Rajashekar, & Ramaiah, Mangayarkarasi. (2017). SEARCH ENGINE OPTIMIZATION.
- VAN DIJCK, José. Search engines and the production of academic knowledge. *International Journal of Cultural Studies*, 2010, 13.6: 574-592.
- What is PPC? Pay-Per-Click Explained. Portent [online]. 2019-05-09 [cit. 2019-11-05]. Dostupné z: <https://www.portent.com/services/ppc/pay-per-click-explained>
- YANOLJA. Microdata. MDN Web docs [online]. 12.1.2020 [cit. 2020-01-15]. Dostupné z: <https://developer.mozilla.org/en-US/docs/Web/HTML/Microdata>
- Začínáme s optimalizací pro vyhledávače (SEO). Google.com [online]. 2020 [cit. 2019-11-10]. Dostupné z: <https://support.google.com/webmasters/answer/7451184?hl=cs>

### **III. Přílohy**

## 16 Přílohy

### 16.1 Ukázka optimalizovaného kódu webové stránky článku

```
<!DOCTYPE html>
<html>
<head>
<meta name="google-site-verification" content="-_Rp_81QRx20hkYdU8oy26Jx4PVEi4aMj6nv8xzm8" />
<meta name="citation_journal_title" content="Slovjani.info">
<meta name="citation_journal_issn" content="ISSN 2570-7108 (print), ISSN 2570-7116 (on-line)">
<meta name="citation_title" content="Prohlášení na pamět 100 lét od osnovenja Českoslovačskej republiky">
<meta name="citation_publication_date" content="2018">
<meta name="citation_publication_year" content="2018">
<meta name="citation_abstract" content="Prohlášení na pamět 100 lét od osnovenja Českoslovačskej republiky i diskusija nad vizijeju novogo strategičnogo čelja s
<meta name="description" content="Prohlášení na pamět 100 lét od osnovenja Českoslovačskej republiky i diskusija nad vizijeju novogo strategičnogo čelja slovja
<meta name="citation_firstpage" content="1">
<meta name="citation_lastpage" content="4">
<meta name="citation_pdf_url" content="http://janjir.slovjani.info/archive/2018-2.pdf">
<title>Prohlášení na pamět 100 lét od osnovenja Českoslovačskej republiky</title>
<meta name="citation_keywords" content="oslava 100 lét">
<meta name="citation_keywords" content="Českoslovačska republika">
<meta name="citation_keywords" content="novy strategičny čelj slovjanskych narodov">
<meta name="citation_keywords" content="slovjanska kulturna diplomacija">
<meta name="citation_author" content="Vojtěch Merunka">
<meta name="citation_author_institution" content="Czech University of Life Sciences Prague, Faculty of Economics and Management, Prague, Czech Republic">
<meta name="citation_author_email" content="vmerunka@gmail.com">
</head>
<body>
<article itemscope itemtype="http://schema.org/ScholarlyArticle">
<h1 itemprop="name">Prohlášení na pamět 100 lét od osnovenja Českoslovačskej republiky</h1>
<span itemprop="headline">Prohlášení na pamět 100 lét od osnovenja Českoslovačskej republiky</span>
<span itemprop="datePublished">2018</span>
<span itemprop="image">http://slovjani.info/images/logo-left.gif</span>
<span itemprop="mainEntityOfPage">http://slovjani.info</span>
<div>
<dt>Publisher</dt>
<dd>
<span itemprop="publisher" itemscope itemtype="http://schema.org/Organization">
<span itemprop="name">Slovanská unie z. s.</span>
<span itemscope itemprop="logo" itemtype="https://schema.org/ImageObject">
<span itemprop="url">http://slovjani.info/images/loga/su-mini.jpg</span>
</span>
</dd>
<dt>Keywords</dt>
<dd>
<ul>
<li itemprop="keywords">oslava 100 lét</li>
<li itemprop="keywords">Českoslovačska republika</li>
<li itemprop="keywords">novy strategičny čelj slovjanskych narodov</li>
<li itemprop="keywords">slovjanska kulturna diplomacija</li>
</ul>
</dd>
<dt>Authors</dt>
<dd itemscope itemtype="http://schema.org/Person" itemprop="author">
<div>
<dt>Vojtěch Merunka</dt>
<dd>
<span itemprop="name">Vojtěch Merunka</span>
<span itemprop="affiliation">Czech University of Life Sciences Prague, Faculty of Economics and Management, Prague, Czech Republic</span>
<link itemprop="sameAs" href="http://orcid.org/0000-0002-9056-1439">
<a href="http://orcid.org/0000-0002-9056-1439" itemprop="url">
<h3 itemprop="givenName">Merunka, V.</h3>
</a>
</dd>
</div>
</dd>
</div>
<div>
<h2>Abstract</h2>
<div itemprop="abstract">
<p>Prohlášení na pamět 100 lét od osnovenja Českoslovačskej republiky i diskusija nad vizijeju novogo strategičnogo čelja slovjanskych narodov do t
</p>
</div>
</div>
<div itemprop="isPartOf" itemscope itemtype="http://schema.org/PublicationIssue">
<div itemscope itemtype="http://schema.org/Periodical">
<h1 itemprop="name">Slovjani.info</h1>
<ul>
<li>ISSN: <span itemprop="issn">2570-7108</span></li>
<li>E-ISSN: <span itemprop="issn">2570-7116</span></li>
</ul>
</div>
<div>
<span itemprop="issueNumber">2</span>
<span itemprop="pageStart">1</span>
<span itemprop="pageEnd">4</span>
<span itemprop="datePublished">2018</span>
<span itemprop="publisher">Slovanská unie z. s.</span>
</div>
</div>
</article>
</body>
</html>
```

Obr. 42 Ukázka optimalizovaného kódu webové stránky článku

## 16.2 Ukázka celého výstupu z testování microdata

ScholarlyArticle		0 CHYB 1 UPOZORNĚNÍ ^
@type	ScholarlyArticle	
name	Prohlášení na pamět 100 lét od osnovanja Češkoslovačskej republiky	
headline	Prohlášení na pamět 100 lét od osnovanja Češkoslovačskej republiky	
datePublished	2018	
image	http://slovjani.info/images/logo-left.gif	
mainEntityOfPage	http://slovjani.info/	
keywords	oslava 100 lét	
keywords	Češkoslovačska republika	
keywords	novy strategičny cělj slovjanskyh narodov	
keywords	slovjanska kulturna diplomacija	
abstract	Prohlášení na pamět 100 lét od osnovanja Češkoslovačskej republiky i diskusija nad vizijeju novogo strategičnogo cělja slovjanskyh narodov do budučnosti 21-ogo stolětija.	
publisher		
@type	Organization	
name	Slovanská unie z. s.	
logo		
@type	ImageObject	
url	http://slovjani.info/images/loga/su-mini.jpg	
author		
@type	Person	
name	Vojtěch Merunka	
sameAs	http://orcid.org/0000-0002-9056-1439	
url	http://orcid.org/0000-0002-9056-1439	
givenName	Merunka, V.	
affiliation		
@type	Organization	
name	Czech University of Life Sciences Prague, Faculty of Economics and Management, Prague, Czech Republic	
isPartOf		
@type	PublicationIssue	
issueNumber	2	
pageStart	1	
pageEnd	4	
datePublished	2018	
publisher		
@type	Thing	
name	Slovanská unie z. s.	
dateModified		<b>Pole <i>dateModified</i> doporučujeme vyplnit. Pokud je to možné, zadejte hodnotu.</b>

Obr. 43 Ukázka celého výstupu testování microdata



## 16.3 Ukázka výsledků vyhledání na Google

The image shows a Google search interface. The search bar contains the text "site:janjir.slovjani.info". Below the search bar, there are navigation links: "Vše", "Obrázky", "Zprávy", "Nákupy", "Mapy", "Více", "Nastavení", and "Nástroje". The search results are displayed below, showing a list of articles from the website "janjir.slovjani.info". Each result includes a breadcrumb trail, a link to translate the page, a title, and a short description.

Přibližný počet výsledků: 86 (0,22 s)

janjir.slovjani.info › scholar › article ▼ Přeložit tuto stránku  
**Jezyk, kultura i historija - SLOVJANI.info**  
Članok govori o odnosah medžu jezykom i kulturoju, o vplyvu jezyka na razvoj kul-tury, pitanje o počatkah jezyka v dalekoj historiji, prëgled pojma.

janjir.slovjani.info › scholar › article ▼ Přeložit tuto stránku  
**Nitra a Bojná – významné slovanské sídla na západnom ...**  
Úrodné nížiny západného Slovenska boli odpradávná vyhľadávané ľudskými spoločenstvami a preto boli husto osídlené. Tu sa v praveku formovali viaceré ...

janjir.slovjani.info › scholar › article ▼ Přeložit tuto stránku  
**Obrazok iz pokryvnoj stranice časopisa - SLOVJANI.info**  
Članok popisuje obrazok, ktory jesm stvoril do oslavnyh dnov na paměť 100 lět od obnovenja samostojnoj Poljsky i osnovanja Češkoslovakije i Kraljevstva ...

janjir.slovjani.info › scholar › article ▼ Přeložit tuto stránku  
**Slovjanstvo dnes i zajutra - SLOVJANI.info**  
Slovjanstvo, točnjeje slovjanska vzajemnost, naleželo do glavnyh motivačnyh faktorov narodnogo razbudženja v 19-om věku i polnoj emancipacije netoliko ...

janjir.slovjani.info › scholar › article ▼ Přeložit tuto stránku  
**Как ние българите разбираме Новославянския език без ...**  
Članok govori o razumivosti novoslovjanskogo jezyka od Bulgarov v latinici i bez nikakogo přëdhodnogo učenja. Anketa iměla 20 pytanj v dvoch grupah.

Obr. 44 Ukázka výsledku vyhledání na Google

## 16.4 Ukázka pořadí výsledků s konkurencí

Proglasenje na pamet 100 let

Vše   Obrázky   Nákupy   Video   Zprávy   Více   Nastavení   Nástroje

Přibližný počet výsledků: 15 200 (0,25 s)

Tip: Hledejte pouze výsledky v češtině. Jazyk vyhledávání můžete nastavit na stránce Nastavení

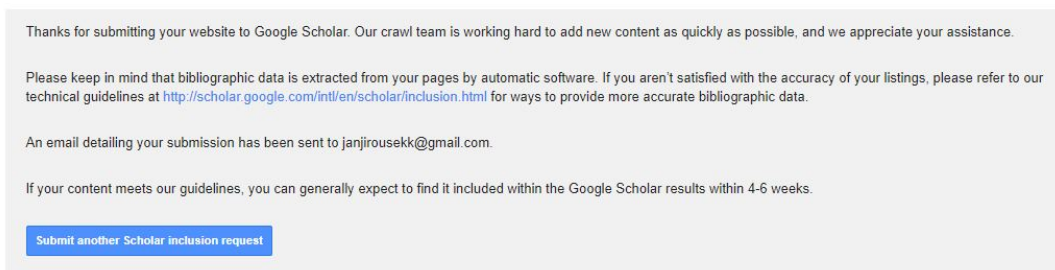
janjir.slovjani.info › scholar › article ▼  
**Proglasenje na pamet 100 let od osnovanja Češkoslovačskoj ...**  
Proglasenje na pamet 100 let od osnovanja Češkoslovačskoj republiky i diskusija nad vizijeju novogo strategičnogo cělja slovjanskyh narodov do budučnosti ...

isv.miraheze.org › wiki › Sbornik:Proglasenje\_na... ▼ Přeložit tuto stránku  
**Sbornik:Proglasenje na pamet 100 let od osnovanja ...**  
Proglasenje na pamet 100 let od osnovanja Češkoslovačskoj republiky :- takože ang.: «Proclamation commemorating 100 years since the founding of the ...

www.ceeol.com › search › article-detail - Přeložit tuto stránku  
**Proglasenje na pamet 100 let od osnovanja Češkoslovačskoj ...**  
autor: V Merunka - 2018  
Proglasenje na pamet 100 let od osnovanja Češkoslovačskoj republiky. Proclamation commemorating 100 years since the founding of the Czechoslovak ...

Obr. 45 Ukázka srovnání pořadí výsledků s konkurencí při hledání "Proglasenje na pamet 100 let"

## 16.5 Ukázka emailu po registraci repositáře do Google Scholar



**Obr. 46 Registrace repositáře - email po registraci**