|  |  |  |
| --- | --- | --- |
|  | | |
| **Závěrečná studijní práce**  **dokumentace** | | |
| **Elasticsearch a jeho využití ve Wordpressu** | | |
| Ondřej Kuděla | | |
|  | | |
|  | |  |
| **Obor:** | 18-20-M/01 INFORMAČNÍ TECHNOLOGIE  se zaměřením na počítačové sítě a programování | |
| **Třída:**  **Školní rok:** | IT4  2021/2022 | |

***Poděkování***

*Rád bych poděkoval panu učiteli Ing. Petru Grussmannovi za pomoc s výběrem projektu, a také za objasnění některých chyb, které mi ve většině případech bránili v postoupení v projektu.*

Prohlašuji, že jsem závěrečnou práci vypracoval samostatně a uvedl veškeré použité   
informační zdroje.

Souhlasím, aby tato studijní práce byla použita k výukovým účelům na Střední průmyslové   
a umělecké škole v Opavě, Praskova 399/8.

V Opavě 31. 12. 2021

*podpis autora práce*

**ANOTACE**

Tento projekt se zabývá Elasticsearchem a jeho využitím ve Wordpressu

OBSAH

[Úvod 6](#_Toc92670100)

[1 Využité technologie 7](#_Toc92670101)

[1.1 Elasticsearch 7](#_Toc92670102)

[1.2 Wordpress 7](#_Toc92670103)

[1.3 phpMyAdmin 8](#_Toc92670104)

[1.4 Kibana 8](#_Toc92670105)

[1.5 Docker 9](#_Toc92670106)

[2 Způsoby řešení a použité postupy 10](#_Toc92670107)

[2.1 Instalace WSL 2 (subsystém) 10](#_Toc92670108)

[2.2 Instalace Docker desktop 10](#_Toc92670109)

[2.3 Konfigurace Wordpressu 12](#_Toc92670110)

[2.4 Pluginy 14](#_Toc92670111)

[2.4.1 ElasticPress 14](#_Toc92670112)

[2.4.2 WooCommerce 14](#_Toc92670113)

[3 ELAsticpress 15](#_Toc92670114)

[3.1 Funkce 15](#_Toc92670115)

[3.2 Nastavení 15](#_Toc92670116)

[3.3 Index health 15](#_Toc92670117)

[3.4 Query log 15](#_Toc92670118)

[3.5 Vyhledávací pole a přidělování váhy 15](#_Toc92670119)

[3.6 Synonyma 16](#_Toc92670120)

[3.7 Vlastní výsledky 16](#_Toc92670121)

[4 cíle a chyby 17](#_Toc92670122)

[4.1 Splněné cíle 17](#_Toc92670123)

[4.1.1 Stav webu 17](#_Toc92670124)

[4.2 Nedokončené cíle 17](#_Toc92670126)

[Závěr 19](#_Toc92670127)

[Seznam použitýCH INFORMAČNÍCH ZDROJů 20](#_Toc92670128)

[Seznam příloh 21](#_Toc92670129)

Úvod

Toto téma projektu jsem si vybral, jelikož jsem si chtěl vyzkoušet práci se vyhledávacím enginem a zjistit, jak funguje a jaké jsou veškerá jeho využití. Protože jsem neměl dostatečné zkušenosti, tak jsem se nepouštěl do složitých projektů jako je například tvorba vlastního vyhledávacího enginu. Místo toho jsem dostal návrh od pana učitele Petra Grussmanna, abych se zaměřil na spojení Elasticseache (ES) a Wordpressu (WP).

Z důvodu přehlednosti pro uživatele si myslím, že by měl být vyhledávací engine skoro až nezbytnou součástí každé stránky, kvůli tomu, aby se v ní uživatel neztratil a stránku pak následně neopustil.

Hlavní cíle projektu byly tedy dva. První bylo pochopení ES jako takového. Druhý cíl byl plně funkční propojení ES s WP a v maximální možné míře využít plugin ElasticPress a všechny jeho funkce. Ten je v mém projektu hostován pomocí ES, a také vytvoření malého e-shopu pomocí pluginu WooCommerce, neboť si myslím, že na e-shopových stránkách se dá ES dobře využít.

# Využité technologie

## Elasticsearch

ES je vyhledávací fulltextový engine, kterých je postavený na Apache Lucene, což je open-source knihovna vydávaná pod licencí Apache. Slouží analýze a vyhledávání dat v reálném čase. K největším přednostem patří právě rychlost vyhledávání a analýzy. Analyzované data jsou uloženy v indexech.

  elasticsearch:

    image: docker.elastic.co/elasticsearch/elasticsearch:7.9.3

    container\_name: sem\_elasticsearch

    restart: always

    ports:

      - "9200:9200"

    # restart: on-failure

    environment:

      - discovery.type=single-node

      - bootstrap.memory\_lock=true

      - "ES\_JAVA\_OPTS=-Xms256m -Xmx256m"

    ulimits:

      memlock:

        soft: -1

        hard: -1

    volumes:

      - ./data/elasticsearch:/usr/share/elasticsearch/data

    logging:

        driver: "json-file"

        options:

            max-size: "10k"

            max-file: "10"

## Wordpress

WP je největší open source redakční a publikační systém naprogramovaný v PHP a MySQL. Je určený pro administraci a tvorbu webových stránek. Mezi jeho klíčové vlastnosti patří převážně uživatelsky přívětivé rozhraní (šablony, widgety, tagy, galerie, GUI), podpora pluginů, dodržování standartů XML, XMTML a CSS a také česká lokalizace.

wordpress:

    depends\_on:

      - db

    image: wordpress:latest

    restart: always

    ports:

      - "80:80"

    environment:

      WORDPRESS\_DB\_HOST: db:3306

      WORDPRESS\_DB\_USER: admin

      WORDPRESS\_DB\_PASSWORD: admin

      WORDPRESS\_DB\_NAME: wpdb

    volumes:

      ["./wordpress:/var/www/html"]

## phpMyAdmin

phpMyAdmin je nástroj napsaný v jazyce PHP, umožňuje jednoduchou správu databáze MySQL prostřednictvím webového rozhraní. Umožňuje databáze vytvářet nebo rušit, vytvářet, upravovat nebo rušit tabulky a také provádět SQL příkazy a spravovat klíče. Jedná se o jeden z nejpopulárnějších nástrojů pro správu databází.

  phpmyadmin:

    depends\_on:

      - db

    image: phpmyadmin/phpmyadmin

    restart: always

    ports:

      - '8070:80'

    environment:

      PMA\_HOST: db

      MYSQL\_ROOT\_PASSWORD: <root\_password>

## Kibana

Kibana je open source software, napsaný v JavaScriptu, slouží pro vizualizaci dat ve webovském prohlížeči. Byl napsán pro vizualizaci dat z vyhledávajícího stroje ES a spolu s nástrojem Logstash tvoří trojici zvanou Elastic Stack (ELK stack).

kibana:

    image: docker.elastic.co/kibana/kibana:7.9.3

    container\_name: sem\_kibana

    restart: always

    environment:

      - "ELASTICSEARCH\_URL=http://elasticsearch:9200"

      - "SERVER\_NAME=127.0.0.1"

    ports:

      - 5601:5601

    depends\_on:

      - elasticsearch

## Docker

Pro instalaci všech výše uvedených jsem využil technologii Docker, což je open source software, jenž vytváří jednotné rozhraní pro izolované kontejnery, ve kterých fungují různé aplikace. V mém případě se jedná o ES, WP, Kibanu a phpMyAdmin. Jednotlivé kontejnery obsahují požadovanou aplikaci spolu s jejími soubory. Mezi výhody patří menší velikost a flexibilita. Největší výhodou je to, že je multiplatformní, tudíž funguje na operačních systémech Windows, Linux i macOS.

# Způsoby řešení a použité postupy

## Instalace WSL 2 (subsystém)

Prvním důležitým krokem, kterým jsem začal byla instalace Windows subsystému pro Linux, jelikož mi usnadňovala práci. Začal jsem příkazem pro instalaci WSL.

**wsl –install**

Následně jsem musel povolit WSL a virtualizaci.

**dism.exe /online /enable-feature /featurename:Microsoft-Windows-Subsystem-Linux /all /norestart**

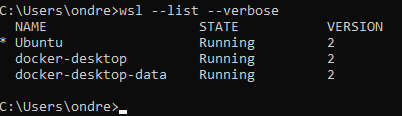
**dism.exe /online /enable-feature /featurename:VirtualMachinePlatform /all /norestart**

Dále jsem nastavil WSL 2 jako základní verzi WSL

**wsl --set-default-version 2**

Poté je potřeba stáhnout Ubuntu, například z Microsoft Store. Pro ověření instalace zadáme příkaz:

**wsl --list –verbose** (zkráceně **wsl -l -v**)



## Instalace Docker desktop

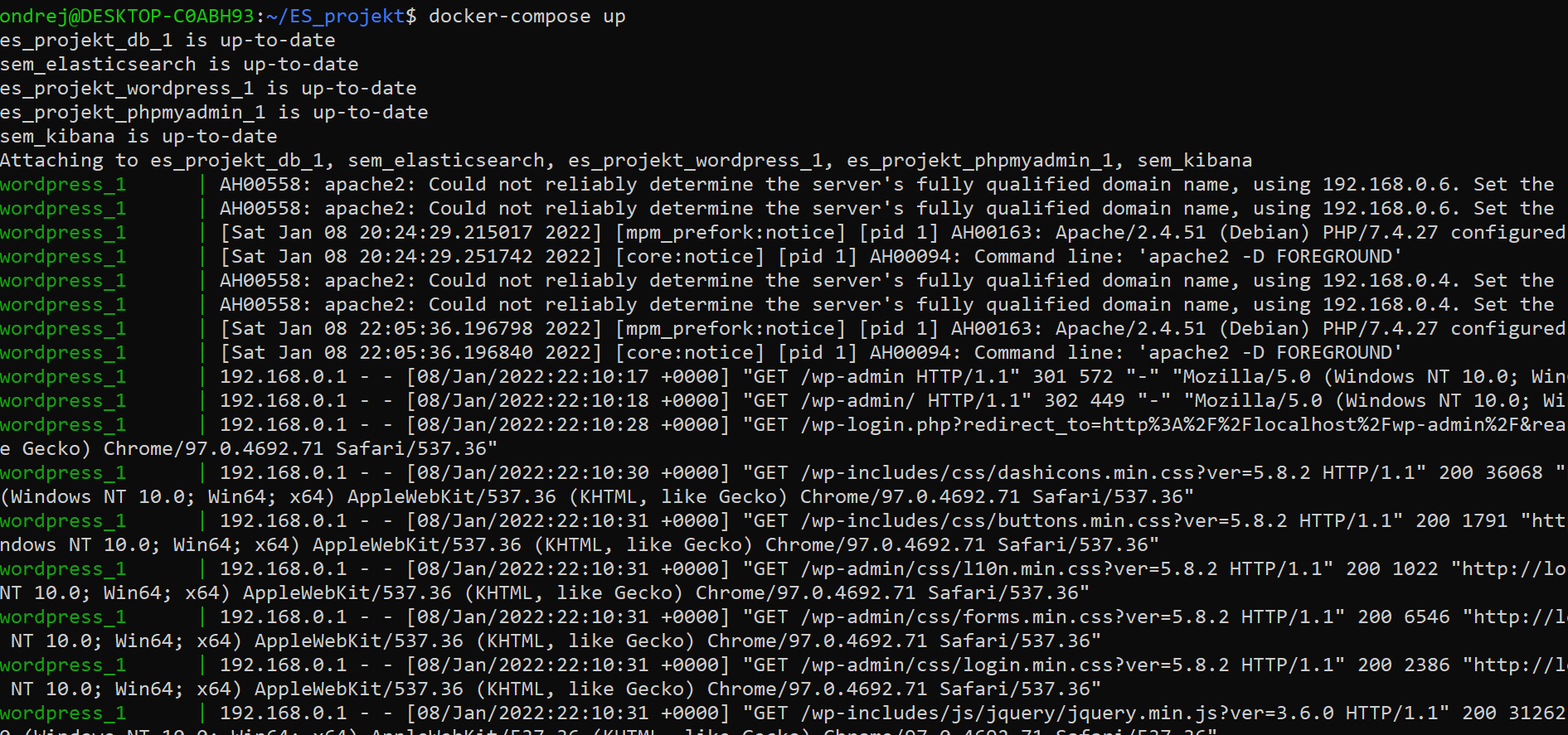
Jako další krok bylo třeba si nainstalovat Docker Desktop a správně ho nakonfigurovat. Musíme se ujistit, že instalace požadavků pro WSL 2 (podtrhnuto červeně) je povolena.



Po instalaci bychom měli mít funkční Docker Desktop, pro ověření můžeme zkusit spustit docker soubor, v mém případě „docker-compose.yaml“ (musíme se nacházet v dané složce) pomocí příkazu:

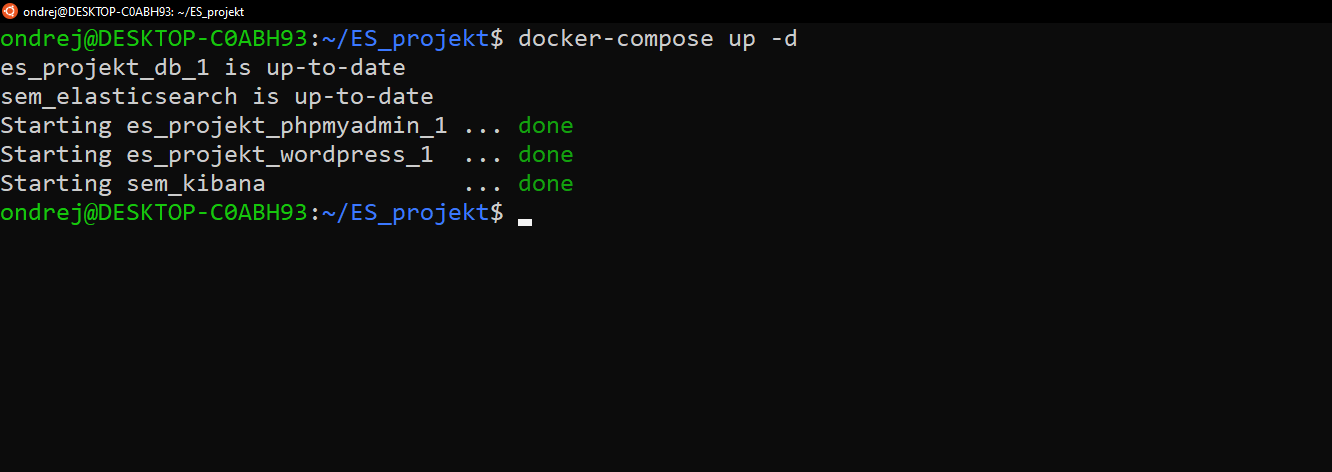
**docker-compose up**

Pro ukončení pak stačí **Ctrl + C**



Pro spuštění kontejneru je také možné za příkaz dodat -d, což nedoporučuji, neuvidíme případné chyby v terminálu a pro vypnutí musíme zadat příkaz:

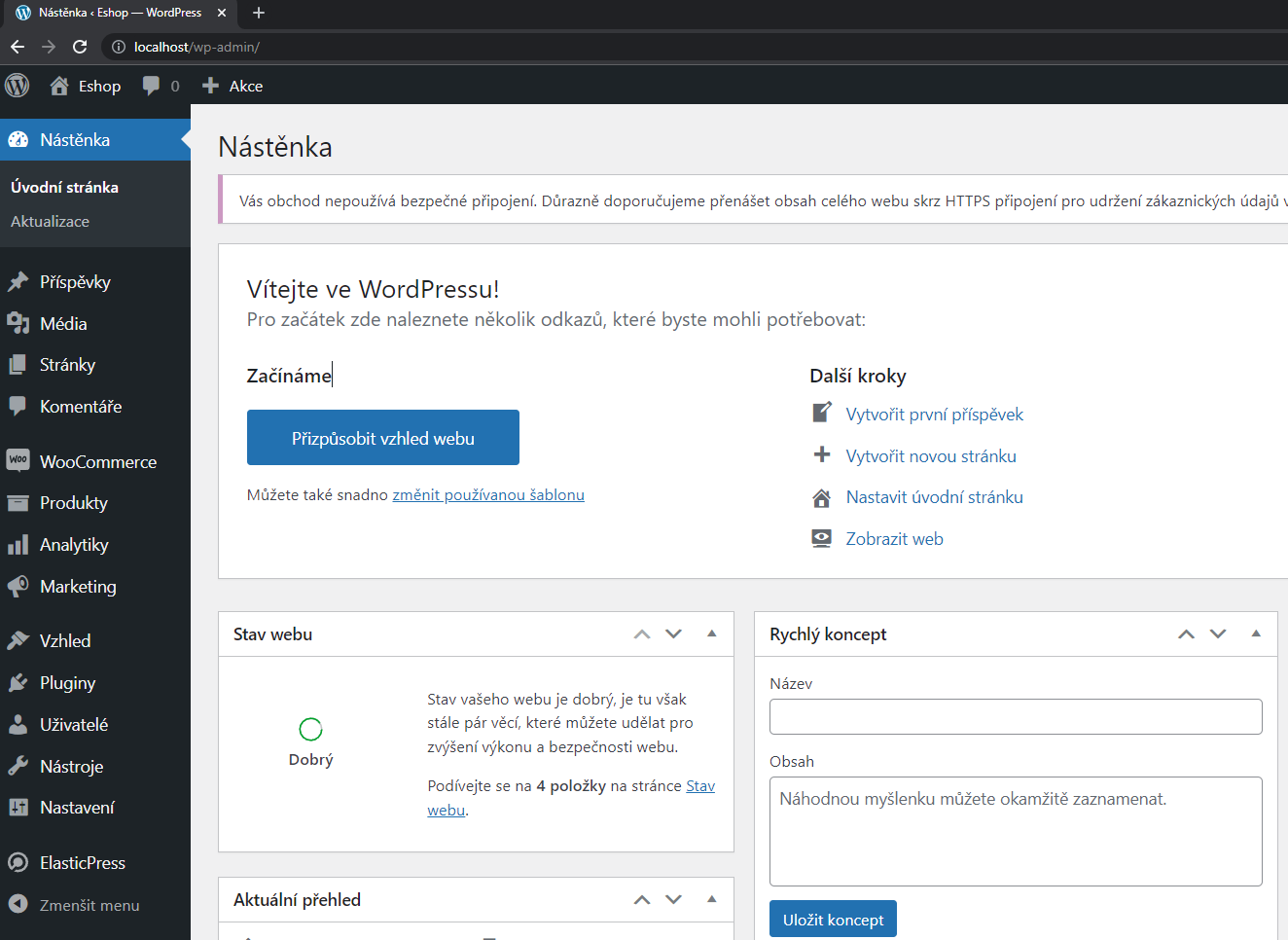
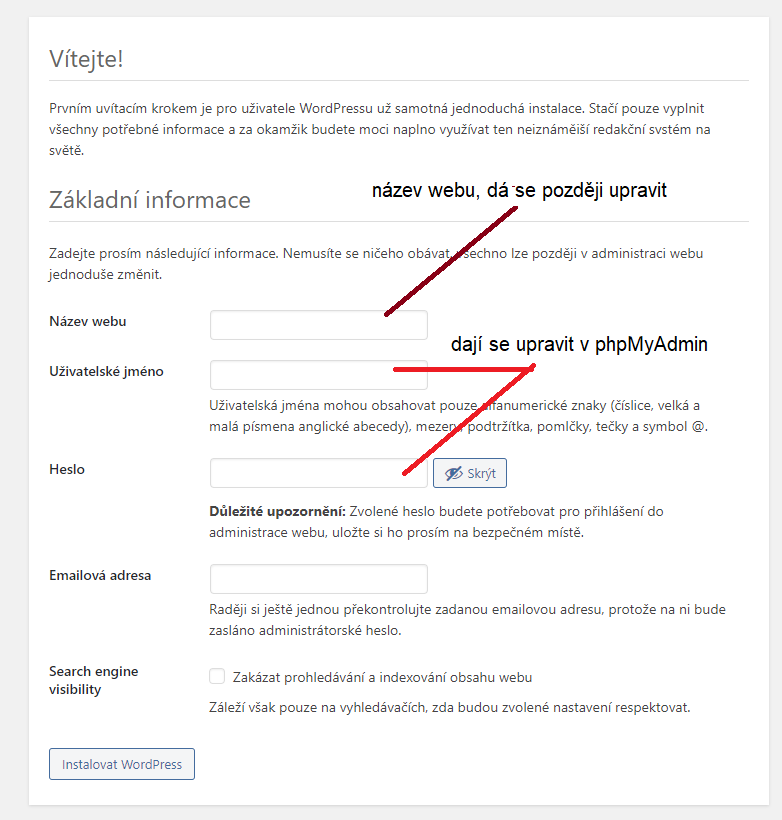
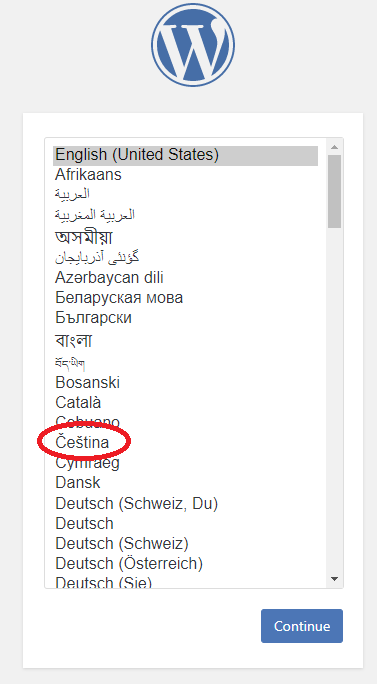
**docker-compose down**



## Konfigurace Wordpressu

První krok, který ve WP musíme udělat je nastavení lokalizace. WP disponuje velkou škálou jazyků mimo jiné i češtinou. V další části je potřeba pojmenovat stránku a vytvořit si administrativní účet, pomocí jej budeme stránku spravovat, je možnost si upravit heslo, v případě velmi slabého hesla musíme zaškrtnout kolonku. Po dokončení všech kroků se dostaneme do menu WP takzvanou nástěnku.

Nalevo se nachází všechny důležité záložky, uprostřed jsou pak různé tipy a možnosti, jak rychle vytvářet stránky, příspěvky či upravovat menu. Při kliknutí na tlačítko s názvem stránky v levém horním rohu se dostaneme na naší webovou stránku.

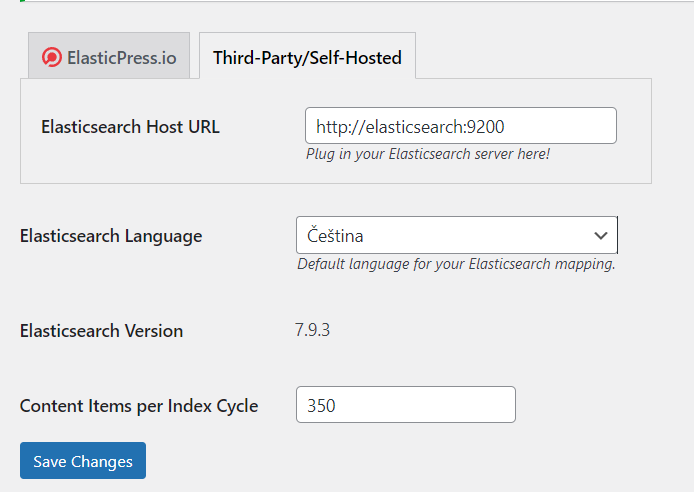


## Pluginy

Instalace pluginů je vskutku jednoduchá v levé liště najedeme na „pluginy“ a rozklikneme „instalace pluginů“, poté stačí napsal do vyhledávání název hledaného pluginu a dát ho instalovat, můžeme taky spustit automatické aktualizace či nepoužívané pluginy smazat.

### ElasticPress

Poté co máme nainstalovaný plugin ElasticPress je třeba v nastavení nastavit správný hosting pro náš plugin, ten se nastavuje tak, že zadáme [*http://[název*](http://[název) *ES kontejneru: a port na kterém ES běží]*, výsledek bude vypadat nějak takto:



ElasticPress nám poté sám nabídne indexaci prvků, poté můžeme již používat jeho funkce a vše co plugin nabízí.

### WooCommerce

Tento plugin nevyžaduje žádný zvláštní hosting, slouží k velmi jednoduché tvorbě e-shopu a jeho provozování. Obsahuje prakticky vše, co pro tvorbu e-shopu potřebujete

# ELAsticpress

## Funkce

ElasticPress má plno dostupných funkcí pro vylepšení vyhledávání na stránce či k přehlednosti minulosti vyhledávání. Většinou funkcí jde jednoduše zapnout tím, že jí přepneme ze stavu „zakázáno“ na stav „povoleno“, ve většině případů je pak funkčnost okamžitá, můžou se zde ale vyskytovat občasné chyby. U některých funkcí to ovšem není tak snadné, je třeba doplnit některé potřebná data jako například enpoint u autossuggestu.

## Nastavení

V nastavení ElasticPressu se nastavuje ES host, dále jazyk pro ES, vypíše se nám zde naše verze.

## Index health

Představuje vizualizaci zdraví naší stránky, příspěvků a termínů. Disponuje taky grafy, ty ukazují celkový počet Query a Indexů a také jejich časy. Dále se zde zobrazuje celkový počet dokumentů jejich velikost a také paměť.

## Query log

## Vyhledávací pole a přidělování váhy

Můžeme zde nastavovat všechny atributy stránek, příspěvků a v mém případě i produktů, zda jsou vyhledatelné a jako mají váhu při hledání. To znamená, že čím větší váhu mají, tím se budou ve vyhledávání řadit na přední příčky. Pokud tedy správce stránky chce, aby něco uživatele viděli při určitém vyhledávání, tak jim jednoduše zvýší váhu. To samé platí, pokud správce vidí, že je něco populární, často vyhledávané, tak se tomu přidá větší hodnota, aby měl uživatel snadnější hledání.



## Synonyma

Dělí se na:

* Sady

Jsou skupiny ekvivalentních hledaných výrazů. Hledání dané položky v řadě vrátí výsledek odpovídající jakékoliv položce v dané řadě.

Pokud tedy máme sadu boty, tenisky, kopačky, tak při hledání jakéhokoliv slova ze sady by se vyhledaly i ostatní slova z dané sady.

* Alternativa

Hledání primárního výrazu vrátí výsledek obsahující kterýkoliv z alternativních výrazů, na rozdíl od sad hledání alternativních výrazů nevrátí výsledky pro primární výraz.

Příkladem primárního výrazu může být například květina a alternativní růže a kaktus. Pokud uživatel vyhledá květinu, ElasticPress vrátí také výsledky obsahující růže a kaktus. Pokud však uživatel vyhledá kaktus nebo růži, tak ElasticPress vrátí výsledek pouze hledané věci.

## Vlastní výsledky

ElasticPress umožňuje přizpůsobení výsledků vyhledávání pro konkrétní dotazy, aby bylo možné zdůraznit, propagovat nebo snížit důraz na konkrétní výsledky vyhledávání.

# cíle a chyby

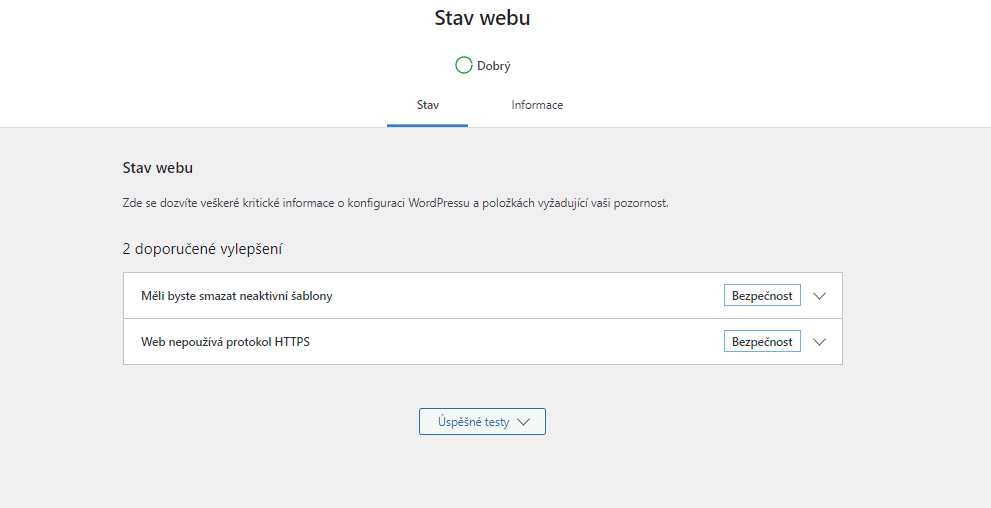
## Splněné cíle

Hlavní cíle splněny byly nebo tedy aspoň z větší části, ať už se jedná o pochopení fungování ES, práce s query a indexy v Kibaně nebo o samotnou integraci s WP. Také se mi podařilo zprovoznit většinu funkcí jenž ElasticPress nabízí, což bych řekl, že je důležité, jelikož se to podepisuje na vizuální stránce při vyhledávání.

### Stav webu

Oprava všech zásadních chyb byla taky splněným cílem, jenž jsem si stanovil, ať už se jednalo o chybu curl 7, ta znamená, že se nepodařilo připojit na hostovací server, tuto chybu jsem opravil tím, že jsem WP, který běžel na portu 8000 přesměroval na port 80, což zmíněnou chybu, ale posléze se hned vyskytla další fatální chyba, a to kvůli mnohonásobnému redirectu neboli že WP byl příliš mnohokrát přesměrován, chybu jsem vyřešil přeinstalováním WP.

## 



## Nedokončené cíle

Jeden z menších cílů, který mi přijde dosti užitečný se mi, ale zprovoznit nepodařilo, jedná se o autosuggest, jedná se o našeptávání uživateli při psaní do vyhledávajícího pole. K využití této funkce je totiž za potřebí nastavit konečný bod, takzvaný endpoint, Díky práci v Kibaně chápu alespoň funkčnost a zásady, jimiž se musíme držet, aby autossuggest fungoval a napovídal tak jak by měl.

Dále bych k mínusům zahrnul logy, jež jsou také součástí ElasticPressu nebo Kibany neboť se mi také nepovedly plně zprovoznit a tudíž se v dashboardu nevyskytují.

# Závěr

Text závěru

* povinná část,
* shrnuje výsledky, hodnotí splnění cíle práce, uvádí možnost uplatnění řešení v praxi a nastínění případných dalších budoucích vylepšení
* kapitola se nečísluje (stejné jako úvod)

Seznam použitýCH INFORMAČNÍCH ZDROJů

[1] BOHMAN, Ludvík. Zákon o pojistné smlouvě. Praha: Linde Praha a. s., 2004. 381 s. ISBN80-7201-504-4

[2] DUCHÁČKOVÁ, Eva. Principy pojištění a pojišťovnictví. 3. aktualizované vydání. Praha: Ekopress 2009. 224 s. ISBN 978-80-86929-51-4

[3] KUBALA, Petr. Planetární dvojcata - Věda a technika (Český rozhlas) [online].   
Č. 2000-2008, poslední revize 19. 3. 2008 [cit. 2008-03-20].  
<http://www.rozhlas.cz/veda/vesmir/\_zprava/435849>.

[4] KULDOVÁ, O., FLEISCHMANNOVÁ, E. Metodická příručka k technice administrativy a obchodní korespondence. 1.vyd. Praha: Fortuna 1998. 111 s.   
ISBN 80-7168-574-7. Kapitola 6, Metody nácviku psaní hmatovou metodou,   
s. 28-29.

[5] VLACH, J. JE Temelín a zásobování teplem. Energetika, 2001, roč. 51, č. 3, s. 84 -85. ISSN 0375-8842.

* musí zahrnovat všechny prameny, knihy, internetové odkazy a další studijní podklady, z nichž jsme čerpali;
* kapitola se nečísluje a zde končí číslování stránek práce;
* jednotlivé publikace se uvádějí v abecedním pořadí podle příjmení autorů a iniciál jeho jména, který se píše za čárkou;
* příjmení autora se píše velkými písmeny;
* název publikace se zvýrazňuje kurzívou;
* jestliže jsou uvedeni více než tři autoři, je možné vypsat hlavního autora s poznámkou „a kol.“(a kolektiv).

Seznam příloh

č. 1 Titulní list

č. 2 Čestné prohlášení

č. 3 Poděkování

Nepovinná část – pokud nemáte žádné přílohy ke své práci, tuto část odstraňte!

* Přílohy se zařazují na konec práce.
* Jsou to texty, obrázky, grafy, tabulky, které by přímo v textu byly zbytečně detailní, ale mají být po ruce k dokreslení východisek i výsledku řešení.
* Jsou číslovány a v textu se na ně může odkazovat.
* Před první přílohu se umisťuje seznam příloh.
* Každá příloha je označena číslem - např. Tabulka č.. 1, Schéma č. 2, Obrázek č. 3.
* Každá tabulka by měla mít i vlastní název, který stručně vystihuje její obsah.
* (Tabulka č. 1 Zakázky stavebních prací v roce 2009-2010).
* Pokud je z tabulky vytvořen graf, umístíme jej na stejné stránce jako tabulku.

**Příloha č. 1: Titulní list**