Západočeská univerzita v Plzni

Fakulta aplikovaných věd

Katedra informatiky a výpočetní techniky

Diplomová práce

Systém pro automatickou kontrolu samostatných prací vytvořených v MS Access

Plzeň, 2018 Vojtěch Kinkor

Zadání

1. Seznamte se s formátem souboru MS Access .accdb a možnostmi jeho čtení.
2. Seznamte se validátorem na portálu ZČU.
3. Vytvořte konfigurovatelný systém pro kontrolu samostatných prací vytvořených v MS Access se zaměřením na splnění zadání a plagiarismus.
4. Část systému pro kontrolu splnění zadání adaptujte pro validátor.
5. Celý systém pečlivě otestujte.

**< tato strana bude nahrazena originálním zadáním / kopií ! >**

Prohlášení

Prohlašuji, že jsem diplomovou práci vypracoval samostatně a výhradně s použitím citovaných pramenů.

V Plzni dne <TODO: 0. 0. 2018>

Vojtěch Kinkor

Abstract

**Systém pro automatickou kontrolu samostatných prací vytvořených   
v MS Access**

text

Abstrakt

**System for Automatic Checking of Student Works Created in MS Access**

text

Obsah

[1 Úvod 1](#_Toc510188371)

[2 Databázový software Microsoft Access 2](#_Toc510188372)

[2.1 Základní informace 2](#_Toc510188373)

[2.2 Prvky databáze 2](#_Toc510188374)

[2.2.1 Tabulky 2](#_Toc510188375)

[2.2.2 Vazby mezi tabulkami **Chyba! Záložka není definována.**](#_Toc510188376)

[2.2.3 Uložené dotazy 6](#_Toc510188377)

[2.2.4 Formuláře 6](#_Toc510188378)

[2.2.5 Sestavy 6](#_Toc510188379)

[2.3 Formát MDB a ACCDB 6](#_Toc510188380)

[2.3.1 Metadata 6](#_Toc510188381)

[2.4 Možnosti čtení souborů ACCDB 6](#_Toc510188382)

[2.4.1 ODBC 6](#_Toc510188383)

[2.4.2 Microsoft Office Interop 6](#_Toc510188384)

[2.4.3 MDBTools 6](#_Toc510188385)

[2.4.4 Jackcess 6](#_Toc510188386)

[2.4.5 Další možnosti (JDBC) 6](#_Toc510188387)

[3 Portál ZČU 7](#_Toc510188388)

[3.1 Základní informace 7](#_Toc510188389)

[3.2 Validátor studentských prací 7](#_Toc510188390)

[3.3 Vytvoření nové validační domény 7](#_Toc510188391)

[4 Analýza řešení 8](#_Toc510188392)

[4.1 Požadavky na řešení 8](#_Toc510188393)

[4.2 Případy užití 8](#_Toc510188394)

[4.3 Validace databáze 8](#_Toc510188395)

[4.4 Vyhodnocení plagiarismu 8](#_Toc510188396)

[5 Implementace 9](#_Toc510188397)

[5.1 Použité technologie 9](#_Toc510188398)

[5.2 Struktura aplikace 9](#_Toc510188399)

[5.3 Validace databáze 9](#_Toc510188400)

[5.4 Implementovaná validační pravidla 9](#_Toc510188401)

[5.5 Hledání podobností a detekce plagiarismu 9](#_Toc510188402)

[5.6 Grafické rozhraní 9](#_Toc510188403)

[5.7 Adaptace pro validátor portálu ZČU 9](#_Toc510188404)

[6 Testování 10](#_Toc510188405)

[6.1 Validační pravidla 10](#_Toc510188406)

[6.2 Konzolová aplikace pro validátor portálu ZČU 10](#_Toc510188407)

[6.3 Grafické rozhraní 10](#_Toc510188408)

[7 Závěr 11](#_Toc510188409)

[Reference 12](#_Toc510188410)

[Přílohy 12](#_Toc510188411)

[A Uživatelská příručka 13](#_Toc510188412)

[Spuštění a kompilace nástroje 13](#_Toc510188413)

[Obsluha nástroje 13](#_Toc510188414)

[B Obsah přiloženého média 13](#_Toc510188415)

# Úvod

# Databázový software Microsoft Access

## Základní informace

Microsoft Access je nástroj řadící se mezi takzvané systémy řízení báze dat (SŘBD či DBMS – database management systém). Jedná se o software, který umožňuje práci s relačními databázemi. Je součástí kancelářského balíku Microsoft Office, případně prodáván i samostatně. Pro vytváření a správu databáze nabízí uživatelům přehledné grafické rozhraní.

Aplikace používá pro ukládání dat technologii Microsoft Jet Database Engine, v novějších verzích poté nazývanou Access Database Engine. Jednotlivé databáze jsou typicky uloženy v jediném souboru ve formátu ACCDB, nebo MDB.

## Objekty uložené v databázi

Dále jsou uvedeny různé objekty, které mohou být součástí databáze.

### Tabulky

Jedná se o stěžejní součást každé databáze. Tabulku lze definovat jako strukturovanou kolekci dat. Skládá se ze sloupců a řádků (též záznamů) a v rámci databáze má unikátní název.

Sloupce tabulky

Struktura tabulky je definována pomocí sloupců, které mají specifikovaný název (unikátní v rámci tabulky) a datový typ. Microsoft Access[[1]](#footnote-2) podporuje následující datové typy [[1](#_toc_1)]:

* **Automatické číslo** – typicky používáno jako primární klíč (viz dále), pro každý nový záznam se automaticky nastaví na následující hodnotu posloupnosti, nebo na náhodné číslo (dle nastavení).
* **Číslo** – rozsah a typ (celočíselné/s desetinnou čárkou) lze zvolit ve vlastnostech sloupce.
* **Krátký text** (dříve Text) – text do délky 255 znaků.
* **Dlouhý text** (dříve Memo) – text do velikosti 1 GB.
* **Datum a čas**.
* **Měna** – specializovaný případ číselného datového typu s fixní desetinnou čárkou (uchovává 4 desetinná místa).
* **Ano/ne** – uchovává hodnotu -1 (Ano) nebo 0 (Ne); v rámci Microsoft Accessu zobrazeno jako zaškrtávací pole (*checkbox*).
* **Hypertextový odkaz.**
* **Objekt OLE** – umožňuje vložit speciální objekty, například obrázek, jiný dokument, či odkaz na soubor.
* **Příloha** – umožňuje vložit libovolný soubor jako součást záznamu. Jedná se o univerzálnější možnost k předchozímu.
* **Počítané** – automatické vložení hodnoty vypočítané na základě zadaného vzorce.

Každému sloupci lze dále nastavit různé vlastnosti dle vybraného datového typu – typicky se jedná o ověřovací pravidla (validace vstupu od uživatele ještě před přidáním záznamu do databáze), výchozí hodnotu a dále nastavení zobrazení v tabulce (formátování, zarovnání, titulek po najetí myší, atp.).

Primární klíč

Tabulka může mít primární klíč – typicky se jedná o sloupec, jehož hodnoty jsou unikátní a vždy zadané (tzv. not null). V případě, že vytvoříme primární klíč pomocí více sloupců, nazýváme jej složeným primárním klíčem.

Primární klíč slouží pro odkázání na jeden konkrétní záznam v tabulce, čehož se využívá při vytváření dotazů nebo tvoření relací mezi tabulkami. Pro vytváření primárních klíčů se obvykle využívá datový typ Automatické číslo, který každému záznamu přiřadí unikátní celé číslo. Často bývá takový sloupec pojmenován „ID“ (*Identification*).

Relace mezi tabulkami a cizí klíče

V případě, že chceme propojit více tabulek mezi sebou, využijeme tzv. relačních vazeb. Jedná se o situaci, kdy záznam v tabulce odkazuje („má referenci“) na jeden konkrétní záznam z druhé tabulky.

Rozlišují se tři druhy relačních vazeb.

* **Relace 1:1** – jednomu záznamu v tabulce A odpovídá žádný, či právě jeden záznam v tabulce B. Pro referencování se využívají pouze primární klíče obou tabulek (mají tedy v obou tabulkách shodnou hodnotu).



Obrázek 1 – model relace 1:1

* **Relace 1:N** – k více záznamům v tabulce A lze přiřadit jeden záznam z tabulky B. To lze zajistit přidáním tzv. **cizího klíče** do tabulky A – sloupce, který bude obsahovat pouze hodnoty primárního klíče z tabulky B (příp. skupiny sloupců, pokud se jedná o složený primární klíč). Jedná se o nejčastěji využívanou vazbu.



Obrázek 2 – model relace 1:N

* **Relace M:N** – k M záznamům v tabulce A lze přiřadit N záznamů z tabulky B. Relace se realizuje pomocí spojové tabulky (též mezitabulky) a dvojicí relací 1:N. Spojová tabulka obvykle obsahuje pouze sloupce cizích klíčů.



Obrázek 3 – model relace M:N

Relace mezi tabulkami mohou zajišťovat **referenčním integritu**. Cílem je zabránit odkazování na neexistující záznam (a rovněž tedy vzniku osiřelých záznamů, na které byly všechny reference zrušeny).

Integritní pravidlo může zajistit kaskádovou aktualizaci polí – pokud se změní hodnota primárního klíče, změní se automaticky hodnota u všech záznamů, které na záznam odkazují. Dále může zajistit kaskádové odstranění souvisejících záznamů – v případě smazání záznamu budou smazány i všechny další, které na právě tento záznam odkazovaly.

### Dotazy

Dotazy slouží k získávání, přidávání, mazání či upravování dat v databázi. Microsoft Access umožňuje ukládání dotazů do databáze – lze tedy vytvořit dotazy pro usnadnění následné práce s daty.

Dotazy mohou mít parametry, které lze využít např. pro filtrování záznamů v rámci tabulky nebo nové hodnoty při vkládání/upravování záznamů. Uživatel je pak při spuštění dotazu vyzván k zadání konkrétních hodnot parametrů.

Podporovány jsou následující druhy dotazů:

* **Výběrové (SELECT)** – jedná se o dotaz, jehož výsledkem je množina vybraných záznamů. Struktura je dána dotazem – jednotlivé sloupce mohou pocházet z různých tabulek, či být spočítané „za běhu“. Obecně lze považovat výběrový dotaz za analogii k databázovým pohledům.
* **Vytvářecí (MAKE TABLE)** – pracuje na stejném principu jako výběrový, výsledek dotazu však není ihned zobrazen uživateli, ale uložen do nové tabulky.
* **Přidávací (INSERT)** – slouží pro vkládání nových záznamů do existujících tabulek.
* **Aktualizační (UPDATE)** – umožňuje úpravu hodnot již existujících záznamů v tabulkách.
* **Křížový (CROSSTAB)** – výsledkem dotazu je tzv. kontingenční tabulka zobrazující data v kompaktní podobě. Typicky se používá například pro sumarizaci hodnot, nalezení průměrů, maximálních hodnot, atp.

### Formuláře

Formuláře poskytují přívětivé rozhraní pro vkládání či editaci záznamů v tabulkách. Grafické rozhraní je plně konfigurovatelné a umožňuje tedy jednotlivá pole záznamů různě seskupovat, přidat popisky, či některá úplně skrýt. Formuláře jsou v databázi opět uloženy pod unikátním názvem.

Access umožňují vytvořit formuláře různých druhů:

* Formuláře pro editaci jednotlivých záznamů (dále označované jako standardní).
* Navigační formuláře, které poskytují možnost přepínání mezi různými formuláři a umožňují tak vytvořit komplexní rozhraní pro správu celé databáze.
* Formuláře zobrazující více položek (záznamů) najednou.
* Datové listy, které vypadají podobně jako zobrazení tabulky (tedy tabulka, kde každý řádek odpovídá jednomu záznamu), ale zachovávají možnost upravovat zobrazená pole.
* Rozdělené formuláře, které jsou kombinaci standardních formulářů v jedné části a datového listu v druhé části obrazovky.
* Modální dialogová okna, která mají stejné možnosti jako standardní formuláře, ale zobrazují se v samostatném okně a jsou

### Sestavy

Sestavy slouží pro vytváření výpisů dat z databáze v přívětivé podobě, zobrazující typicky více záznamů na jedné straně, na rozdíl od formulářů ale neumožňuje editaci dat. Často se využívá pro následné vytisknutí. Při návrhu se definuje záhlaví a zápatí stránek a rozložení prvků pro každý záznam („řádek“ sestavy).

### Skryté systémové tabulky

Jak se jmenují, jak k nim lze přistoupit, co obsahují/lze z nich zjistit.

## Metadata databázových souborů

Datum a čas vytvoření/editace databázového souboru, tabulek, jméno autora/organizace.

## Formáty ACCDB a MDB

Nativním formátem pro ukládání databází je od verze 2007 ACCDB, v předchozích verzích byl hlavním formátem MDB. Oba jsou založeny na technologii Jet (u formátu ACCDB také označované jako Access Database Engine) a jsou si tedy technologicky podobné. Z uživatelského hlediska jsou rozdíly zejména v různých možnostech zabezpečení dat.

## Možnosti čtení souborů ACCDB

Jedná se o proprietární binární formát vyvíjený společností Microsoft bez dostupné specifikace. Jediným oficiálním nástrojem pro správu je právě Microsoft Access, pro přístup k datům je dostupné ODBC.

To velmi omezuje možnosti programového přístupu k databázím – pokud bychom vzali v potaz pouze oficiální nástroje, jsme limitováni na systémy s nainstalovanou aplikací Microsoft Access (a tím pádem i operačním systémem). V současné době jsou však dostupné i nástroje vzniklé na základě reverzního inženýrství formátů MDB/ACCDB bez závislosti na programovém vybavení počítače.

Dále jsou zmíněny všechny možnosti čtení souborů ACCDB včetně výhod a nevýhod, jaké přináší.

### ODBC

ODBC (Open Database Connectivity) je standardizované API pro přístup k datům uloženým v databázích. Připojení ke konkrétním databázím je zajištěno speciálními ovladači, které lze do systému doinstalovat. Pro komunikaci skrze ODBC se typicky využívá jazyk SQL (Standard Query Language), ovladač poté zajistí vykonání příkazu nad konkrétní databází.

Pro přístup k ACCDB databázím v rámci OS Microsoft Windows se využívají ovladače Access Database Engine nainstalované spolu s aplikací Microsoft Access, případně ze samostatného distribučního balíku. Pro další platformy existují komerční Access ODBC ovladače. Vzniká zde tedy závislost na dostupnosti ovladače, přičemž v určitých případech může být problém jej do systému doplnit.

Zásadní nevýhodou přístupu k datům přes ODBC API jsou omezení vyplývající z univerzálnosti metody. Jednoduše lze pracovat pouze s daty v tabulkách a není možné přímo přistupovat k dalším uloženým objektům. Jedinou možnost je využít skryté systémové tabulky, pomocí kterých lze zjistit alespoň existenci objektů.

Novější obdobnou technologií je OLE DB (*Object Linking and Embedding, Database*), vyvinuté firmou Microsoft původně jako nástupce ODBC. Z hlediska způsobu použití a nabízených funkcí pro čtení souboru ACCDB jsou technologie shodné.

### Microsoft Office Interop

Aplikace z balíku Microsoft Office lze programově ovládat pomocí technik obecně označovaných jako *interoperability* (zkráceně *interop*). Typicky se využívají proprietární technologie COM (Common Object Model) a OLE (Object Linking and Embedding) vyvinuté firmou Microsoft. Dále jsou poskytovány *Primary Interop Assemblies* – knihovny určené pro použití na platformě .NET (tedy v tzv. řízeném kódu) obalující COM volání do objektového rozhraní. V současné době poskytují nejjednodušší možnost pro programové ovládání aplikací Microsoft Office (mj. se využívají i pro psaní doplňku, *plug-inů*, pro jednotlivé Office aplikace).

Tato technika oproti ODBC umožňuje kompletní správu databáze vč. všech dostupných objektů a bez nutnosti analyzovat obsah systémových tabulek. Avšak zůstává zde nutnost spouštět kód v systému, kde je nainstalovaný Microsoft Access. Jedná se rovněž o poměrně pomalý přístup, jelikož interop kód de-facto jen ovládá aplikaci Microsoft Access spuštěnou v systému na pozadí.

### MDB Tools

Jedná se o open-source sadu nástrojů pro práci se soubory Microsoft Access, respektive Jet databázemi ve formátu MDB, jejíž vývoj započal již v roce 2000[[2]](#footnote-3). Vzhledem k uzavřenosti formátu vznikla většina nástrojů technikami reverzního inženýrství, a jak vyplývá z odezvy uživatelů, objevují se případy, kdy nástroje nefungují zcela správně[[3]](#footnote-4).

Nástroje jsou napsány v jazyce C a mají konzolové rozhraní, existuje ale i několik grafických nadstaveb pro prohlížení Access souborů. Součástí projektu je i dokument popisující strukturu a klíčové části Jet databází. V posledních letech probíhá vývoj pomalým tempem a podpora novějších verzí Access databází včetně formátu ACCDB není zaručena. Hlavní výhodou je nezávislost na konkrétní platformě a externích knihovnách.

### Jackcess

Jackcess je Java knihovna poskytující čisté objektové rozhraní pro práci s Microsoft Access databázemi. Její vývoj započal v roce 2005 v rámci open-source projektu OpenHMS zaštítěného firmou Health Market Science, Inc.

Kromě čtení dat z tabulek umožňuje i základní editaci struktury databáze, výpis všech relací mezi tabulkami a výpis uložených dotazů. Díky přístupu ke skrytým systémovým tabulkám lze vyhledat i uložené formuláře a sestavy (reálně však lze zjistit pouze jejich existenci). Knihovna podporuje Access databáze ve verzích 2000 až 2016 (ve formátu MDB i ACCDB) a ve verzi 97 v režimu pro čtení. Knihovna neobsahuje rozhraní pro spouštění SQL dotazů, neumožňuje tedy ani vyhodnocení uložených dotazů.

Zásadního výhodou pro potřeby této práce je přenositelnost knihovny (nezávislost na platformě) a aktivní vývoj – tím pádem i podpora nejnovějších verzí Access databází. Vzhledem k distribuci v podobě samostatné Java knihovny je rovněž její použití ve vlastní aplikaci jednoduché.

Pro knihovnu existuje rozšíření nazvané Jackcess Encrypt umožňující správu databází opatřených heslem. Podporuje „některé formy šifer aplikací Microsoft Access a Microsoft Money“[[4]](#footnote-5).

### Další dostupné metody

MDB Tools Java

V roce 2004 začala *portace* nástrojů MDB Tools pro jazyk Java[[5]](#footnote-6), vývoj však již po roce ustal. Následně vzniklo několik *forků* (projektů založených na kódu původního projektu), nejaktuálnější z nich lze nalézt pod názvem „ome-mdbtools“*[[6]](#footnote-7)*. Vývoj těchto projektů je ale spíše pomalý – poslední větší aktualizace proběhla v roce 2016. Oproti knihovně Jackcess nabízí méně možností a použité je značně komplikované[[7]](#footnote-8). Překážkou je chybějící dokumentace jak použití nástrojů, tak samotného programového kódu.

JDBC

Java Database Connectivity je API pro přístup k relačním databázím a tedy obdobou technologie ODBC pro programovací jazyk Java. API je standardní součástí platformy Java SE. Připojení ke konkrétní databázi je opět zajištěno speciálními ovladači určenými pro konkrétní typ databází.

Microsoft neposkytuje

Dále existují ovladače umožňující použití ODBC ovladačů, tzv. *JDBC-ODBC bridge*. V rámci platformy Java SE byl standardní součástí do verze 1.7; do novějších verzí jej lze překopírovat (jedná se však o neoficiální postup bez záruky na funkčnost) nebo nahradit komerčními produkty.

# Portál ZČU

## Základní informace

## Validátor studentských prací

## Vytvoření nové validační domény

# Analýza kontroly prací

## Požadavky na řešení

## Případy užití

## Validace databáze

## Vyhodnocení plagiarismu

# Implementace systému pro automatickou kontrolu prací

## Použité technologie

## Struktura aplikace

## Validace databáze

## Implementovaná validační pravidla

## Hledání podobností a detekce plagiarismu

## Grafické rozhraní

## Adaptace pro validátor portálu ZČU

# Testování vytvořeného systému

## Validační pravidla

## Konzolová aplikace pro validátor portálu ZČU

## Grafické rozhraní

# Závěr

# Reference

[1] Data types for Access desktop databases. *Microsoft Office help and training - Office Support.* [Online] [Citace: 22. 3. 2018]. Dostupné z: <https://support.office.com/en-us/article/data-types-for-access-desktop-databases-df2b83ba-cef6-436d-b679-3418f622e482>

[2] Create and use an index to improve performance. *Microsoft Office help and training - Office Support.* [Online] [Citace: 14. 3. 2018]. Dostupné z: <https://support.office.com/en-us/article/create-and-use-an-index-to-improve-performance-0a8e2aa6-735c-4c3a-9dda-38c6c4f1a0ce>

[3] Which Access file format should I use? *Microsoft Office help and training - Office Support.* [Online] [Citace: 20. 3. 2018]. Dostupné z: <https://support.office.com/en-us/article/which-access-file-format-should-i-use-012d9ab3-d14c-479e-b617-be66f9070b41>

[4] Microsoft Access ACCDB File Format Family. *Digital Preservation at the Library of Congress.* [Online] [Citace: 20. 3. 2018]. Dostupné z: <https://www.loc.gov/preservation/digital/formats/fdd/fdd000462.shtml>

[5] ADAMSKI, Joseph J.; FINNEGAN, Kathy T. ; SCOLLARD, Sharon. *New perspectives on Microsoft Access 2013: comprehensive.* Stamford, CT: Cengage Learning, 2014. ISBN 978-1-285-09920-0.

# Přílohy

## A Uživatelská příručka

### Spuštění a kompilace nástroje

### Obsluha nástroje

Typický postup práce s nástrojem je následující:

1. ...

## B Obsah přiloženého média

Součástí práce je přiložené paměťové médium (DVD) obsahující tyto adresáře a soubory:

* Project/ – adresář obsahující projekt vytvořené aplikace,
* Kinkor\_A16N0040P\_DP.pdf – text této práce ve formátu PDF,
* readme.txt – textový soubor obsahující popis struktury DVD.

Obsah přiloženého média (včetně aktuální verze nástroje) je možné najít též v repozitáři projektu v rámci služby GitHub na adrese:

<https://github.com/ikeblaster/access-validator/>

1. Aktuálně ve verzi 2016. [↑](#footnote-ref-2)
2. Viz repozitář projektu: https://github.com/brianb/mdbtools [↑](#footnote-ref-3)
3. Viz *issue tracker* projektu: <https://github.com/brianb/mdbtools/issues> [↑](#footnote-ref-4)
4. Viz WWW: <http://jackcessencrypt.sourceforge.net/index.html> [↑](#footnote-ref-5)
5. Viz WWW: <https://sourceforge.net/p/mdbtools/discussion/6688/thread/a543445a/> [↑](#footnote-ref-6)
6. Viz repozitář: <https://github.com/ome/ome-mdbtools> [↑](#footnote-ref-7)
7. Viz ukázkový kód: [https://github.com/ome/ome-mdbtools/blob/master/src/main/ ↩ java/mdbtools/tests/ColumnTest.java](https://github.com/ome/ome-mdbtools/blob/master/src/main/%20↩%20java/mdbtools/tests/ColumnTest.java) [↑](#footnote-ref-8)