



# Posudek oponenta diplomové práce

**Jméno studenta:** **Bc. Michal Poul**  
**Téma práce:** **Vícevláknová serverová aplikace parsující veřejně dostupné webové stránky**

Primárním cílem diplomové práce bylo provedení analýzy požadavků, návrh a implementace multiplatformní serverové aplikace umožňující skenování a následné parsování informací ze zadaných, veřejně dostupných webových stránek. Takto získané informace poté ukládá na domácí server do vhodné databáze.

Diplomant se ve své práci nejdříve zaměřuje na popis programovacího jazyka C++ a jeho standardů. Následně popisuje nástroj CMake, možnosti využití souběžnosti a značkovací jazyky. V neposlední řadě uživatel ukazuje základní práci s knihovnamy, které používá v rámci aplikace. Diplomant ve své práci využívá znalosti z oblasti pokročilých programovacích technik, programovacího jazyka C++ a databázových systémů.

V teoretické části diplomant provedl podrobný popis týkající se možnosti přenositelnosti kódu v programovacím jazyce C++ a vybraných části standardu C++11. V následující části autor popisuje práci s nástrojem CMake pro automatizaci překladu v různých operačních systémech nezávisle na typu kompilátoru. Dále se diplomant zaměřil na využití souběžnosti v programovacím jazyce C++, značkovací jazyky se zaměřením na HTML a XML, a na vybrané knihovny. Zde jsou uvedeny knihovny Boost, kolekce moderních knihoven, které doplňují standardy programovacího jazyka C++, knihovna cURL pro přenos dat přes internetové protokoly, nástroj Google C++ Testing Framework, který slouží pro unit testování kódu napsaného v programovacím jazyce C++, knihovnu MySQL C++ Connector pro komunikaci s databází MySQL, logovací knihovna Log4Cplusplus a knihovna PugiXML, která je využívána pro práci se značkovacím jazykem XML. Na závěr teoretické části diplomant popisuje průběh zpracování webové stránka a jednotlivé části aplikace. Teoretická část diplomové práce je zbytečně velká vůči praktické části a to v poměru 51 stránek k 16 stránkám. Diplomant se zde zbytečně moc zabývá některými kapitolami (Kapitola 3 a Kapitola 6). Tyto kapitoly mohly být spíše zpracovány v praktické části na ukázce použití a výběru. Autor dále v závěru zmiňuje problém ohledně kódování textu, který nikde v práci nebyl popsán.

V praktické části DP diplomant implementoval aplikaci pro parsování webových stránek elektronického obchodu, která splňuje veškeré kladené požadavky. Aplikace je vytvořena v programovacím jazyce C++ a využívá techniky nejnovějšího standardu C++11. Za vývojové prostředí bylo zvoleno Microsoft Visual Studio 2013 Professional pro možnost

kompilování na ostatních platformách diplomant použil nástroj CMake. Aplikace dále využívá všechny knihovny, které byly popsány v teoretické části DP. Aplikace nenabízí žádné možné ovládání pro uživatele a pracuje pouze s konkrétní webovou stránkou a databází. Pro možnosti přidání vlastní konfigurace je potřeba přistoupit do zdrojového kódu aplikace a daný plugin zde vytvořit. Průběh provádění parsování a zachytávání změn do databáze je zachycen v logovacím souboru.

Stanovené cíle podle osnovy DP byly splněny s tím, že analýza požadavků zde byla provedena částečně. Diplomant analýze požadavků vyhradit pouze výčet požadavků na straně 63. Řídící program je funkční a byl úspěšně otestován.

Text DP po formální stránce obsahuje následující nedostatky:

- Absence odkazů na některé obrázky a tabulky (např. Obrázek 1 a Tabulka 1).
- U některých zkratk chybí vysvětlení, co znamenají (např. ISO, ANSI).
- V textu se nachází řada gramatických chyb, chybějících čárek a překlepů (např. „Boost knihovny má na starost komunita, která je složena z velké skupiny. Vývojářů C++ z celého světa.“, slovo „zezačátku“, název vývojového prostředí název „Microsoft Visual Studio Profesione 2013“).
- Některé věty v textu nedávají smysl (např. „V těle lambdy je přístup ke všem kopiím proměnným.“, „Ve funkci *uziti* je deklarován *unique\_ptr*, ...“ – kde *unique\_ptr* není proměnná ale šablonová třída).
- Na straně 66 se odkazuje na kapitolu 0, která se v práci nenachází.
- Špatné názvy tříd (např. Třída ThreadGroup je na některých místech zmíněna špatně jako Thread group).
- Nadpisy 3. a 4. úrovně jsou modrou barvou oproti ostatním nadpisům.
- Ve zdrojích se vícekrát uvádí za zdroj Wikipedie.

Text práce je doplněn vhodně uspořádanými přílohami. Na přiloženém CD můžete nalézt podrobnou dokumentaci ve formátu HTML. Řešení *web-parser-msvc12* drží odkazy na neplatné soubory (*F:\Dropbox\...*) *projektům*.

Otázka k práci

- Jaké jsou rozdíly mezi kódováním UTF-8 a UTF-16?

**Doporučení k obhajobě: ANO**

**Hodnocení: VELMI DOBŘE MINUS**

Oponent: Ing. Emil Řezanina  
Katedra softwarových technologií FEI, Univerzita Pardubice

V Pardubicích dne: 5. června 2014

Podpis: Emil Řezanina