

# SEHR

datové úložiště osobní elektronické zdravotní dokumentace

Tým: Trained Monkeys

Členové:

|         |           |                          |
|---------|-----------|--------------------------|
| Martin  | Graubner  | graubnerm@email.cz       |
| Markéta | Wolfová   | market.wolfova@gmail.com |
| Adam    | Barák     | light.fire@seznam.cz     |
| Marek   | L'uptáčík | maraluptacik@gmail.com   |

## Obsah

|   |           |
|---|-----------|
| <b>1. ZADÁNÍ.....</b>   | <b>3</b>  |
| 1.1 CÍL PROJEKTU .....  | 3         |
| 1.2 ROZSAH PROJEKTU .....                                     | 3         |
| 1.4 SLOVNÍČEK DEFINIC, POJMŮ A ZKRATEK .....                  | 3         |
| 1.5 PROVOZNÍ PROSTŘEDÍ .....                                  | 3         |
| 1.5.1 Omezení návrhu a implementace .....                     | 4         |
| <b>2. KONTEXT A ARCHITEKTURA SYSTÉMU .....</b>                | <b>4</b>  |
| 2.2 ARCHITEKTURA SYSTÉMU, PŘEHLED PODSYSTÉMŮ .....            | 4         |
| 2.3 ZVOLENÁ TECHNOLOGIE, PROGRAMOVACÍ JAZYK AD., DŮVODY ..... | 4         |
| <b>3. TYPY INFORMACÍ ZPRACOVÁVANÉ SYSTÉMEM.....</b>           | <b>5</b>  |
| <b>4. PODSYSTÉMY.....</b>                                     | <b>5</b>  |
| 4.1 KNIHOVNA PRO KONVERZI ARCHETYPŮ .....                     | 5         |
| 4.1.1 Přehled tříd.....                                       | 5         |
| 4.1.2 Třída Loader.....                                       | 5         |
| 4.1.3 Třída DataParser.....                                   | 5         |
| 4.1.4 Třída Converter .....                                   | 6         |
| 4.2 KNIHOVNA PRO KOMUNIKACI S ELASTICSEARCH.....              | 6         |
| 4.2.1 Přehled tříd.....                                       | 6         |
| 4.2.2 Třída Elasticsearch.....                                | 6         |
| 4.3 KLIENT – KOMUNIKACE S UŽIVATELEM.....                     | 6         |
| 4.3.1 Přehled tříd.....                                       | 7         |
| 4.3.2 Třída ArchetypeManager.....                             | 7         |
| 4.3.3 Třída DatabaseManager .....                             | 7         |
| <b>5. TESTOVÁNÍ.....</b>                                      | <b>7</b>  |
| <b>TESTOVACÍ PROTOKOL.....</b>                                | <b>7</b>  |
| INFORMACE O TESTOVÁNÍ.....                                    | 7         |
| TESTOVACÍ PROSTŘEDÍ .....                                     | 7         |
| TESTOVACÍ SCÉNÁŘE .....                                       | 8         |
| Test ve Windows 7 .....                                       | 8         |
| Test ve Windows 8 .....                                       | 9         |
| Test v Debianu .....  | 9         |
| <b>7. EXTERNÍ MODULY .....</b>                                | <b>10</b> |
| 7.3. Postman .....  | 10        |
| 7.3. HDC.....   | 10        |
| 7.3. Curl.....  | 10        |
| 7.3. Hadoop .....   | 10        |

# 1. Zadání

Vytvořte datové úložiště osobní elektronické zdravotní dokumentace, které umožní zpracovávat archetypy openEHR a vytvářet omezení v databázi elasticsearch.

## 1.1 Cíl projektu

Cílem projektu je vytvořit funkční systém osobní elektronické zdravotní dokumentace, tedy jednotného úložiště, kam si uživatel-pacient ukládá svá naměřená data. Těmito daty mohou být výsledky vyšetření, např. váha, tlak, EKG, ale i hodnoty naměřené aplikacemi jako jsou „sporttrackery“, krokoměry, spánkové monitory. Systém je zaměřen ryze na pacienty, nikoliv lékaře.

Tento systém bude schopen vypořádat se s heterogenitou dat a automaticky se přizpůsobit jejich struktuře. Je nutné, aby veškerá data byla anotována a byl tak jasně zachován jejich význam a příslušnost ke konkrétnímu archetypu (více o archetypech dále). Pro tyto účely byla zvolena nerelační databáze Elasticsearch.

Systém bude založen na standardu openEHR<sup>a)</sup>. Jedná se o otevřený standard elektronické zdravotní dokumentace a je tvořen právě archetypy, které zajišťují jednoznačný popis jednotlivých medicínských domén.

Vize do budoucna tedy je, že si uživatel-pacient propojí jím využívanou aplikaci s databází (nebo data zadá ručně pomocí vytvořeného GUI), kde se z poskytnutých ověřených archetypů automaticky připraví indexy pro ukládaná data a jejich mapping. Data jsou poté uložena, anotována ontologiemi a připravena k případné analýze či sdílení.

## 1.2 Rozsah projektu

Cílem naší práce bude vytvořit prototyp úložiště osobní elektronické zdravotní dokumentace. První část práce bude načíst archetypy a pomocí již existujících knihoven konvertovat jejich formát z ADL na JSON<sup>b)</sup>, který poté upravíme tak, aby se podle něj dala připravit databáze na ukládání nových dat. Minimální rozsah této části je zpracování archetypů typu item cluster, item tree, entry evaluation a entry observation. Zpracování zbylých typů není prioritou, ale jsou samozřejmě vítány, pokud na ně zbyde čas.

Druhá část obsahuje klient pro zajištění komunikace s databází (pomocí dotazů). Program umožní přidání nového archetypu a po jeho zpracování, které zajišťuje první část práce, vytvoří pro nová data v databázi indexy a jejich mapping. Tento program zároveň bude plnit kontrolní činnost, tedy zajistí, že se nevytvoří duplicitní indexy podle stejného, či jen částečně pozměněného archetypu.

Třetí část zahrnuje klient, který bude výše uvedené programy využívat ve formě knihoven. Pomocí konzole bude také komunikovat s uživatelem.

Tento prototyp bude po dokončení dále rozvíjen skupinou Neuroinformatics Research Group.

## 1.4 Slovníček definic, pojmů a zkratk

- a) EHR (Electronic Health Record): elektronická zdravotní dokumentace
- b) JSON (JavaScript Object Notation): datový formát pro výměnu dat
- c) Archetypy: soubory ve formátu ADL, který zajišťují jednoznačný popis jednotlivých medicínských domén

## 1.5 Provozní prostředí

Systém poběží na běžném stolním počítači bez specifických požadavků na hardware. Bude optimalizován pro operační systémy Windows verze 7, 8 a Linux Debian. Databáze pro účely našeho prototypu poběží na virtuálním stroji. Nutná je komunikace s databází Elasticsearch verze 1.3.x/1.4.x.

### 1.5.1 Omezení návrhu a implementace

- Programovací jazyk: Java
- Formáty dat: archetypy jsou ve formátu ADL a zprávy pro ElasticSearch v JSON formátu
- Archetypy: budeme pracovat s již existujícími archetypy, nebudeme vytvářet nové
- Ukládání dat: nerelační databáze ElasticSearch a její API

## 2. Kontext a architektura systému

### 2.2 Architektura systému, přehled podsystémů

Systém je rozdělen do tří částí. Prvním podsystémem je program, který zajistí načtení a konverzi archetypů, druhým pak klient pro komunikaci s databází ElasticSearch, který provede kontrolu archetypu a vytvoří indexy pro uložení dat. Tyto dvě části jsou na sobě nezávislé. Třetím podsystémem je klient zajišťující komunikaci s uživatelem pomocí konzole, který zároveň bude využívat předchozí dva podsystémy formou knihoven.

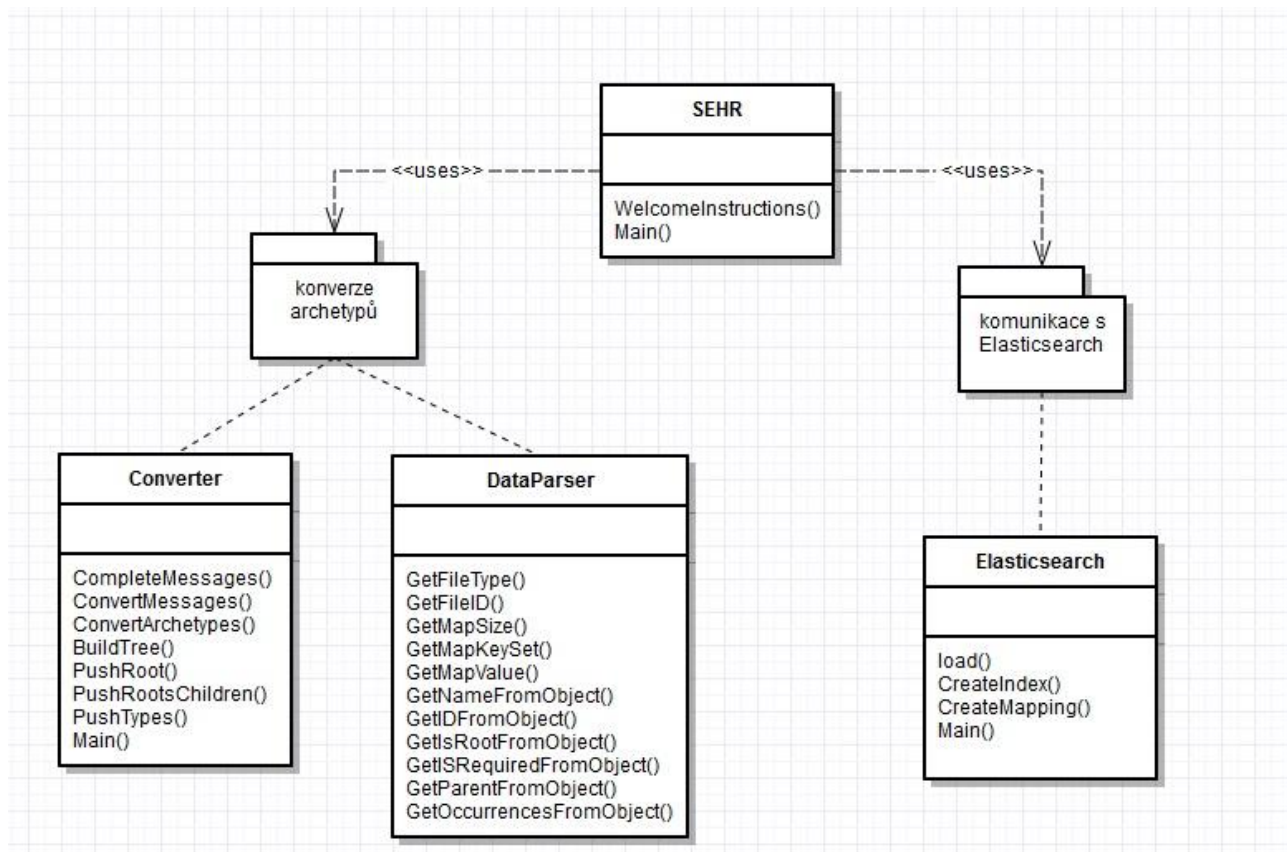


Diagram tříd

### 2.3 Zvolená technologie, programovací jazyk ad., důvody

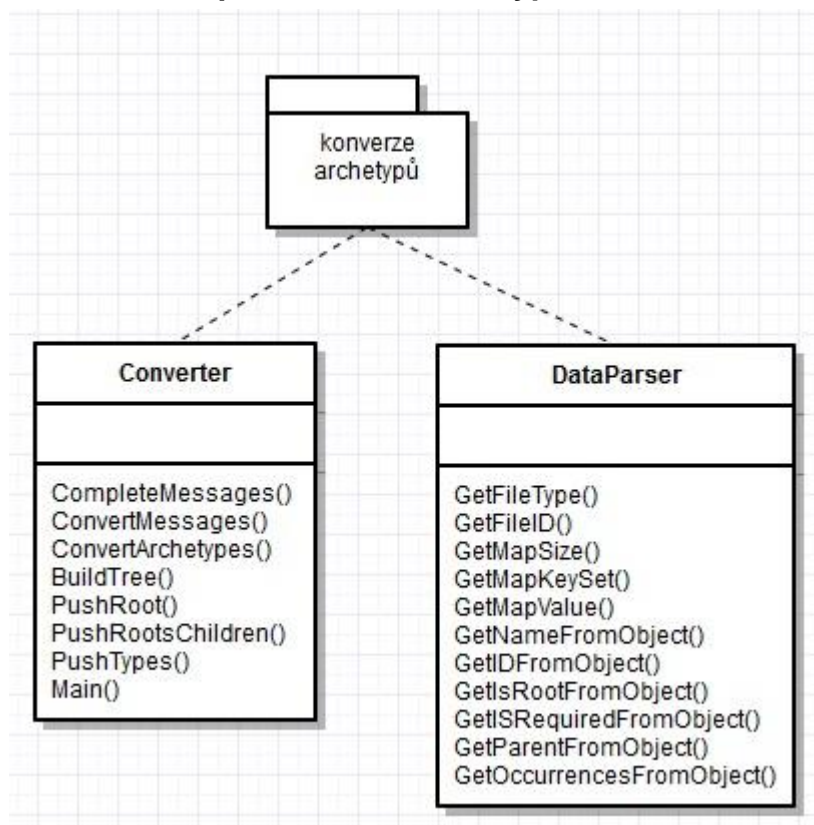
Systém je vyvíjen v programovacím jazyce Java. Využívá archetypy standardu openEHR<sup>a)</sup> a pracuje s nerelační databází ElasticSearch. Všechny technologie byly zvoleny dle požadavku zadavatele.

### 3. Typy informací zpracovávané systémem

Vstupními soubory jsou archetypy standardu openEHR v textovém formátu ADL. S těmito soubory bude pracovat první podsystém, který daný soubor rozparsuje. Z dat, která z archetypů získá, doplněných o příkazy vytvoří soubor formátu JSON<sup>b)</sup> (tento formát je nutností pro komunikaci s databází ElasticSearch). Ten bude poté načten druhým podsystémem a budou podle něj vytvořeny indexy v databázi.

## 4. Podsystémy

### 4.1 Knihovna pro konverzi archetypů



#### 4.1.1 Přehled tříd

- Loader
- DataParser
- Converter

#### 4.1.2 Třída Loader

Tato třída načte archetypy ze složky zadané uživatelem a poté je pomocí externí knihovny rozparsuje na objekty.

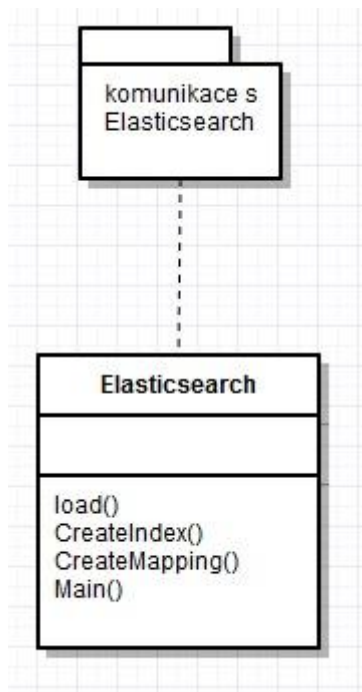
#### 4.1.3 Třída DataParser

Tato třída získá relevantní data z rozparsovaného souboru a uloží je v textovém formátu.

#### 4.1.4 Třída Converter

Tato třída k uloženým datům doplní příkazy pro vytvoření indexů v databázi a uloží výsledný soubor ve formátu JSON.

### 4.2 Knihovna pro komunikaci s Elasticsearch



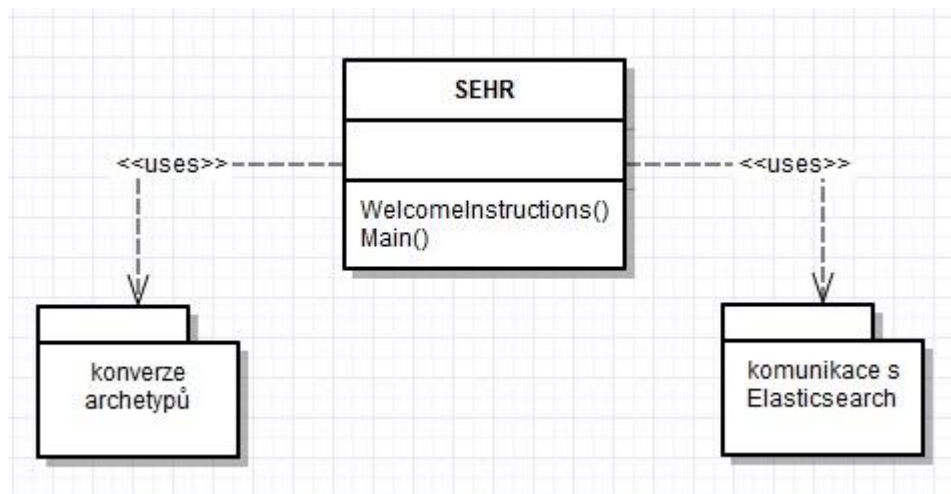
#### 4.2.1 Přehled tříd

- Elasticsearch

#### 4.2.2 Třída Elasticsearch

Tato třída zajišťuje veškerou komunikaci s databází ElasticSearch.

### 4.3 Klient– komunikace s uživatelem



### 4.3.1 Přehled tříd

- ArchetypeManager
- DatabaseManager

### 4.3.2 Třída ArchetypeManager

Tato třída zajistí pomocí knihovny pro konverzi archetypů načtení a rozparsování souborů. Přitom bude zajišťovat komunikaci s uživatelem.

### 4.3.3 Třída DatabaseManager

Tato třída zajistí pomocí knihovny pro komunikaci s databází vytvoření indexů pro nová data. Přitom bude zajišťovat komunikaci s uživatelem.

## 5. Testování

### Testovací protokol

#### Informace o testování

|                            |                                    |
|----------------------------|------------------------------------|
| <b>Testované programy:</b> | Converter<br>Sehr<br>Elasticsearch |
|----------------------------|------------------------------------|

#### Testovací prostředí

|                               |   |
|-------------------------------|---|
| <b>Testovací stroj č. 1:</b>  | HW: Intel Core i3-3240 CPU @ 3.40GHz, 4GB RAM DDR3,<br>harddisk 150GB SATA<br>OS: Win 7 ultimate x64<br>SW: Java 1.8.0_31 |
| <b>Testovací stroj č. 2 :</b> | HW: Intel Core i3-4030U CPU @ 1.90GHz, 4GB RAM<br>DDR3, harddisk 500GB SATA<br>OS: Win 8.1 x64                            |
| <b>Testovací stroj č. 3:</b>  | HW: Intel Core i3-3240 CPU @ 3.40GHz, 4GB RAM DDR3,<br>harddisk 150GB SATA<br>OS: Win 7 ultimate x64<br>SW: Java 1.8.0_31 |
| <b>Testovací stroj č. 4:</b>  | HW: Intel Core i3-4030U CPU @ 1.90GHz, 4GB RAM<br>DDR3, harddisk 500GB SATA<br>OS: Win 8.1 x64                            |

|   |   |
|---|---|
| <b>Virtuální stroj č. 1<br/>běžící na stroji č. 3</b> | OS: Linux Debian 6.0.6 alias squeeze<br>SW: Java 1.8.0_31 |
|---|---|

| Kompatibilita                         |  |  |
|---------------------------------------|--|--|
| <b>Podporované operační systémy:</b>  | Windows 7<br>Windows 8<br>Linux Debian |  |
| <b>Podporované webové prohlížeče:</b> | není relevantní                        |  |
| <b>Další vyžadovaný SW:</b>           | Java 7 nebo 8<br>Elasticsearch         |  |

## Testovací scénáře

### Test ve Windows 7

Datum testování: 24. 4. 2015

Použitý testovací stroj: č. 1

|     | Činnost                                     | Výsledek |
|-----|---|----------|
| 1.  | Spuštění sehr                               | ok       |
| 2.  | Načtení správné cesty ke složce s archetypy | ok       |
| 3.  | Načtení špatné cesty ke složce s archetypy  | ok       |
| 4.  | Manuální kontrola 3 náhodných archetypů     | ok       |
| 5.  | Spuštění es                                 | ok       |
| 6.  | Vytvoření indexu                            | ok       |
| 7.  | Vytvoření typu                              | ok       |
| 8.  | Vytvoření mapping                           | ok       |
| 9.  | Přidání dokumentu                           | ok       |
| 10. | Dotaz na dokument                           | ok       |



### Test ve Windows 8

Datum testování: 20. 4. 2015

Použitý testovací stroj: č. 2

|     | Činnost                                     | Výsledek |
|-----|---|----------|
| 1.  | Spuštění sehr                               | ok       |
| 2.  | Načtení správné cesty ke složce s archetypy | ok       |
| 3.  | Načtení špatné cesty ke složce s archetypy  | ok       |
| 4.  | Manuální kontrola 3 náhodných archetypů     | ok       |
| 5.  | Spuštění es                                 | ok       |
| 6.  | Vytvoření indexu                            | ok       |
| 7.  | Vytvoření typu                              | ok       |
| 8.  | Vytvoření mapping                           | ok       |
| 9.  | Přidání dokumentu                           | ok       |
| 10. | Dotaz na dokument                           | ok       |

### Test v Debianu

Datum testování: 7. 5. 2015

Použit virtuální stroj

|    | Činnost                                     | Výsledek |
|----|---|----------|
| 1. | Spuštění sehr                               | ok       |
| 2. | Načtení správné cesty ke složce s archetypy | ok       |
| 3. | Načtení špatné cesty ke složce s archetypy  | ok       |
| 5. | Spuštění es                                 | ok       |
| 6. | Vytvoření indexu                            | ok       |
| 7. | Vytvoření typu                              | ok       |
| 8. | Vytvoření mapping                           | ok       |

## 7. Externí moduly

### 7.3. Postman

Client slouží k jednodušší práci s databází(rest client)

<https://chrome.google.com/webstore/detail/postman-rest-client-packa/fhbjgbiflinjbdggehcddcbncdddomop>

Client slouží k jednodušší práci s databází

### 7.3. HDC

Client slouží k jednodušší práci s databází(rest client)

<https://chrome.google.com/webstore/detail/dhc-resthttp-api-client/aejoelaoggembcahagimdiliamlcdmfm>

### 7.3. Curl

Slouží k komunikaci s elasticsearch přes příkazový řádek

<http://curl.haxx.se/download.html>

### 7.3.Hadoop

Slouží k vizualizaci a dotazování na data

<https://www.elastic.co/downloads/hadoop>