

Vysoká škola ekonomická v Praze
Fakulta informatiky a statistiky



Návrh báze znalostí skautských programů

AUTOREFERÁT BAKALÁŘSKÉ PRÁCE

Studijní program: Aplikovaná informatika

Autor: Šimon Trousil
Vedoucí práce: prof. Ing. Václav Řepa, CSc.

Praha, květen 2024

Cíl práce

Cílem mé práce bylo navrhnout bázi znalostí skautských programů. S tím, že báze je specifikována jako systém umožňující co nejsnazší sdílení (zapisování) záznamů o programech připravených v rámci činnosti našeho oddílu a zároveň nabízející možnost správného a rychlého filtrování zobrazení (čtení) uložených záznamů, podle definovaných částí jejich obsahu. Tento systém by zároveň, vzhledem k neziskové a dobrovolné povaze skautingu, nevyžadoval lidské ani finanční zdroje na svůj provoz či údržbu konzistence uložených dat.

Obsahem záznamů jsou pak chápány konkrétní části a parametry jednotlivých zaznamenaných programů (např. "Pro kolik dětí je hra vhodná?", "Kolik vedoucích je na hru potřeba?", ...). Proto bylo třeba definovat, co jsou to 'skautské programy' a jaké údaje, s nimi asociované, je vhodné v navrhované bázi ukládat právě jako obsah jednotlivých záznamů, který pak umožní samotné jejich filtrování při prohledávání (čtení) v bázi uloženého obsahu. Pro vyřešení tohoto dílčího problému byl v práci stanoven 2. dílčí cíl, který si klade za úkol vybrání výšece reality, která by v bázi měla být zaznamenávána.

Aby však, na základě získané definice domény skautských programů, bylo možné sestavit konkrétní strukturu navrhované báze, bylo nutné mít zvolený konkrétní softwarový nástroj, který by tvořil její infrastrukturu. A protože požadavky na bázi nabízené možnosti čtení a zapisování i vyžadované zdroje, stanovené hlavním cílem, jsou všechny záležitostí primárně zvolené infrastruktury, byl v práci stanoven 1. dílčí cíl, který si klade za úkol vybrání právě takové infrastruktury, která by nejlépe splňovala stanovené požadavky.

S vybranou infrastrukturou i definovanou doménou, jenž má být bází reprezentována, je pak posledním krokem, potřebným k dosažení hlavního cíle práce, sestavení schématu struktury pro údaje, které byly vybrány k ukládání ve vybraném databázovém nástroji. Proto je v práci stanoven 3. a zároveň poslední dílčí cíl, který si klade za úkol tuto strukturu získat.

Použité metody

1. dílčí cíl Při výběru nejvhodnější infrastruktury jsou v práci posuzovány čtyři softwarové nástroje, každý využívající odlišnou datovou strukturu pro uložení záznamů.

Hodnocení nástrojů z hlediska nabízených možností zapisování obsahu je provedeno pouze pomocí jednoduchého kritéria, které přijímá pouze nástroje umožňující zapisování bez nutnosti znalosti jiného než přirozeného jazyka, v němž jsou zápisy realizovány.

Kritérium tak efektivně zajišťuje, že navržená báze bude umožňovat uživatelům zapisovat její obsah pomocí rozhraní nepříliš odlišného od Dokumentů Google - relativně k posuzovaným nástrojům databázového typu.

Z hlediska nabízených možností čtení uloženého obsahu je v práci využito hodnocení posuzovaných nástrojů pomocí tří kritérií a výběrem jediného nástroje, který je shledán z tohoto hlediska nejlepším. Základním kritériem byl požadavek na schopnost posuzovaného nástroje zobrazovat záznamy odpovídající zadanému dotazu, bez chyb prvního i druhého typu.

Druhým kritériem hlediska možností čtení byl odhadovaný průměrný počet znaků vyžadovaný k napsání dotazu danému nástroji o zobrazení jen určité podmnožiny z uloženého obsahu.

Posledním kritériem tohoto hlediska pak bylo, zda nástroj v základu poskytuje grafické webové rozhraní umožňující s ním interagovat. Všechna kritéria tak mají za úkol pomoci určit, který nástroj nabízí nejsnadnější, zároveň bezchybné filtrování zobrazení uloženého obsahu.

Pro nástroje vybrané hodnocením, podle jednoho i druhého hlediska, je následně provedena analýza obsahu literatury publikované o daném nástroji. Účelem této analýzy bylo podrobnější charakterizování datových struktur v nichž nástroje ukládají zapsaný obsah, spolu s možnostmi programové komunikace s danými nástroji.

Posledním krokem při výběru infrastruktury je pak již jen identifikování konkrétního softwarového řešení pro použití v návrhu a ověření realizovatelnosti tohoto řešení pomocí implementace jeho klíčových funkcionalit.

2. dílčí cíl Co se obsahu navrženého k uložení týče, k jeho určení jsou v práci využity tři centralizovaná úložiště záznamů o skautských programech označovaná dále jako existující báze.

Pro existující báze, jsou za prvé identifikovány kvality zaznamenaného obsahu a za druhé vlastnosti báze, které jsou problematické z hlediska jejího využití jako interního nástroje oddílu. Tím je tak zároveň argumentována volba danýchází jako výchozího zdroje pro definování výšece reality navržené k zaznamenávání.

Obsah těchto existujícíchází byl následně analyzován s primárním záměrem nalezení pojmů, asociovaných danou existujícíází s uloženými záznamy, které vyznačí výšece reality jenž bude dále modelována. Konkrétní způsoby získání hledaných pojmů se nicméně mírně lišily v závislosti na míře strukturovanosti jednotlivýchází. Pokud báze nabízela jasný popis struktury pro záznamy o programech byla identifikována konkrétní část té báze, ze které byly pojmy vybrány. V případě, že jasná struktura chyběla, byla provedena stručná rešerše daného obsahu a pojmy byly následně získány z ní.

Pro dokončení definice vybrané výšece reality je v práci využito diagramu tříd podle standardu UML. Konkrétně byla nejprve sestavena 'základní definice', která modeluje vztahy mezi množinou pojmů vzniklou sjednocením jednoho, až dvou pojmů, z každé z existujícíchází, které nejvýstižněji reprezentují, v ní uložené záznamy. Následně byly všechny, z analýzy obsahuází, získané pojmy roztříděny, s pomocí vlastních doménových znalostí, podle jejich asociovatelnosti k jednotlivým třídám v základní definici.

Dále byly modelovány vztahy mezi pojmy v rámci jednotlivých skupin přiřazeným třídám v základní definici, tím bylo docíleno minimálního počtu tříd nutných asociovat se základními třídami a navíc přesnějšího zobrazení reality. Na závěr konceptuálního modelování pak byly modelovány vhodné asociace mezi strukturami z pojmů získaných analýzou existujícíchází a jednotlivými třídami v základní definici. Tím tak byla sestavena 'rozšířená definice' domény skautských programů.

Popsanou metodikou získaný diagram tříd je nakonec porovnán s modelem tvořeným rovněž pro navrhovanou bázi, nicméně primárně s využitím mých vlastních doménových znalostí, nikoliv systematickým zpracováním existujících zdrojů.

3. dílčí cíl Aby byla podpořena šance na navržení takové struktury, která by byla v souladu s nejlepšími praktikami identifikovanými pro zvolený databázový nástroj, byla nejprve provedena rešerše publikované literatury k danému nástroji s cílem identifikovat tyto praktiky.

Samotný návrh struktury báze znalostí skautských programů byl realizován transformací diagramu tříd tvořícího 'rozšířenou definici' modelované domény s ohledem na dříve identifikované nejlepší praktiky. Pro zakreslení navrženého schématu byla využita stejná notace, jakou běžně využívá i oficiální dokumentace zvolené databáze.

Správnost navrženého schématu byla verifikována pomocí několika kompetenčních otázek, které vycházejí z podmnožiny pojmů z množiny všech získaných pojmů, podle kterých dovedly existující báze filtrovat zobrazované záznamy. Kompetenční otázky byly následně přeloženy do jazyka používaného databází a byla pro ně provedena vizuální kontrola porovnáním oproti navrženému schématu, zda jsou v něm obsaženy všechny údaje nezbytné pro zodpovězení daného dotazu.

Dosažené výsledky

Výsledky jsou v práci prezentovány ve dvou, na sebe navazujících částech, rozdělených podle toho, zda se jedná o mezivýsledky vytvořené jako prostředek pro dosažení stanovených dílčích cílů nebo o finální výsledky stanovených dílčích cílů a ověření úspěšnosti jejich dosažení.

Teoretická část Mezivýsledky využití k dosažení dílčích cílů práce jsou souhrnně uvedeny právě v teoretické části.

Konkrétně pro 1. dílčí cíl se jedná o výsledky z hodnocení metodikou vybraných softwarových nástrojů z hlediska jimi nabízených možností čtení a zapisování uloženého obsahu. Následují výsledky z analýzy nástrojů přijatých v rámci, buď hlediska čtení, nebo hlediska zapisování obsahu. Výstupem této analýzy byla detailněji popsaná datová struktura a možnosti integrovatelnosti, hodnocením přijatých, nástrojů.

Z výsledků získaných v rámci 2. dílčího cíle jsou v této části prezentovány pouze ty z hodnocení vybraných existujícíchází. A pokud daná existující báze neměla jasně stanovenou strukturu pro záznamy, pak i rešerše obsahu dané báze. Samotné pojmy, asociované se záznamy o programech analyzovanýmiázemi, jsou představeny až v praktické části. Jelikož během sestavování definice modelované domény jsou stejné pojmy detailně popisovány, není proto nutné, získané pojmy popisovat v obou částech.

V rámci posledního dílčího cíle jsou jediným mezivýsledkem popsané nejlepší praktiky pro modelování dat ve zvoleném databázovém nástroji získané rešerší publikované literatury o konkrétním nástroji.

Praktická část Na úvod této části jsou interpretovány výsledky z hodnocení a analýzy posuzovaných nástrojů. K interpretaci bylo využito jednoduchých modelů, hodnocením přijatých nástrojů, reprezentujících integrovatelnost (možnosti programového přístupu) jednotlivých z nich. Ze zjištěných informací byla následně vyvozena konkrétní tvrzení, která musí být pravdivá, má-li navržená infrastruktura splňovat všechny požadavky hlavního cíle práce. Posledním výsledkem získaným v rámci 1. dílčího cíle, je tak už jen implementace konkrétních funkcí, identifikovaných jako nutnost, pokud má návrh mít šanci naplnit práci stanovený cíl.

Výsledkem 2. dílčího cíle je primárně diagram tříd, jak základní (1-2 pojmy z každého analyzovaného zdroje), tak i rozšířený (základní + ostatní získané pojmy z modelované domény). Tyto představují vybranou výseč reality navrženou k zaznamenávání. Diagramy jsou rovněž doplněny o vysvětlení jednotlivých, v modelu zachycených, pojmů i jejich vztahů. Úspěšnost dosažení tohoto dílčího cíle je pak vyjádřena porovnáním modelovaného obsahu s dříve vytvořeným diagramem tříd, jenž byl sestaven primárně pomocí vlastních doménových znalostí.

Závěrečnými výsledky 3. dílčího cíle je prvně návrh schématu pro vybranou bázi. Pro názorné vysvětlení postupu transformace diagramu tříd podle nejlepších praktik pro modelování dat ve vybrané databázi jsou prezentovány výstupy několika iterací transformování schématu. Bylo při tom započato iterací s minimálními změnami oproti původnímu diagramu a pokračováno několika dalšími, které dále postupně optimalizují a upřesňují nakreslené schéma. Druhým výsledkem určeným k ověření správnosti navrženého schématu jsou sestavené ukázkové dotazy pro databázi, které vycházejí z pojmů s jejichž pomocí umožňují filtrování svého obsahu existující báze.

Vlastní přínos autora

Prvním přínosem mé práce je výběr softwarového řešení, jenž splňuje podmínky stanovené s ohledem na snadné a nenáročné využívání navrženého systému. Pro toto softwarové řešení jsem také ověřil jeho realizovatelnost pomocí implementace jeho klíčových funkcionalit.

Druhým přínosem je sestavení základní definice domény skautských programů s využitím diagramu tříd podle UML. Tuto definici jsem poté rozšířil o pojmy, které jsou vhodné pro asociování se skautskými programy v případě vytváření záznamů o jednotlivých instancích základních tříd z této domény.

Třetím přínosem je navržení schématu pro vybraný databázový nástroj. Toto schéma vychází ze sestaveného diagramu tříd, představujícího rozšířenou definici modelované domény. Navíc jsem ověřil kompetence navrženého schématu na zodpovídání otázek, jako dovedou zodpovídat existující báze.

A konečně, mým přínosem je i relativně podrobné popsání postupu, který jsem v rámci práce využil pro získání výsledků.