

Divize environmentální informatiky a modelování
Institut biostatistiky a analýz
Masarykova univerzita

Sborník příspěvků

3. letní škola aplikované informatiky

Editoři

Jiří Hřebíček, Jaroslav Ráček

Bedřichov, 25.-27. srpna 2006

© Masarykova univerzita, 2006

ISBN XXXXXXXXXXXXXXXXXX

Obsah

<i>Jiří Hřebíček, Jaroslav Ráček</i> Environmentální informatika a její role v aplikované informatice.....	1
<i>Jana Andrášková</i> Návrh a pojetí poradenského systému podle metodiky WASP	7
<i>Tomáš Foltýnek</i> Evaluace položek didaktických testů pomocí koeficientu korektnosti.....	15
<i>Michal Hejč</i> Metody hodnocení kvality dat a softwarové možnosti jejího zlepšení.....	22
<i>Jiří Hřebíček, Danka Némethová, Jaroslav Ráček</i> Informační potřeby státní správy, nezávislých organizací, podniků, škol a veřejnosti v oblasti životního prostředí.....	28
<i>Karel Kiswa, Matěj Štefaník</i> Výpočetní knihovna pro systém hodnocení ekologického stavu povrchových vod.....	35
<i>Lukáš Kokrment</i> Ontologie v bioinformatice	39
<i>Miroslav Kubásek</i> Optimalizace SOA pomocí cache.....	46
<i>Jan Ministr</i> Metodický rámec stanovení kvality informačního systému pomocí vrstveného modelu.....	53
<i>Petra Misařová</i> Implementace EMS v národních parcích USA	59
<i>Jan Pavlovič</i> Servisně-orientovaná architektura systému na výběr dekontaminační strategie	65
<i>Miroslav Petráš, Jaroslav Ráček</i> Komunikační systém NOP On-line	73
<i>Tomáš Pitner, Pavel Drášil</i> Moderní metodiky, technologie a nástroje pro podporu výuky (E-learning 2.0)	79
<i>Jaroslav Ráček, Jiří Hřebíček</i> Informační systém environmentálního reportingu	85
<i>Jana Sedláčková, František Huňka</i> Odhady nákladů workflow projektů metodou FPA.....	90
<i>Jana Soukopová</i> Analýza a návrh matematického modelu na podporu rozhodování ve veřejných projektech	95
<i>Michal Škrdla</i> Matematický model růstu populací mšic a afidofágů na kukuřičném poli.....	101
<i>Matěj Štefaník, Karel Kiswa</i> Komunikační rámec pro výměnu environmentálních informací.....	106
<i>Vladimír Žák</i> Rychlé grafické algoritmy a jejich využití se systémem Maple.....	117

Environmentální informatika a její role v aplikované informatice

Jiří Hřebíček, Jaroslav Ráček

Masarykova univerzita, Institut biostatistiky a analýz,
Kamenice 126/3, 625 00 Brno
hrebicek@iba.muni.cz, racek@iba.muni.cz

Abstrakt

Informační a komunikační technologie jsou nástrojem pro tvorbu a distribuci informací, z nichž některé se mohou týkat životního prostředí v celosvětovém měřítku. Letní školy aplikované informatiky kombinují témata životního prostředí s rozšiřováním potenciálu informatiky pro zpracování dat o životním prostředí a k zpřístupnění poznatků výzkumu prostřednictvím Internetu. Tento článek se snaží o objasnění, jak výměna výsledků výzkumu v oblasti environmentální informatiky na bázi odborných konferencí, veřejně přístupných informačních zdrojů a nově vyvinutých nástrojů environmentální informatiky podněcuje integraci znalostí a podporu rozhodování v životním prostředí. Shrňme v něm rovněž témata, kterými se zabývala 3. letní škola aplikované informatiky, která se konala ve dnech 25. až 27. srpna 2006 v Bedřichově.

Abstract

Information and Communication Technology is a tool for generating and distributing information, some of which can be environmental in nature and worldwide in scope. Summer schools of applied informatics combine environmental topics with the dissemination potential of informatics to implement access to environmental data and research findings as information offerings on the Internet. This paper seeks to clarify how the conference-based exchange of research results, information sources and newly developed applications stimulates knowledge integration and competence building. We summarize here also topics of the 3th Summer School of Applied Informatics, which was held in Bedřichov 25. - 27 August 2006.

Klíčová slova

Environmentální informatika, environmentální informace, environmentální informační systém, letní škola, aplikovaná informatika.

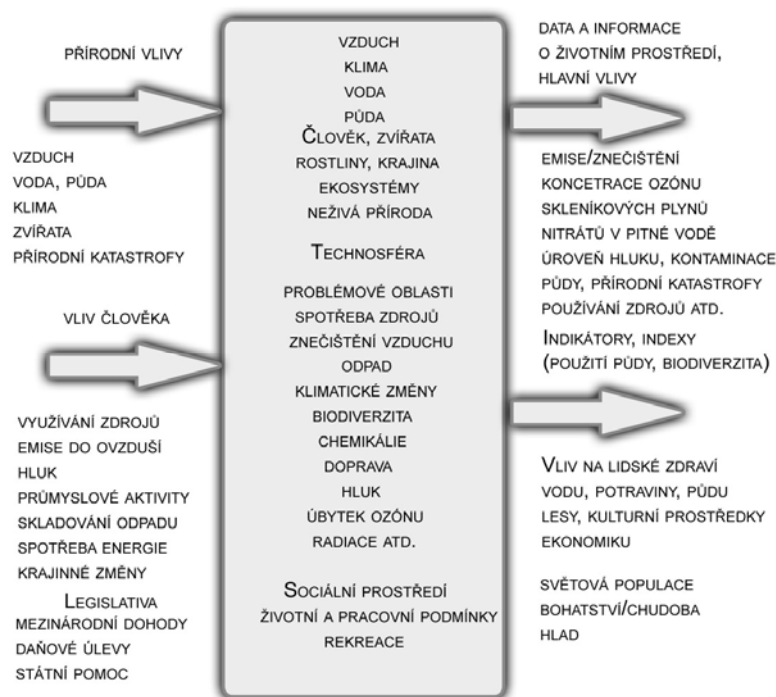
Keywords

Environmental informatics, environmental information, environmental information systems, summer school, applied informatics.

1 Výzvy životního prostředí informační společnosti

Monitoring a sběr dat v oblasti životního prostředí přináší neustále rostoucí množství informací. Když se Valné shromáždění organizace spojených národů rozhodlo uspořádat Světový summit Rio+10 v Johannesburgu (2002) o trvale udržitelném rozvoji, bylo jasné, že mezi jejími účastníky bude nutná nová úroveň dialogu. Mimo proslovy a fráze byli účastníci summitu přinuceni v interaktivním dialogu čelit potřebám a argumentům ostatních činitelů, jako bojovníkům za lidská práva, zástupcům komerční sféry a průmyslu.

Jednotlivé druhy v současnosti používaných informačních toků zahrnují: *osobní komunikaci; veřejné diskuse; debaty a úvahy; publikování; komunikace masmédií; vzdělávání a digitální komunikaci prostřednictvím Internetu.* Evropská agentura životního prostředí (EEA) vypracoval ve zprávě „Nový model environmentální komunikace pro Evropu od spotřeby až po využití informací“ (1998) návrh na změnu komunikace o životním prostředí takovým způsobem, že informace se stávají zdrojem pro podporování znalostí o životním prostředí a schopnosti rozhodování vedoucí k udržitelnosti. Jedním ze struktur. závěrů zprávy je, že by mělo být zvýšeno úsilí směrem k rozvoji vhodných komunikačních technologií a struktur.



Obr. 1: Systém vstupů/výstupů životního prostředí: vlivy a účinky
Zdroj: (Pillmann 2005), upraveno

Na obr. 1 je znázorněn systémový diagram vstupů a výstupů životního prostředí. Lidské vlivy a přírodní účinky způsobují změny v životním prostředí v různých sférách, které mohou být pozorovány a agregovány do podoby informací a indikátorů. Jako indikátor životních podmínek je zde uvedeno „Sociální prostředí“, kvůli svému důležitému vlivu na životní prostředí ve smyslu chudoby, hladu, vyloučení ze společnosti, stárnutí a nezaměstnanosti, atd.

Komunikace mezi vědci zabývajícími se problematikou životního prostředí byla usnadněna schopností získávat informace z Internetu a většina vědců se na něj již napojila on-line. Environmentální informatika sloučila telekomunikace a zpracování informací o životním prostředí s vědeckým výzkumem a jejich využitím ve veřejné správě a obchodu. Během 4. až 6. rámcových programů Evropská komise podpořila vývoj informačních a komunikačních technologií (ICT) v oblasti, které byly později nazvány „Telematické“ aplikace. Činnost EEA se zaměřila na distribuci vědeckých a aplikačně orientovaných informací o životním prostředí implementovaných pomocí webových technologií v EIONETu (Evropská monitorovací a pozorovací síť informací o životním prostředí). Informační zdroje EEA (<http://dataservice.eea.eu.int/dataservice/>) nebo systém Envirowindows (<http://ewindows.eu.org>) jsou příklady veřejného přístupu k informacím o životním prostředí.

Vědní obor v aplikované informatice zabývající se výměnou rozsáhlých souborů dat a informací o životním prostředí lze nazývat zkráceně jako „*Environmentální informatika*“ (z angl. Environmental Informatics). Detailní přehled tohoto vědního oboru lze nalézt např. ve sbornících ze série konferencí EnviroInfo (www.enviroinfo2005.org) známé jako „Informatika pro ochranu životního prostředí“.

2 Meta data a sémantický web

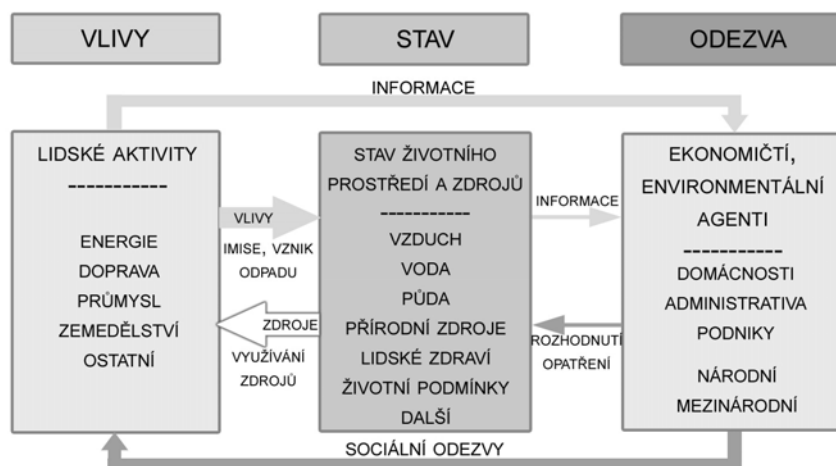
Za účelem zlepšení vyhledávání nepřetržitě rostoucích množin dat a informací bylo v informatice zavedeno pojetí tzv. meta dat (popisných dat). Meta data poskytují informace o datech, ale neobsahují data samotné. Jednou z možností vyvinutí inovačního Internetu v rámci vývoje web 2.0 je sémantický web. Sémantika je tradičně definována jako studium významů slov, frází, vět a textů. Podle toho je sémantický web projektem, který má v úmyslu rozšiřovat vzhled Internetu způsobem

založeným na významech, ne na pouhém kombinování písmen. S pomocí standardů a nástrojů ke zpracování jde o pokus vytvořit všestranné médium pro výměnu informací (http://en.wikipedia.org/wiki/Semantic_web).

Sémantický web má zajistit běžný rámec, který umožňuje sdílet údaje a opětovně je užívat napříč aplikacemi, podniky a hranicemi společností (<http://www.w3.org/2001/sw/>). Cílem Iniciativy sémantického webu je vytvořit všestranné médium pro výměnu údajů. Jde o údaje v současnosti příbuzných databázích, XML dokumenty, tabulkové editory a datové soubory patentovaného formátu – všechny, pro které by bylo užitečné mít přístup k takovéto velké databázi (<http://www.consortiuminfo.org/bulletins/semanticweb.php>).

3 Přechod od informací ke znalostem životního prostředí

Jedním způsobem běžně používaným pro popsání nepříznivých dopadů na životní prostředí a lidské zodpovědnosti je model *Hnací síla – vliv – stav – dopad – odezva* nazývaný *DPSIR* (angl. Driving force-Pressure-State-Impact-Response). Často je používán EEA a mezinárodními organizacemi zabývajícími se informacemi o životním prostředí. Jiné členění *Hnací síla – vliv – stav – expozice – účinek – činnost* nazývané *DPSEEA* (angl. Driving force-Pressure-State-Exposure-Effect-Activity) je používáno ve vědách o veřejném zdraví.



Obr. 2: Indikátory životního prostředí

Zdroj: OECD základní soubor, 1994; Fletcher J. (1999), upraveno

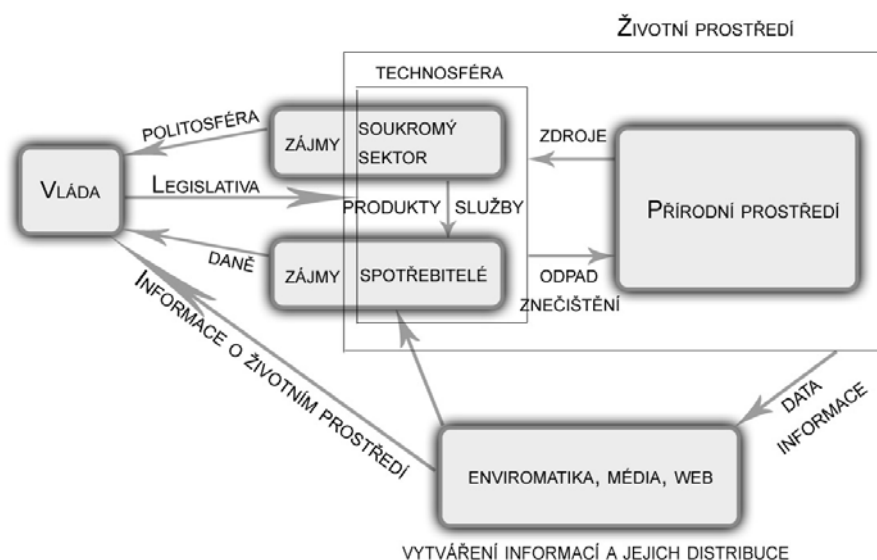
Podíváme-li se detailněji např. na základní model PSR *Vliv – stav – odezva* (angl. Pressure-State-Response) na obr. 2, tak k tomu, aby se daly zkratky D,I,E,A do spojitost s reakcí, zodpovědností a strukturou stimulovaného procesu bychom potřebovali strukturální/terminologické vysvětlení.

Informace spojené s životním prostředím vytvořily a mají v environmentální informatice následující rysy:

1. *Odvolávají se na typ informací nebo oblast použití* (energie, doprava, zemědělství, klimatické změny, odpad a tok materiálu atd.)
2. *Mají prostorový (územní) rozměr* (souřadnice bodových zdrojů, prostředí, města, oblasti, povodí řek, atd.)
3. *Mají časový rozměr* (čas, datum, rok, pozorovací interval, časový sled, atd.)
4. *Jsou prezentovány v určité formě* (osobní komunikace na konferencích, psané materiály, digitální informace: data, databáze, obrázky a mapy, multimediální materiály, webové stránky atd.)
5. *Většinou byly shromážděny odesílatelem a jsou určeny pro příjemce* (úřady, agentury, vlády, vědce, NGO, veřejnost, atd.)

Diskuze na téma udržitelného rozvoje zahrnuje obrovské množství informací. Údaje o životním prostředí a s nimi spojené údaje z národních agentur životního prostředí, výzkumných institutů, EEA (indikátory, základní soubory, Reportnet, EIONET, atd), OECD dokumenty (Náhled na životní prostředí) atd. nabízí základní bázi pro identifikaci a analýzu dynamiky a nezávislosti spojených systémů životního prostředí.

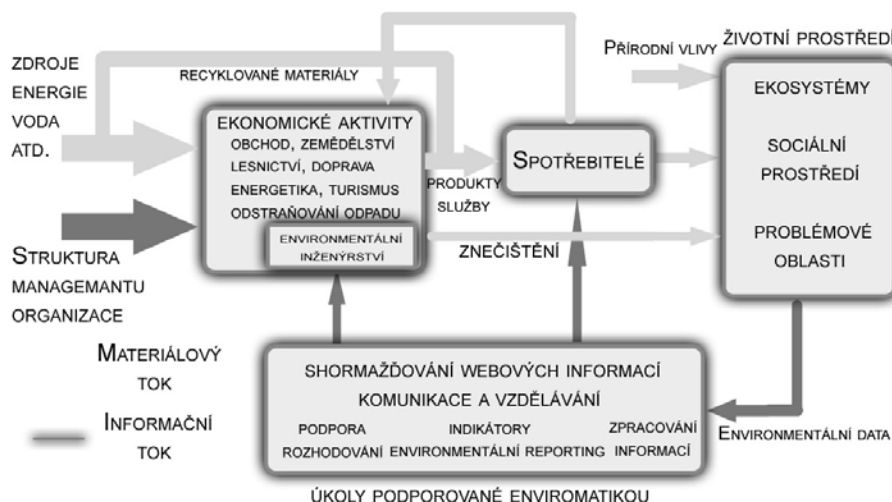
Obr. 3 se pokouší popsat „udržitelný rozvoj“ z hlediska životního prostředí, za použití struktury systémů ve znázornění modelu vstupu/výstupu. Indikátory směřují k vnějším spojením znázornění komunikace mezi ekonomikou, přírodním a umělým životním prostředím (technosférou) a spotřebiteli/voliči.



Obr. 3: Systémový diagram udržitelného rozvoje se vzájemně propojenou výměnou informací
Zdroj: (Pillmann 2005), upraveno

Stejný model technosféry/životního prostředí vzhledem k zdůraznění antropogenních zásahů do životního prostředí, je uveden na obr. 4. Struktura systému zde ukazuje četné zpětné vazby. Cíle životního prostředí (např. snížení znečištění, národní akční plány pro životní prostředí, Místní Agenda 21, Integrovaná výrobní politika (IPP), Systémy environmentálního managementu (EMS) a auditu (EMAS)) ovlivňují politiky, vládní administrativu, průmysl, služby a domácnosti. Vhodná legislativa a politická měřítká, stejně jako daně nakloněné technologiím životního prostředí spolu s informacemi, výzkumem a komunikací životního prostředí limitují dopad podniků, služeb a občanů na životní prostředí.

Environmentální informatika v kombinaci s mnoha dalšími obory pomáhá shromažďovat údaje o prvcích systému ukázaných na obr. 4, aby je pak analyzovala a sdělila těm, co rozhodují. Některé z používaných metod jsou měřítka (znečištění ovzduší/vody, meteorologie atd.), statistika životního prostředí, dálkové snímání parametrů ze senzorů, databáze a GIS aplikace.



Obr. 4: Role environmentální informatiky v podpoře rozhodování
Zdroj: (Pillmann 2005), upraveno

Porozumění rozdílu mezi informacemi a znalostmi o životním prostředí vyžaduje podle Radermachera (2002) nahléd jaký je v jeho „Globálním ekosociálním trhu“. Schmidt-Bleek (2002), dále uvádí, že „budoucí růst ekonomiky závisí na dovednosti v nahrazování hmoty znalostmi“.

Právě environmentální informatika jako významná vědní oblat aplikované informatiky pomáhá zlepšit znalosti o životním prostředí spojováním klíčových výzkumných témat, podporou vědecké spolupráce a vzdělání a vyvíjením nových cest ke sdělování informací mezi technickou společností, politikou a těmi, co rozhodují a veřejností. Získávání znalostí o životním prostředí je individuálním procesem. Inovační učební základny, budování kapacity, účast a sociální učení podporované ICT a environmentální informatikou může proces významným způsobem urychlit. K tomu sloužila 3. letní škola aplikované informatiky, kterou pořádala divize environmentální informatiky a modelování Institutu biostatistiky a analýz v letošním roce.

4 Shrnutí 3. letní školy aplikované informatiky

V této kapitole shrneme témata, kterými se zabývala 3. letní škola aplikované informatiky, která se konala ve dnech 25. až 27. srpna 2006 v Bedřichově. Tato letní škola navázala na předchozí letní školy, které se zde konaly v letech 2002 a 2003 a workshop v Šubířově v roce 2004. V loňském roce byla hlavní akcí Masarykovy university v oblasti environmentální informatiky 19. mezinárodní konference *Informatics for Environmental Protection - EnviroInfo 2005* s nosným tématem *Networking environmental information*, kterou hostila ve dnech 7. až 9. září 2005 se v brněnském hotelu Voroněž.

Všemi přednáškami 3. letní školy prolínala skutečnost, že vývoj ICT pro životní prostředí (potažmo environmentální informatiky) v České republice (ČR) i mezinárodně se zaměřuje na podporu implementace politiky životního prostředí EU a ČR. Monitorovaná data o atmosféře, povrchových i podzemních vodách, biodiverzitě, atd. umožňují efektivnější a přesnější sledování aktuálního stavu životního prostředí a simulaci jeho dalšího vývoje.

Za hlavní přínos 3. letní školy považujeme skutečnost, že na letošním ročníku jsou v příspěvcích zastoupeni jak studenti, doktorandi a učitelé jak z Masarykovy university (Přírodovědecká fakulta, Ekonomicko správní fakulta a Fakulta informatiky), tak i z Mendlovy zemědělské a lesnické university, Vysokého učení technického v Brně, Ostravské university a Technické university – Vysoké školy báňské v Ostravě a jejich příspěvky ve sborníku rozhodně nelze považovat za náhodné. Vyplývá z nich, že oproti minulým ročníkům byla letní škola i výrazněji interdisciplinární. Těžiště projednávaných otázek nebylo jen v aplikované informatice (jako v uplynulých ročnících), ale i v jejích nástrojích a jejím významu pro podnikovou sféru a veřejnou správu, jak vyplývá z Tab. 1.

Tab. 1: Autoři a témata příspěvků

Jana Andrýsková	Návrh a pojetí poradenského systému podle metodiky WASP
Tomáš Foltýnek	Evaluace položek didaktických testů pomocí koeficientu korektnosti
Michal Hejč	Metody hodnocení kvality dat a softwarové možnosti jejího zlepšení
Jiří Hřebíček, Danka Némethová, Jaroslav Ráček	Informační potřeby státní správy, nezávislých organizací, podniků, škol a veřejnosti v oblasti životního prostředí
Karel Kiswa, Matěj Štefaník	Výpočetní knihovna pro systém hodnocení ekologického stavu povrchových vod
Lukáš Kokrment	Ontologie v bioinformatice
Miroslav Kubásek	Optimalizace SOA pomocí cache
Jan Ministr	Metodický rámec stanovení kvality informačního systému pomocí vrstveného modelu
Petra Misařová	Implementace EMS v národních parcích USA
Jan Pavlovič	Servisně-orientovaná architektura systému na výběr dekontaminační strategie
Miroslav Petráš, Jaroslav Ráček	Komunikační systém NOP On-line
Tomáš Pitner, Pavel Drášil	Moderní metodiky, technologie a nástroje pro podporu výuky (E-learning 2.0)
Jaroslav Ráček, Jiří Hřebíček	Informační systém environmentálního reportingu
Jana Sedláčková, František Huňka	Odhady nákladů workflow projektů metodou FPA
Jana Soukopová	Analýza a návrh matematického modelu na podporu rozhodování ve veřejných projektech
Michal Škrdla	Matematický model růstu populací mšic a afidofágů na kukuřičném poli
Matěj Štefaník, Karel Kiswa	Komunikační rámec pro výměnu environmentálních informací
Vladimír Žák	Rychlé grafické algoritmy a jejich využití se systémem Maple

Literatura

- [1] Berners-Lee T.: The Semantic Web - Interview 2005. <http://www.consortiuminfo.org/>
- [2] OECD: Organization for Economic Cooperation and Development: Core Set of Indicators for Environmental Performance Reviews. Monograph no. 83. A Synthesis Report by the Group on the State of the Environment. OECD, Paris, 1993.
- [3] Pillmann, W.: The Integration of Environmental Knowledge. Environmental Informatics Provides Information for Communication. Sborník: EnviroInfo 2005, Brno, 2005.
- [4] Radermacher F.J.: Balance or Destruction: A Global Eco-social market Economy as a Key to Sustainable Development. Sborník: EnviroInfo 2002, Vienna, 2002.
- [5] Schmidt-Bleek, F.: Environmental Information and Sustainability in Perspective. Sborník: EnviroInfo 2002, Vienna, 2002.

Návrh a pojetí poradenského systému podle metodiky WASP

Jana Andryšková

Mendelova zemědělská a lesnická univerzita v Brně, Provozně ekonomická fakulta
Zemědělská 1, 613 00 Brno
xandrysk@pef.mendelu.cz

Abstrakt

Tento příspěvek se zabývá úvodem do metodiky WASP. Jeho cílem je přiblížit životní cyklus a ve stručnosti odpovědět na základní otázky Kdo?, Co?, Kdy? a Jak? Jsou zde zmíněni účastníci projektu a vstupně-výstupní elementy, včetně formálního popisu. Závěr příspěvku se věnuje filozofii zahajovacího a rozvojových cyklů, jejichž logika je zachycena ve formě algoritmu.

Abstract

The paper deals with introduction to WASP methodology. The aim of this paper is to focused the life cycle and answer basic questions Who?, What?, When? and How? in brief. There are discussed project's participants and input/output elements including formal description. Philosophy of starting and evolution cycles and their logic in the form of algorithm are described in the end of paper.

Klíčová slova

Web advisory system process, government advisory portal, metodika, standard, zahajovací cyklus, rozvojové cykly.

Keywords

Web advisory system process, government advisory portal, methodology, standard, starting cycle, evolution cycles.

1 Úvod

Tento příspěvek se věnuje úvodu do metodiky pro vývoj poradenských portálů veřejné správy. Obecně metodik, které se zabývají budováním softwaru je velké množství. Problém spočívá v tom, že nejsou dostatečně a jednotně popsány ani rozumně korigovány [5]. Rovněž nejsou definována jasná kritéria pro výběr vhodné metodiky, pro určitý typ softwarového produktu, ani postupy pro její přizpůsobení na konkrétní podmínky organizace či projektu. Standard ISVS 005/02.01 [7] sice zmiňuje obecné metodiky, které je vhodné pro tvorbu informačních systémů veřejné správy použít, ale ty jsou převážně zaměřeny na vývojové projekty a neřeší například fáze provozu či údržby. Standard zároveň nelze použít jako metodiku [5]. Definuje pouze minimální náležitosti, které by měla každá metodika, životní cyklus informačního systému nebo výstupní dokumenty projektů obsahovat. Musí být „vždy“ používán ve spojení s určitou vybranou metodikou. Nepředepisuje ani specifický model životního cyklu, specifickou metodiku vývoje ani metody pro vypracování jednotlivých položek dokumentace.

Na tomto dokumentu lze rovněž vidět problém nedostatku tradičních českých metodik. Většina předkládaných přístupů je totiž v angličtině a nemá českou lokalizaci, což je v podmínkách veřejné správy závažný nedostatek. Řada kvalitních a propracovaných metodik je šířena na komerční bázi. České firmy, a instituce veřejné správy zvláště, nechtějí nebo nemohou vydávat finanční prostředky na jejich nákup. Lze konstatovat, že pro oblast eGovernmentu České republiky neexistuje žádná metodika pro webové poradenské systémy veřejné správy, standardizace internetového poradenství zde chybí úplně.

2 Úvod do metodiky WASP

Úkolem metodiky WASP (*Web Advisory System Process*) je poskytnout potřebné komplexní postupy a pravidla pro vývoj poradenského portálu veřejné správy, tzv. GAP (*Government Advisory Portal*) systému, a tvůrcům těchto systémů podat odpověď na otázky Kdo?, Co? Kdy? a Jak? Metodika WASP

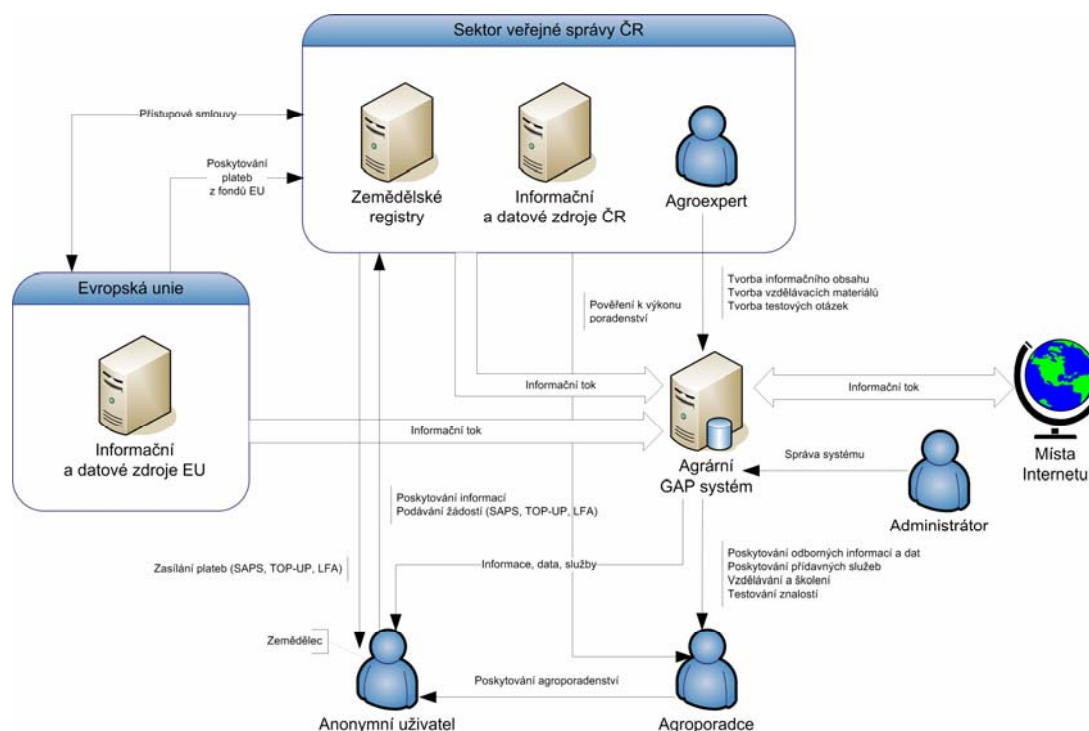
vychází z tradičních i agilních přístupů k vývoji softwaru, je zaměřena na specifické druhy projektů eGovernmentu [3] a do budoucna může být šablonou specifické aplikace *Standardu ISVS 005/02.01 (pro náležitosti životního cyklu informačního systému)* pro veřejné poradenské systémy. Disponuje určitými charakteristikami, které ji odlišují od jiných, existujících metodik:

- GAP systém má charakter informačního systému veřejné správy a podléhá specifickým pravidlům a standardům v této oblasti;
- GAP systém je procesně i funkčně zaměřen na specifickou oblast poradenství, pro kterou zatím neexistují žádná využitelná pravidla či metodiky;
- GAP systém má úzce profilovaný segment uživatelů podle své typové orientace;
- GAP systém vykazuje znaky znalostního systému, což vyžaduje nasazení a účast velkého počtu expertů podle typové orientace;
- GAP systém operuje v prostředí s mezinárodní působností a dosahem (zejména vazby na EU a její legislativu);
- GAP systém využívá moderních webových technologií a portálového řešení informací a služeb.

2.1 Účastníci (Kdo?)

Pod účastníkem metodiky WASP je třeba si představit roli, kterou daná osoba či subjekt v projektu zastává. Metodika WASP rozlišuje následující účastníky projektu vývoje GAP systému:

- **Klient** – rolí typu *Klient* je každá instituce veřejné správy, která vystupuje v pozici zadavatele projektu a která je zastupována pověřenou osobou či osobami (např. ministerstvo, krajský úřad);
- **Uživatel** – rolí typu *Uživatel* je každá osoba, kterou lze označit za potenciálního uživatele systému a která anonymně využívá informačních či jiných systémových služeb bez nutnosti autentizace. Třída osob s touto rolí je přímo závislá na typové orientaci poradenského systému (např. zemědělství, právo, ekologie);
- **Poradce** – rolí typu *Poradce* je každá taková osoba, která disponuje specifickými znalostmi a schopnostmi, které získává využíváním nadstandardních informací a služeb GAP systému a zároveň nezávisle na něm poskytuje přímé poradenství třetím osobám (např. agroporadce, finanční poradce). Podle A. G. Wattse [10] patří tato forma poradenství do modelu *samostatného způsobu využití IT*, kdy ICT se používají odděleně od jiných poradenských aktivit;
- **Expert** – rolí typu *Expert* je každá taková osoba, která má specifické znalosti z daného oboru a která se přímo účastní tvorby obsahové části GAP systému. Metodika WASP zahrnuje do této třídy rolí také tzv. tutorů vzdělávacích materiálů či examinátory, kteří sestavují didaktické testy pro evaluaci znalostí poradců. Možnostmi kvantifikace znalostí poradců se zabývá článek J. Andráskové a T. Foltýnka [2];
- **Tvůrce** – rolí typu *Tvůrce* je každá taková osoba, která se přímo podílí na analýze, návrhu, implementaci, testování, provozu a údržbě GAP systému. Jde o množinu osob, kterou metodika WASP označuje pojmem vývojový tým – VT (např. analytik, programátor, tester). Zároveň platí, že VT je vlastní podmnožinou množiny RT (realizační tým), $VT \subset RT$. Tato inkluze nepřipouští rovnost, neboť metodika WASP předpokládá, že $VT \neq RT \wedge |VT| < |RT|$;
- **Správce** – rolí typu *Správce* je každá taková osoba, která se přímo nepodílí na vývoji GAP systému, ale zajišťuje řízení projektu nebo jeho správu (např. projektový manažer, účetní). Jde o třídu osob, pro které platí $a \in RT \wedge a \notin VT$.



Obr. 1: Participanti GAP systému pro oblast agroporadenství [4]

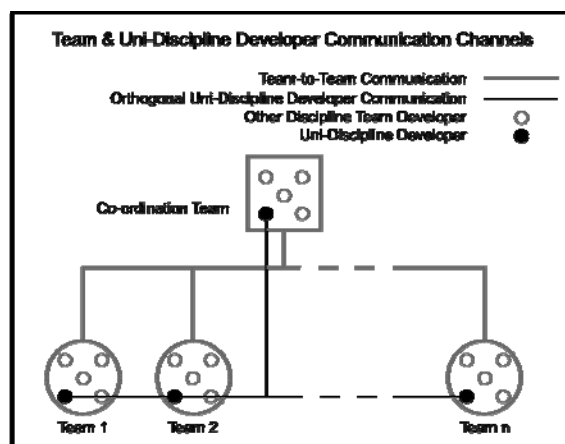
Metodika WASP předpokládá pouze jeden realizační tým, který lze definovat jako uspořádanou trojici $RT = (ET, VT, ST)$ pro $ET \cap VT \cap ST = \emptyset$, kde ET je množina všech osob s rolí *Expert*, VT je množina všech osob s rolí *Tvůrce* a ST je množina všech osob s rolí *Správce*.

Metodiku WASP lze přizpůsobit také projektům většího rozsahu a vytvářet několik tzv. *rozvojových cyklů* (viz dále) najednou. To je možné realizovat pouze za předpokladu změny v organizaci řízení, kdy dochází k dekompozici vývojového i expertního týmu.

Označme $VT_{\Sigma} = \bigcup_{i=1}^n VT_i$ a $ET_{\Sigma} = \bigcup_{i=1}^n ET_i$, kde VT_i je i -tý vývojový tým ($VT_{\Sigma} = \{VT_i \mid i = 1, \dots, n\}$),

ET_i je i -tý expertní tým ($ET_{\Sigma} = \{ET_i \mid i = 1, \dots, n\}$) a n je celkový počet podřízených týmů rozšířeného projektu metodiky WASP. Platí, že $VT_i \cap VT_j = \emptyset$ pro $i \neq j$ a $ET_i \cap ET_j = \emptyset$ pro $i \neq j$.

Rovnice $|VT_{\Sigma}| = |ET_{\Sigma}|$ je důležitým předpokladem pro definici realizačního týmu rozšířeného projektu, který předpokládá, že každý vývojový tým bude spolupracovat s jedním týmem expertním. Zavedme zobrazení $\alpha: VT_{\Sigma} \rightarrow ET_{\Sigma}$, které každému vývojovému týmu přiřadí jeden tým expertů. Pravidlo přiřazení konkrétního expertního týmu vývojovému týmu nebude uvedeno, neboť se předpokládá, že všechny vývojové i expertní týmy jsou stejně schopné a vyřeší zadaný úkol se stejnou pravděpodobností.



Obr. 2: Projektové řízení rozsáhlých projektů (zdroj AWE [6])

Za předpokladu optimálního stavu, kdy každý z týmů je plně využit, lze konstatovat, že uvedené zobrazení je bijekcí, protože splňuje podmínky injekce (dva vývojové týmy nemohou zatěžovat jeden tým expertů a tedy nemůže nastat situace, kdy by jeden expertní tým byl nevyužitý) a také surjekce (neexistuje expertní tým, jehož informace by neměl kdo zpracovat).

Nechť $RT = (ET, VT, ST)$ je realizační tým základního projektu metodiky WASP. Pro rozšířený projekt definujeme realizační tým jako $RT' = (T, ST)$, kde T je množina všech týmů, které jsou řízeny

účastníky s rolí Správce z množiny ST . Pro množinu T platí, že $T = \bigcup_{i=1}^n T_i$, kde n je celkový počet

týmů a T_i je i -tý tým, definovaný jako uspořádaná dvojice $T_i = (VT_i, ET_i)$ pro $VT_i \in VT_\Sigma$ a $ET_i \in ET_\Sigma$.

2.2 I/O Dokumenty a I/O Kódy (Co?)

Vstupy a výstupy jednotlivých činností a cyklů mohou mít dvojí charakter:

- **I/O Dokument** – procesní element typu *I/O Dokument* je každý dokument, smlouva či protokol, který je vyhotovený v tištěné podobě a který odpovídá standardům ISVS. Každý Dokument může obsahovat text či vizuální schémata v podobě tabulek, grafů, obrázků či modelů. Podle obsahové náplně rozlišuje metodika WASP *Dokumenty systémové* (např. Detailní analýza a návrh) a *Dokumenty řídicí* (např. Podrobný plán cyklu).
- **I/O Kód** – procesní element typu *I/O Kód* je každá spustitelná aplikace nebo zdrojový kód zapsaných pomocí příkazů daného programovacího jazyka, který je vytvořen v rámci jednoho cyklu.

Cílem uvedených procesních elementů je udržovat přehled o provedené analýze, návrhu a vývoji, požadavcích účastníků, provedeném kódování, testech, změnách, akceptaci či zamítnutí výstupů klientem. Metodika WASP přesně stanovuje jejich obsahovou náplň i fázi iterace, ve které se tyto vstupně-výstupní elementy realizují. Na konci každého cyklu se realizují tzv. dodávky cyklu, která je předána klientovi. Předmět dodávky lze formalizovat následovně.

Nechť E_D je neprázdná konečná množina všech procesních elementů typu *I/O Dokument* a nechť E_K je neprázdná konečná množina všech procesních elementů typu *I/O Kód*. Za předpokladu, že výstupem každého cyklu je právě jeden prvek množiny E_K , lze dodávku i -tého cyklu (DC_i) pro $i \in \{k \in \mathbb{N} - \{0\} \mid k \leq n\}$, kde n je celkový počet vývojových cyklů, definovat jako podmnožinu kartézského součinu:

$$DC_i \subseteq \prod_{j=1}^2 E_j, \text{ kde } E_1 = E_K \text{ a } E_2 = 2^{E_D} - \emptyset \text{ je potenční množina množiny } E_D \text{ bez prázdné množiny}$$

(neboť dodávka jakéhokoli cyklu metodiky WASP musí obsahovat alespoň jeden procesní element typu *I/O Dokument*).

Uvedená relace platí pro dodávky:

- **Dodávka Zahajovacího cyklu:** DC_1
- **Dodávky Rozvojových cyklů:** DC_i pro $i \in \langle 2, n \rangle$, kde n je celkový počet cyklů.

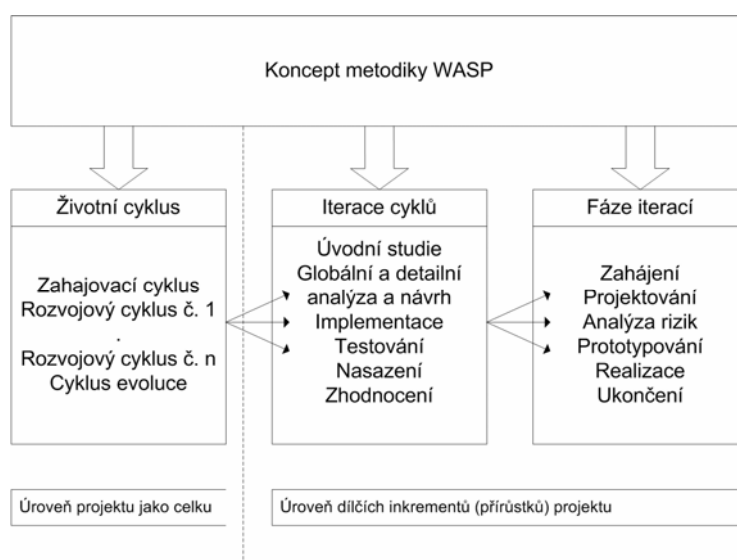
Výše uvedený formalismus platí za předpokladu, že je znám celkový počet prvků množin E_D a E_K . Tento údaj je reálně dostupný až po skončení vývojové etapy metodiky WASP. Proto je vhodné definovat dodávku každého cyklu vztahenou k danému časovému okamžiku a vytvářet relace na základě hodnot, které v daný okamžik reálně existují. Definovat hodnoty E_1 a E_2 pak lze následovně:

- $E_1 = E_{K_i}$, kde E_{K_i} je neprázdna končená množina všech procesních elementů typu I/O Kód, které reálně existují v době předání dodávky i -tého cyklu,
- $E_2 = 2^{E_{D_i}} - \emptyset$, kde E_{D_i} je neprázdna končená množina všech procesních elementů typu I/O Dokument, které reálně existují v době předání dodávky i -tého cyklu,

Zároveň platí: $E_{K_{i+1}} = E_{K_i} \cup \{k\}$, kde k je element typu I/O Kód vytvořený v průběhu cyklu $i+1$,
 $E_{D_{i+1}} = E_{D_i} \cup D$, kde D je množina všech elementů typu I/O Dokument vytvořených v průběhu cyklu $i+1$.

2.3 Zahajovací a rozvojové cykly (Kdy? a Jak?)

Odpovědi na otázky *Kdy?* a *Jak?* vychází z návrhu a popisu životního cyklu metodiky WASP, který je rozdělen na **etapu vývoje** (akvizice, vývoj) a **etapu evoluce** (provoz, údržba). Navržené přístupy a pravidla není možné pojmut v rozsahu tohoto článku, proto zde budou nastitěny pouze základní principy. Etapa vývoje je realizována inkrementálním a zároveň iterativním způsobem. Inkrementální přístup umožňuje vývoj jádra aplikace nezávisle na jednotlivých přírůstcích (modulech). Označme první cyklus metodiky WASP jako **Zahajovací cyklus** (Starting Cycle) a další cykly jako **Rozvojové cykly** (Evolution Cycles).



Obr. 3: Základní koncept životního cyklu metodiky WASP

Je zřejmé, že počet rozvíjecích cyklů přímo závisí na rozsahu konkrétního projektu, a celý vývoj aplikace se bude odvíjet od činností a úkolů definovaných v Zahajovacím cyklu. Současně s každým dalším přírůstkem (Rozvojovým cyklem) je potřeba určit také jeho vazby na jádro portálu, případně na již vytvořené moduly. Aby byl životní cyklus metodiky WASP kompletní, je nutné definovat ještě tzv. **Cyklus evoluce**, který by pokryl etapu evoluce.

Kromě plánovaných cyklů se mohou vyskytovat také cykly neplánované. Typickým příkladem je implementace neplánovaného modulu, který reaguje na aktuální požadavky klienta. Z těchto poznatků plyne, že rozvíjecí cykly metodiky WASP mohou mít dvojí charakter:

- **Plánovaný inkrement** (naplánovaný v Zahajovacím cyklu),
- **Neplánovaný inkrement** (stanovený v některé další generaci cyklů).

Logiku cyklů metodiky WASP je možné definovat jednoduchým algoritmem. Nechť I je konečná neprázdná množina všech iterací každého cyklu, F je konečná neprázdná množina všech fází každé iterace, n je celkový počet rozvojových cyklů metodiky WASP (počet přírůstků jádra), E_D je neprázdná konečná množina všech předpokládaných procesních elementů typu I/O Dokument a $P(E_D)$ její potenční množina, E_K je neprázdná konečná množina všech předpokládaných procesních elementů typu I/O Kód a $P(E_K)$ její potenční množina.

```
WASP(n, I, F, P(ED), P(EK))
{
  Vykonej_Cyklus(cislo_cyklu)
  D ← (D ∩ P(ED) ∧ D je množina dokumentů definovaná pro cyklus: cislo_cyklu)
  K ← (K ∩ P(EK) ∧ K je množina kódů definovaná pro cyklus: cislo_cyklu)
  for každou iteraci i ∈ I do
    for každou fázi f ∈ F do
      d ← (d ∈ D ∧ d je dokument definovaný pro danou fázi)
      k ← (k ∈ K ∧ k je kód definovaný pro danou fázi)
      Proved_Fázi_Iterace(cislo_cyklu, i, f);
      dout ← VytvořDokument(d);
      kout ← ImplementujKód(k);
      Out ← Out ∪ {dout} ∪ {kout}
    od
  od
  return Out;
```

Algoritmus:

```
1. Out ← ∅; //Výstup cyklu - dodávka
2. Ink ← 1;
3. OpakujCyklus ← False;
4. Vykonej_Cyklus(Ink); //Zahajovací cyklus - vytvoření jádra
5. while Ink ≤ n+1 do
6.   if OpakujCyklus = False
7.     then Ink ← Ink + 1
8.   fi
9.   VykonejCyklus(Ink); //Rozvojové cykly - vytvoření přírůstků
10.  if klient akceptuje Out
11.    then OpakujCyklus ← False //Instance plánovaného inkrementu
12.    else OpakujCyklus ← True //Instance neplánovaného inkrementu
13.  fi
14. od
15. VykonejCyklus(n+2); //Cyklus evoluce - provoz a údržba
}
```

3 Závěr

V průběhu posledních desetiletí začaly vlády do své práce postupně zavádět informační a komunikační technologie. Každý úřad chtěl dosáhnout určitého stupně automatizace a požadoval určitý systém pro zpracování dat, který nakonec vybudoval. Do nově vzniklého softwaru se však zakódovaly všechny staré postupy, procesy a organizační formy. Důsledkem toho jsou dnes vládní organizace zcela zablokovány ve starých strukturách a ve starých metodách práce, přičemž každá z nich pracuje se svým vlastním, izolovaným ostrůvkem informačních a komunikačních technologií [9]. Takováto

situaci přímo volá po zavedení nových přístupů k technologiím, nových standardech a především metodikách, bez kterých by tyto přístupy nebyly účelné.

Společnou charakteristikou všech metodik je jejich široký záběr a prakticky univerzální nasazení [5]. Systémy veřejné správy, a poradenské systémy zvláště, však vyžadují jistý odlišný přístup, neboť musí vycházet ze specifické legislativy a standardů [1]. V tomto směru nelze považovat vývoj systémů veřejné správy za uspokojivý. Například Standard ISVS 005/02.01 [7] ukládá pouze povinnost dokumentovat průběh životního cyklu informačních systémů veřejné správy, ale již neřeší podrobnější problémy spojené s volbou konkrétních prostředků a implementací. Dalším specifickým jevem je zapojení poradenství do informačních systémů veřejné správy. Využití informačních a komunikačních technologií se v tomto procesu jeví jako účelné, nutno podotknout, že pro autoregulované internetové poradenství [8] metodické pokyny ani standardy neexistují. V reálných projektech je postrádána metodika, která by zohledňovala jak oblast eGovernmentu, tak oblast autoregulovaných poradenských procesů a portálových řešení webových systémů. Do budoucna by mohla být řešením právě metodika WASP.

Tento příspěvek stručně popisuje základní principy životního cyklu metodiky WASP, jehož úkolem je podat realizačním týmům odpověď na čtyři základní otázky: Kdo?, Co?, Kdy? a Jak? Odpovědi na otázku Kdo? je charakteristika hlavních účastníků a koncept realizačního týmu základního a rozšířeného projektu. Dokumenty a zdrojové kódy mají charakter vstupně-výstupní a odpovídají na otázku Co?. Tyto elementy jsou vždy vztaženy k určitému cyklu metodiky WASP tak, aby bylo možné formálně definovat tzv. dodávku cyklu. Základem celého konceptu životního cyklu je stanovení cyklů, které prostřednictvím iterací (činností) a jejich etap podávají odpověď na otázky Kdy? a Jak?. V metodice WASP, která je realizována inkrementálně-iteračním přístupem, mají význam především tzv. *Zahajovací cyklus*, ve kterém je realizováno jádro, a tzv. *Rozvojové cykly* pro implementaci přírůstků. Logika obou cyklů je zachycena jednoduchým algoritmem, lze ji také zobecnit například formálními jazyky a automaty. V tomto příspěvku není bohužel prostor pro vymezení metodických zásad a postupů internetového poradenství na portálech eGovernmentu ČR, které tato metodika rovněž pokrývá.

Význam metodiky WASP spočívá v rozšíření metodické oblasti eGovernmentu České republiky, neboť pro takto specifický druh projektu byla podobná metodika zatím postrádána. Standardy ISVS i tradiční metodiky nelze účelně použít, neboť jsou příliš obecné a ve své podstatě nemohou nahradit tuto úzce profilovanou metodiku. Tato tvrzení vychází ze zkušeností reálných projektů poradenských systémů (viz projekt eFarmer [4]).

Literatura

- [1] Andrásková, J.: Standardizace systémů veřejné správy. In sborník Firma a konkurenční prostředí 2005. Konvoj, Brno, 2005.
- [2] Andrásková, J., Foltýnek, T.: Vícerozměrný model pro stanovení AAC. In IMEA 2006. Gaudeamus, Hradec Králové, 2006.
- [3] Andrásková, J., Hřebíček, J.: Advisory Web Portals in the eGovernment of the Czech Republic. In IADIS International Conference e-Society 2005. IADIS Press, Malta, Qawra, 2005.
- [4] Andrásková, J., Mišovič, M.: vzdělávání zemědělských poradců států V4 v systému eFarmer. In sborník Agrární perspektivy XIV. ČZU, Praha, 2005.
- [5] Buchalceková, A.: Metodiky vývoje a údržby informačních systémů. Grada Publishing, Praha, 2005.
- [6] McDonald, A., Welland, R.: Agile Web Engineering (AWE) Process. Dostupné na: <http://www.dcs.gla.ac.uk/publications/PAPERS/7087/TR-2001-98%5B1%5D.pdf>.
- [7] Ministerstvo informatiky: Standard ISVS pro náležitosti životního cyklu informačního systému – 005/02.01. Dostupné na: http://www.micr.cz/files/457/uvis-S005.02.01_V2002c5-20021218.pdf.
- [8] Klærke, L.: Study of Optimal IT Using in Education and Employment Guidance. Interprint, København, 1999.

- [9] Tapscott, D.: The Digital Economy: Promise and Peril in the Age of Networked Intelligence. McGraw-Hill, New York, 1996.
- [10] Watts, A. G.: Computer in Guidance. In Rethinking Careers Education and Guidance: Theory, Policy and Practice. Routledge, New York, 1996.

Evaluace položek didaktických testů pomocí koeficientu korektnosti

Tomáš Foltýnek

Mendelova zemědělská a lesnická univerzita v Brně, Provozně ekonomická fakulta
Zemědělská 1, 613 00 Brno
foltynek@pef.mendelu.cz

Abstrakt

Příspěvek popisuje zobecnění metod skórování položek didaktických testů pomocí koeficientu korektnosti. Přináší komplexní formalizaci procesu skórování, zasazuje současné používané metody do obecného kontextu a přináší metody nové. Detailně jsou rozebrány metody skórování seřadovacích položek a korekce hádání. Poznatky popsané v tomto příspěvku mohou pomoci examinátorům pro přesnější stanovení výsledků didaktických testů.

Abstract

This paper deals with achievement test item scoring generalization with the help of correctness coefficient. The paper brings complex formalisation of scoring process, contextualises contemporary used methods to the general context and brings new methods. The scoring methods of sorting items, and guessing penalty is described in detail. Observations described in this paper can help to examiners with more accurate assessment of achievement test results.

Klíčová slova

Didaktický test, skórování, elektronické testování, koeficient korektnosti.

Keywords

Achievement test, scoring, electronic testing, correctness coefficient.

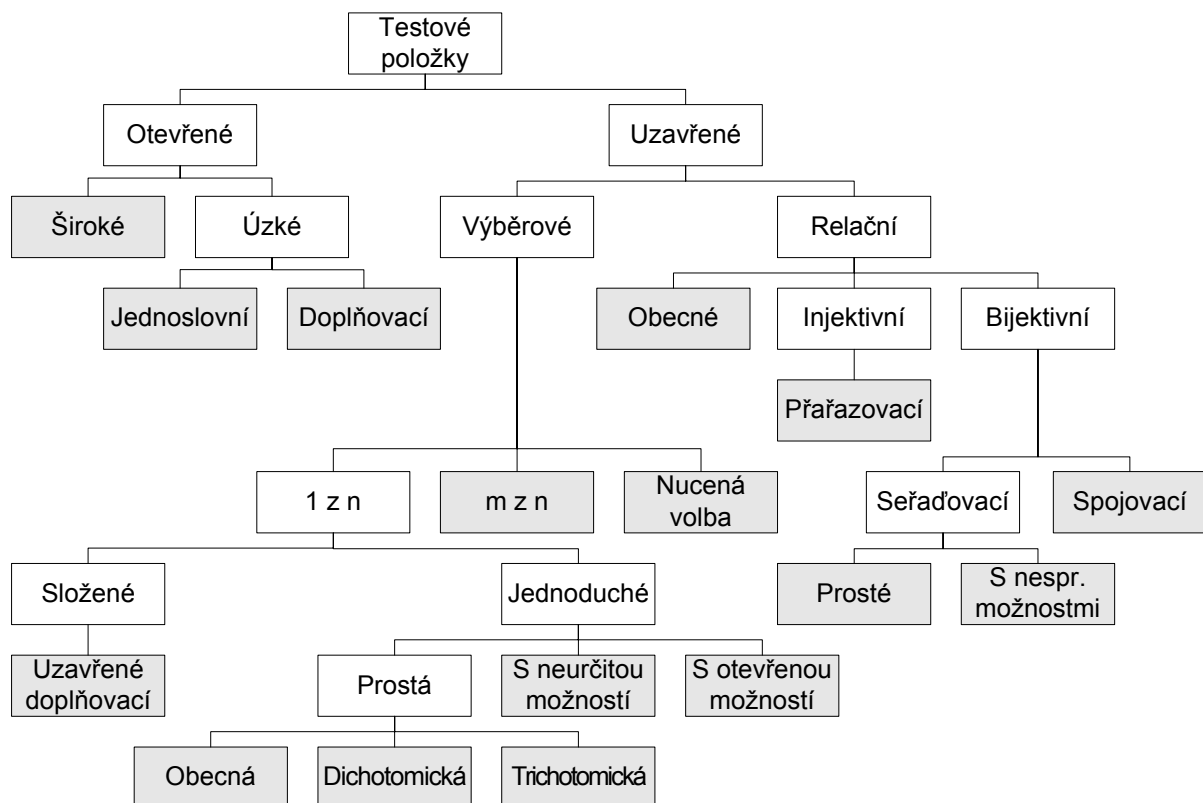
1 Úvod

Didaktický test je považován za nejobektivnější nástroj pedagogické evaluace [8]. Jednotlivé testové otázky – položky – mohou však být vyhodnocovány – skórovány – rozdílnými metodami, při nichž je dosahováno různých výsledků. Tento problém vyvstává zejména u položek složitějších typů, kde nejsou skórovací metody zřejmé na první pohled [1]. V následujícím textu zobecníme skórovací metody a analyzujeme možnosti jejich konkretizace s ohledem na jejich metodickou vhodnost v závislosti na typu zkoušeného učiva, očekávané hloubce poznatků studentů, smyslu testu a dalších charakteristikách. Ve světě, kdy automatizované počítačem podporované testování nabývá stále většího významu a výsledky didaktických testů jsou považovány za absolutně správný ukazatel znalostí zkoušeného [1], je namístě věnovat skórovacím metodám adekvátní péči pro zachování objektivitu didaktických testů.

V následujícím textu zavedeme formalizaci evaluace položek didaktických testů pomocí koeficientu korektnosti, který – v závislosti na typu testové položky – vyjadřuje míru studentovy chyby. Stanovení skóre testu je tak daleko přesnější, než by tomu bylo v případě, kdybychom uvažovali jen mezní hodnoty správně – špatně.

2 Typy testových položek

Testové položky (častěji označované jako *otázky*) lze rozdělit na mnoho typů [4]. Strom rozdělení testových položek je znázorněn na následujícím obrázku:



Obr. 1: Rozdělení typů testových položek

3 Matematický základ skórování

Testová položka je určena těmito prvky:

- Textovým řetězcem **Z**, tvořícím zadání otázky, které specifikuje, co má zkoušený provádět. Na zadání jsou kladeny určité metodické požadavky, na způsob vyhodnocování však nemá vliv.
- U výběrových otázek konečnou množinou alternativ (tedy nabízených možností) **M**, která je zkoušenému předložena ve zcela náhodném uspořádání.
- U relačních otázek množinou objektů **A** a množinou skupin **B**, které jsou zkoušenému také předloženy ve zcela náhodném uspořádání

Dále pak definujeme prostor řešení Ω jako množinu všech možných odpovědí na danou otázku. Pro Ω přitom platí:

$\Omega = \{\omega; \omega \text{ je textový řetězec}\}$ pro otevřené otázky

$\Omega = P(M)$, tedy potenční množina množiny **M** pro výběrové otázky

$\Omega = P(A \times B)$, tedy množina všech relací mezi množinami **A** a **B** pro relační otázky

Odpověď zkoušeného, tedy prvek množiny Ω budeme označovat ω . Nyní hledáme funkci, která z odpovědi zkoušeného vypočítá dílčí skóre položky. Nazvěme tuto funkci *evaluační funkcí* a označme ji f . Definičním oborem funkce je prostor řešení Ω a oborem hodnot je interval $<-1,1>$. Platí tedy

$$f: \Omega \rightarrow <-1,1>.$$

4 Koeficient korektnosti κ

Abychom mohli formalizmus skórování zpracovat co nejobecněji, je třeba skórování položek každého typu rozložit na co nejelementárnější prvky.

- U otevřených položek uvažujeme textové řetězce
- U výběrových položek uvažujeme jednotlivé alternativy (tedy prvky množiny $\mathbf{M}$)
- U injektivních a obecných relačních položek uvažujeme prvky kartézského součinu $\mathbf{A} \times \mathbf{B}$ (tedy uspořádané dvojice (a, b) , kde $a \in \mathbf{A}$ a $b \in \mathbf{B}$).
- U bijektivních položek uvažujeme jednotlivé bijekce mezi množinami $\mathbf{A}$ a $\mathbf{B}$. V dalším textu budeme množinu všech takovýchto bijekcí označovat symbolem $Bij(\mathbf{A}, \mathbf{B})$.

Pro každý z těchto elementů definujeme jeho koeficient korektnosti, který označíme κ , a který bude základem pro výpočet dílčího skóre položky. Koeficient korektnosti bude mít stejný obor hodnot jako dílčí skóre položky; jak ukážeme později, v některých případech si tyto hodnoty budou rovný. Z metodického hlediska se v koeficientu korektnosti skrývá míra správnosti (případně míra špatnosti) každé dílčí odpovědi. Vysvětlíme nyní sémantiku koeficientu korektnosti u jednotlivých skupin typů testových položek.

U otevřených položek je koeficient korektnosti přiřazen každému textovému řetězci. Je zřejmé, že pro většinu řetězců bude platit, že $\kappa = 0$, to budou obecně špatné či nesmyslné odpovědi. Pro zcela správnou odpověď bude $\kappa = 1$. V případě, že je zcela správných odpovědí více, budou mít všechny koeficient korektnosti roven jedné. Mohou však existovat pouze částečně správné odpovědi. V tom případě nabývají na významu hodnoty zevnitř intervalu $(0, 1)$, které budou určovat, jak moc se zkoušený správné odpovědi přiblížil. Otevřené položky jsou příkladem položky, pro něž platí rovnost mezi koeficientem korektnosti a dílčím skóre položky.

U výběrových položek je koeficient korektnosti přiřazen každé alternativě. Nezáleží přitom na tom, o který typ výběrové položky se jedná. Je totiž potřeba si uvědomit, že výběrové položky typu m z n jsou de facto soustavami dichotomických položek, které jsou samy o sobě zvláštním případem výběrových položek typu 1 z n . Stačí tedy numerizaci skórování zpracovat pouze pro výběrové položky typu 1 z n a získané poznatky pak bude možné aplikovat i na další typy.

Je tedy dána množina alternativ, z nichž má zkoušený za úkol vybrat právě jednu. V nejjednodušším případě přiřadíme jediné správné alternativě $\kappa = 1$ a všem ostatním alternativám $\kappa = 0$. Nyní je ale třeba uvažovat případ, kdy je třeba odlišit chybu plynoucí ze zvolení jednotlivých špatných alternativ. Ne všechny mohou totiž být stejně špatné a může se dokonce stát, že jedna z nich je „skoro správná“. V takovém případě je tedy, podobně jako tomu bylo u otevřených položek, na místě uvažovat i hodnoty koeficientu korektnosti zevnitř intervalu $(0, 1)$. Dále se může stát, že mezi položkami jsou takové, jejichž zvolení poukazuje na hrubou neznalost, kterou je třeba potrestat záporným bodovým hodnocením a můžeme tak uvažovat i hodnoty z intervalu $(-1, 0)$.

Vzhledem k tomu, že položky typu m z n jsou, jak již bylo řečeno, soustavami položek typu 1 z n se dvěma (případně třemi) konstantními alternativami, je třeba u výběrových položek nadefinovat koeficient korektnosti u každé alternativy jak pro případ, kdy zkoušený danou alternativu zvolí, tak pro případ, kdy ji nezvolí.

Injektivní položky jsou de facto soustavou výběrových položek typu 1 z n s konstantní množinou alternativ rovnou množině $\mathbf{B}$. Koeficient korektnosti tedy definujeme pro každou dvojici kartézského součinu $\mathbf{A} \times \mathbf{B}$ a jeho sémantika tkví v ohodnocení správnosti zařazení objektu ze skupiny $\mathbf{A}$ do skupiny $b \in \mathbf{B}$. Hodnoty 0 nebo 1 , respektive hodnoty zevnitř intervalu $(0, 1)$, případně hodnoty záporné mají stejný význam jako tomu bylo u výběrových položek typu 1 z n .

Na obecné relační položky lze pohlížet jako na soustavu výběrových položek typu m z n , kdy u každého objektu z množiny $\mathbf{Q}$ zkoušený rozhoduje, do kterých skupin z množiny $\mathbf{B}$ daný objekt patří. Protože víme, že položky typu m z n jsou vlastně soustavami dichotomických položek, lze i na obecné

relační položky pohlížet jako na soustavu $|\mathbf{A}| \cdot |\mathbf{B}|$ dichotomických položek. Koeficient korektnosti tedy určujeme zvlášť pro každou dvojici (a, b) , kde $a \in \mathbf{A}$ a $b \in \mathbf{B}$ a každý z případů zvoleno – nezvoleno.

Největší problém při stanovování koeficientu korektnosti je u bijektivních položek, a to jak u spojovacích, tak u seřadovacích. Na bijektivní položky totiž nelze pohlížet jako na soustavu výběrových položek, neboť tyto položky by nebyly vzájemně nezávislé. Ukořím zkoušeného sice je, stejně jako u všech ostatních relačních položek, přiřadit každý objekt do správné skupiny, v případě bijektivních položek navíc požadujeme, aby byla každá skupina použita právě jednou (množiny $\mathbf{A}$ a $\mathbf{B}$ jsou v tomto případě zřejmě stejně početné). Nejmenší nezávislý element, jenž jsme schopni v takovémto případě ohodnotit koeficientem korektnosti, je tak celá bijekce. Vzhledem k jejich velkému počtu ($k!$ pro $k = |\mathbf{A}| = |\mathbf{B}|$) je tak potřeba vybrat (kromě správné bijekce) ty, které poukazují na určité specifické chyby a těm přiřadit individuální koeficient korektnosti. U ostatních bijekcí nám pak nezbude, než přiřadit jim nějakou vhodnou implicitní hodnotu κ (např. $\kappa = 0$).

5 Skórování položek testu

Máme zaveden koeficient korektnosti κ pro různé typy položek. Nyní tohoto koeficientu využijeme pro stanovení dílčího skóre položek. Rozebereme si zvlášť všechny typy a možné metody skórování.

5.1 Skórování uzavřených výběrových položek

Nejprve se podívejme na nejjednodušší případ uzavřených výběrových položek, tedy položky typu 1 z n , u nichž zkoušený vybírá z nabízené množiny alternativ *právě jednu*. Připomeňme, že pracujeme s množinou alternativ $\mathbf{M}$. Pro každé $m \in \mathbf{M}$ je definován koeficient korektnosti $\kappa(m)$. Nechť $m_o \in \mathbf{M}$ je odpověď zkoušeného, tedy zvolená alternativa. Dílčí skóre položky je pak dáno vztahem

$$\sigma = \kappa(m_o)$$

Stejný vztah platí i pro zvláštní případy uzavřených výběrových položek, tedy pro položky dichotomické a trichotomické.

Na výběrové položky typu m z n pohlížíme jako na soustavu dichotomických položek. Je-li tedy $\mathbf{M}$ množina alternativ, definujeme koeficient korektnosti pro každou alternativu a každou z možností zvoleno – nezvoleno (0 – 1). Označme $\mathbf{Z} \subseteq \mathbf{M}$ množinu položek, které zkoušený zvolil (označil, zaškrtnl) a $\check{\mathbf{Z}} \subseteq \mathbf{M}$ množinu položek, které nezvolil. Nutně přitom platí, že množiny $\mathbf{Z}$ a $\check{\mathbf{Z}}$ tvoří rozklad množiny $\mathbf{M}$ (tedy $\mathbf{Z} \cup \check{\mathbf{Z}} = \mathbf{M}$ a $\mathbf{Z} \cap \check{\mathbf{Z}} = \emptyset$). Pro dílčí skóre položky pak platí

$$\sigma = \frac{\sum_{m \in \mathbf{Z}} \kappa(m, 1) + \sum_{m \in \check{\mathbf{Z}}} \kappa(m, 0)}{|\mathbf{M}|}$$

5.2 Skórování injektivních a obecných relačních položek

Vyjděme z toho, že úkolem zkoušeného je nalézt relaci mezi dvěma množinami. U relace jako u podmnožiny kartézského součinu platí, že daná dvojice do hledané relace buďto patří, nebo ne. Rozložíme-li tedy hledanou relaci na jednotlivé prvky a uvažujeme-li každý prvek zvlášť jako elementární dichotomickou otázku, lze při skórování relačních otázek aplikovat metody uzavřených výběrových položek.

Nechť $\mathbf{A}$, $\mathbf{B}$ jsou množiny, mezi nimiž se hledá relace. Pak platí, že odpověď $\mathbf{O} \subseteq \mathbf{A} \times \mathbf{B} = \{(a, b) \mid a \in \mathbf{A}, b \in \mathbf{B}\}$. Pro každý prvek kartézského součinu $\mathbf{A} \times \mathbf{B}$ a každý z případů zvoleno, nezvoleno, definujeme koeficient korektnosti κ , tedy zobrazení $\kappa: \mathbf{A} \times \mathbf{B} \times \{0, 1\} \rightarrow \langle -1, 1 \rangle$.

Dílčí skóre uzavřené relační otázky je tedy v takovémto nejobecnějším případě dáno vztahem

$$\sigma = \frac{\sum_{(a,b) \in \mathbf{O}} \kappa(a, b, 1) + \sum_{(a,b) \in \mathbf{A} \times \mathbf{B} \setminus \mathbf{O}} \kappa(a, b, 0)}{|\mathbf{A} \times \mathbf{B}|}$$

5.3 Skórování úzkých otevřených položek

I v tomto případě lze problematiku skórování rozšířit z klasického pohledu „správná odpověď / špatná odpověď“ na obecnější případ, kdy rozlišujeme správnost jednotlivých možných odpovědí. Je zde však potřeba určit implicitní hodnotu, která se použije v případě, že koeficient korektnosti není pro odpověď zkoušeného definován. Výsledné hodnocení úzké otevřené otázky, na níž zkoušený odpověděl slovem s , je pak dáno jednoduchým vztahem

$$\sigma = \kappa(s)$$

5.4 Skórování spojovacích položek

Nyní se budeme věnovat bijektivním položkám kromě položek seřadovacích. Tedy takovým položkám, v nichž je třeba nalézt bijekci mezi dvěma množinami. Připomeňme, že $\mathbf{A}$ a $\mathbf{B}$ jsou množiny objektů, mezi nimiž zkoušený hledá bijekci. Koeficient korektnosti je definován pro každou bijekci, tedy, pro každý prvek množiny $Bij(\mathbf{A}, \mathbf{B})$. Je-li tedy odpověď zkoušeného $\mathbf{O} \in Bij(\mathbf{A}, \mathbf{B})$, je dílčí skóre spojovací položky dáno vztahem

$$\sigma = \kappa(\mathbf{O})$$

Zamysleme se nyní nad tím, jak vypočítat κ pro danou bijekci, která představuje odpověď zkoušeného a rozeberme si různé metody výpočtu. Ve Symbolem $\mathbf{P}$ budeme dále označovat správnou odpověď, tedy správnou bijekci. Platí tedy, že $\mathbf{O} \in Bij(\mathbf{A}, \mathbf{B})$ a zejména $\kappa(\mathbf{P}) = 1$.

Označme symbolem k počet prvků v množině $\mathbf{A}$, resp. $\mathbf{B}$. Platí tedy $k = |\mathbf{A}| = |\mathbf{B}|$. Zkoušený tedy hledá k dvojic. Použijeme-li jako jediné měřítko počet dvojic, které má správně, lze koeficient korektnosti dané bijekce stanovit podle vztahu

$$\kappa(\mathbf{O}) = \frac{|\mathbf{O} \cap \mathbf{P}|}{k},$$

tedy jako podíl velikosti průniku bijekcí (a tedy množin) $\mathbf{O}$ a $\mathbf{P}$ představující počet správných dvojic a počtu všech dvojic.

5.5 Skórování seřadovacích položek

Koeficient korektnosti stanovujeme pro každou permutaci množiny alternativ zvlášť. Označíme-li tedy $Per(\mathbf{M})$ množinu všech permutací množiny $\mathbf{M}$ a $\mathbf{O} \in Per(\mathbf{M})$ odpověď zkoušeného, tedy zvolenou permutaci, dílčí skóre položky je určeno zřejmým vztahem

$$\sigma = \kappa(\mathbf{O})$$

Dále označme $\mathbf{P} \in Per(\mathbf{M})$ správnou odpověď, tedy permutaci představující správné setřídění. Pro tuto permutaci pak zřejmě platí

$$\kappa(\mathbf{P}) = 1$$

Naším cílem bude dále definovat takovou operaci na množině všech permutací konečné podmnožiny přirozených čísel, která by vyjadřovala míru jejich odlišnosti a odpovídala tak míře odlišnosti odpovědi zkoušeného od správné odpovědi. Nazvěme tuto operaci *odlišnost* a označujme ji dále symbolem d .

Definici této operace nyní ponechme stranou a věnujme se nejprve přepočtu jejího výsledku do intervalu $\langle 0, 1 \rangle$, tedy stanovení dílčího skóre položky. Ten je realizován vzorcem

$$\sigma = 1 - \frac{d(\mathbf{O}, \mathbf{P})}{d_{\max}(\mathbf{P})},$$

kde $d(\mathbf{O}, \mathbf{P})$ je odlišnost permutace $\mathbf{O}$ od permutace $\mathbf{P}$ a $d_{\max}(\mathbf{P})$ je odlišnost nejodlišnější permutace od permutace $\mathbf{P}$.

Nyní tedy zbývá definovat zmíněnou operaci odlišnosti. Tu lze definovat mnoha způsoby, které budou korespondovat s různými způsoby vyhodnocování. Metody jsou podrobně rozebrány v [10], na tomto místě uvedeme pouze dvě nejpoužívanější.

5.5.1 Součtová metoda

Nechť $\mathbf{A} = (a_1, a_2, \dots, a_n)$ a $\mathbf{B} = (b_1, b_2, \dots, b_n)$ jsou dvě permutace množiny $\mathbf{N}_n$. Jejich odlišnost d definujeme vztahem

$$d = \sum_{i=1}^n |a_i - b_i|$$

Pro $\mathbf{B} = \mathbf{P} = (1, 2, \dots, n)$ pak dostáváme zjednodušený vztah

$$d = \sum_{i=1}^n |a_i - i|$$

Tato metoda pro každý prvek množiny možností vypočítá, jak daleko je tento prvek umístěn od své správné pozice. Součet těchto rozdílů (vzdáleností) pak vypovídá o celkové míře shody, resp. odlišnosti celých permutací.

Maximální možnou hodnotu $d_{\max}$ dostáváme pro opačné permutace a její hodnota je rovna

$$d_{\max} = \sum_{i=1}^n (n - (n - i + 1)),$$

odkud po úpravě dostaneme

$$d_{\max} = \frac{n^2 - (n \bmod 2)}{2}$$

5.5.2 Metoda sekvencí

Nechť $\mathbf{A} = (a_1, a_2, \dots, a_n)$ je permutace množiny $\mathbf{N}_n$. Její odlišnost d od permutace $\mathbf{P} = (1, 2, \dots, n)$ definujeme vztahem

$$d = (n-1) - \sum_{i=1}^{n-1} j(a_{i+1} - a_i),$$

kde

$$j(x) = \begin{cases} 1 & \text{pro } x = 1 \\ 0 & \text{pro } x \neq 1 \end{cases}$$

Výsledkem této metody je součet délek správných sekvencí bez ohledu na jejich umístění v posloupnosti. Protože maximální možná sekvence má délku $n-1$, vyskytuje se tento výraz na začátku pravé strany a zajišťuje, že výsledkem operace d bude pro maximální možnou sekvenci hodnota 0. Pro permutaci neobsahující žádnou správnou sekvenci dostaneme výsledek $d_{\max} = n-1$.

6 Závěr

V tomto příspěvku jsme zavedli rozdělení typů běžně používaných položek didaktických testů a sumarizovali teoretická východiska skórování těchto položek. Z tohoto základu jsme vybudovali formalizaci skórování položek didaktických testů na základě koeficientu korektnosti.

Ačkoliv je nalezený způsob velmi obecný a poskytuje examinátorovi značnou svobodu, s touto svobodou je také spojená velká práce při stanovování koeficientů korektnosti u všech relevantních entit a s tím související riziko chyby. Teprve budoucí pedagogická praxe tak ukáže, zda jsou způsoby skórování nadnesené v tomto článku vhodné či nikoliv.

7 Literatura

- [1] Foltýnek, T.; Motyčka, A.: K metodám skórování položek didaktických testů. In Efficiency and responsibility in education – proceedings of papers 2006. ČZU, Praha, 2006.
- [2] Hniličková, J.; Josífk, M.; Tuček, A.: Didaktické testy a jejich statistické zpracování. SPN, Praha, 1972.
- [3] Chrástka, M.: Didaktické testy. Paido, Brno, 1999.
- [4] Mužic, V.: Testy vědomostí. SPN, Praha, 1971.
- [5] Payne, D.A.: The specification and measurement of learning outcomes. Blaisdell Publishing Company, Waltham, 1968.
- [6] Řešátko, M.: Didaktické testy ve školní praxi. Praha: SNTL, 1975.
- [7] Smékal, V.; Švec, V.; Zajac, J.: Didaktické testy a jejich vyhodnocování. Brno: Středisko pro výzkum učebních metod a prostředků, 1973.
- [8] Švec, V.: Didaktický test jako účinný prostředek pedagogické diagnostiky. Gottwaldov: Socialistická akademie, 1973.
- [9] Turek, I.: Didaktické testy (Kapitoly z didaktiky). Metodické centrum v Bratislave, Bratislava, 1995.
- [10] Foltýnek, T.: Metody evaluace seřadovacích položek v didaktických testech. In Sborník konference Firma a konkurenční prostředí 2006. MZLU, Brno, 2006.

Metody hodnocení kvality dat a softwarové možnosti jejího zlepšení

Michal Hejč

Masarykova univerzita v Brně, Fakulta informatiky
Botanická 68a, 602 00 Brno
bobson@fi.muni.cz

Abstrakt

V příspěvku jsou popsána úskalí pojmového aparátu v dané oblasti a vytyčeny základní definice, dále jsou shrnuty metody reprezentace dat za účelem hodnocení jejich kvality a také možnosti zlepšení kvality dat za přispění softwarových metod. Navíc jsou prezentovány předběžné informace k zamýšleným případovým studiím.

Abstract

The paper describes problem of terms and definitions in given area, then it summarizes the methods of data representation for the purpose of measurement of their quality. Software options for improving of data quality are discussed. Paper contains some information from supposed case studies.

Klíčová slova

Data, informace, kvalita dat, nejistota v datech, reprezentace dat, metriky, pravděpodobnost, statistika, neurčitost, soft systémy.

Keywords

Data, information, data quality, data uncertainty, representation of the data, metrics, probability, statistics, fuzziness, soft systems.

1 Motivace

Analýzy problémů se dříve často prováděly s velmi omezenými znalostmi, intuitivně, na základě zkušeností nebo jiným ne zcela spolehlivým a současně automatizovatelným způsobem. Vezměme si například problém stanovení výše povinných dávek do státní „pokladny“ v tzv. době předhistorické, v Mezopotámii. Odvody se prováděly v naturáliích, nebylo písmo a tedy ani nebyla žádná evidence. Vládci šlo samozřejmě o to, aby vybrané odvody byly co nejvyšší, ale současně mu bylo zřejmé, že je nutno zachovat určitou udržitelnost, aby poddaní příliš nezchudli a měli dál z čeho odvádět dávky. O tomto problému nikde neexistují žádné záznamy, ale vzhledem k tomu, že dávky se zcela jistě odváděly, problém stanovení jejich výše (pokud nebyla dobrovolná) určitě v té době vyvstal a někdo ho musel řešit (pravděpodobně rádce panovníka).

Naštěstí bylo vynalezeno písmo, věci se daly do pohybu a doba, kdy data a informace byly nedostatkovým zbožím, je již dávno pryč. Přistoupilo se k shromažďování všech možných i nemožných údajů, jako záznamové médium sloužil materiál, na který bylo možno psát. Nejprve hlína a kámen, později papyrus, pergamen a papír.

S jídlem ale roste chuť a požadavky na přesnost analýzy si často vynutily rozsáhlejší sběr dat, který zvládat a hlavně později využívat k analýze nebylo úplně jednoduché. Svoji roli sehrála zejména roztříštěnost sebraných informací, velká pracovní náročnost procesu zpracování, možnost chyby (zejména vlivem lidského faktoru) vzhledem k velkému zpracovávanému objemu, atd.

V moderní době došlo v souvislosti s rozvojem ICT k posunu v této oblasti. Byly odstraněny všechny nedostatky ručního zpracování dat kromě jediného – možnosti chyby v datech. Zpočátku nebyl problém tak významný, protože samotný přechod na využití ICT byl dostatečným motorem pro zlepšování kvality výsledků analýz. Vývoj se ubíral zejména směrem ke zvyšování zpracovávaných objemů a druhů dat. Co bylo možno zjišťovat se zjišťovat skutečně začalo, vrcholem se stalo globální sdílení dat na Internetu.

Zde se ale znovu dostává do popředí problém chyb v datech. Kvalitu analýz již není stávajícími prostředky jak zlepšovat, výrazně větší množství dat už k dispozici nebude a i kdyby bylo, vzorky jsou již natolik reprezentativní, že nárůst kvantity není v mnoha případech třeba.

Dostáváme se tedy k současné situaci, která je motivací pro tuto práci. Je k dispozici velké množství zdrojů dat a informací. Jsou to různé komerční i vládní databáze, encyklopedie, soubory dat a v neposlední řadě obrovský prostor Internetu, na který je možno si strojově „sáhnout“ a získat z něj určitý požadovaný typ dat a informací a dále s ním pracovat. Rozvíjí se služby, které práci s informacemi na Internetu takto strojově organizují (Google).

Stále ale zůstává problém chybovosti veškerých těchto zdrojů. V jednotlivých případech je zkušený analytik schopen jasně rozhodnout, jak s daným údajem nebo informací naložit – tedy jestli mu věřit, nevěřit, věřit částečně, atd. Jak se ale zachovat v případě hromadného strojového zpracování? To je otázka, kterou bych se chtěl zabývat v následujícím příspěvku.

2 Cíle

Již název příspěvku stanovuje dva základní cíle:

- automatizované určení kvality dat (výstupem budou metriky hodnocení kvality dat),
- možnosti automatizovaného zlepšení této kvality dat (výstupem budou postupy pro vylepšení kvality dat, vycházející z nějaké metriky).

K dosažení základních cílů bylo nutno vytyčit několik dílčích cílů:

- pojmový aparát,
- způsoby reprezentace dat s ohledem na možnosti kvantifikace jejich kvality,
- metriky kvality dat,
- možnosti zásahu do dat,
- algoritmy pro některé možnosti zásahu do dat,
- případové studie.

V následujícím textu budou jako etapy podrobněji popsány jednotlivé zamýšlené dílčí cíle a také předpokládané dosažené výsledky z těchto etap.

2.1 Pojmový aparát

Základem veškeré následující teorie je dobře zmapovaný a případně upravený či nově navržený pojmový aparát. Zvolená oblast je velice citlivá na volbu (a interpretaci) správných odborných termínů. V praxi jsou oproti jiným vědním oblastem pojmy častěji zaměňovány, slučovány nebo jinak špatně využívány, takže by bylo velkým paradoxem, aby teorie na odstranění chybovosti nebyla sama prostá chyb již ve své pojmové podstatě.

V dané problematice je možno identifikovat čtyři charakteristické základní množiny pojmů, obsahující konkrétní pojmy s příbuzným významem. Množiny je možno charakterizovat jako:

- vstupy pro analýzu (data, údaje, informace, atd.),
- chyby v datech (nepřesnosti, záměrné lži, chybějící údaje, atd.),
- kvalita dat (důvěryhodnost, přesnost, reprezentativnost, atd.),
- akce nad daty (zásahy, validace, nahrazení, zneplatnění, verifikace, atd.).

Další pojmy (zejména z oblasti metrik a algoritmů) jsou již obecně známy a nejsou pro počáteční pochopení problematiky natolik podstatné. Proto budou uvedeny až v příslušných kapitolách, ke kterým se budou vztahovat.

2.1.1 Vstupy pro analýzu

Vstupy pro analýzu jsou subjekty, nad kterými se bude provádět hodnocení a zlepšování kvality. V našem případě se bude konkrétně jednat o data, ale zde je nutno mít na paměti jaká definice dat je pro nás směrodatná. V souvislosti s tím bude nutno ujasnit vztah mezi daty a informacemi, chápání pojmů metadata, ontologie, údaj, záznam, atd.

Podrobnější definice těchto pojmů budou dále zkoumány, v této fázi zpracování řešení problému je k dispozici zatím jen obecný předpoklad, jak bude situace pravděpodobně vypadat.

Data (množné číslo) budou chápána jako soubor elementárních měřitelných *údajů* (jednotné číslo) nebo *záznamů*. Příkladem budiž databáze hlášení o produkci a nakládání s odpadem jako data a jedna položka (řádek) databáze jako údaj. *Ontologie* je chápána jako popis významu dat (vazba na jiná data) a *metadata* vzniknou sloučením ontologie a dat. *Informace* je konkrétní závěr, který jsme schopni z dat (metadat) nějakým způsobem vyvodit. Analýzu dat provádíme právě za účelem vyvození konkrétní informace, takže požadovaná informace bude jedním z měřítek kvality dat.

2.1.2 Chyby v datech

Na základě předchozích definic je již zřejmé, co je myšleno pod pojmem data. V těchto datech je pak možno identifikovat určité nedostatky. Tyto nedostatky, *chyby* je nutno pro naše účely dobře rozlišit – kategorizovat. Určité rozlišení je v této oblasti definováno, ale již nyní je zřejmé, že rozlišení není kompletní ani dostatečně detailní.

Dalším úkolem je tedy toto rozlišení kompletně zmapovat a přidat (nedefinovat) veškeré chybějící *kategorie chyb*.

Jednotné číslo je možno u tohoto pojmu (chybná data) definovat jako *chybný údaj*.

2.1.3 Kvalita dat

Tento pojem přímo souvisí s kategorií chyb v datech. Určitou *kategorii kvality dat* je možno odvodit od míry, v jaké je daná kategorie chyb v datech zastoupena. Dále je možno kvalitu odvodit od účelu dat (tedy jako *kvalitu informace*, která vznikne odvozením z dat). Zde se ovšem jedná o rozdílný pojem – kvalita informace. Zmapování vztahu takto chápané kvality dat a informace bude důležitým úkolem dalšího výzkumu. Vztah je dosti složitý, přičemž z velké části bude záviset na kategorii kvality dat a na typu informace.

Jednotné číslo je možno u kvality dat definovat jako *kvalita údaje*. Kvalita informace je chápána jako jednotné číslo, pokud bychom měli hodnotit *kvalitu více informací*, hodnotili bychom již například *kvalitu studie*. Takto vysokou úroveň hodnocení kvality se tedy nebudeme dále zabývat.

Pokud kvalitu dat zatím neznáme, můžeme jednotlivým údajům před započítáním dalších akcí nastavit výchozí hodnotu, kterou definujeme jako *implicitní kvalitu*.

2.1.4 Akce nad daty

Nejprve je nutno stanovit kvalitu jednotlivých údajů. Stanovení kvality je prvotní *operace nad daty*, kdy údajům s neznámou kvalitou je přiřazena kvalita na základě zvolené metriky.

Aby bylo možno zlepšit kvalitu dat, je nutné v datech provést nějaké změny nebo je nutné kvalitu *ověřit* a tím ji *zvýšit*. Přímou s jednotlivými údaji je tedy možno provádět tři základní akce: *nahrazení*, *vyřazení* a *doplnění*. Ověření (které ručně provede pověřená osoba) pak při nezměněných údajích může zvýšit (ale i potvrdit nebo snížit) jejich kvalitu.

Nahrazení, zneplatnění a doplnění se provede v případě, že provedená akce slibuje zvýšení kvality dat. Ověření se provede v ideálním případě vždy, ale protože tato operace je časově a zdrojově náročná, je nutno pro údaje stanovit kritéria, za kterých se ověření bude vyžadovat, a tím snížit množství provedených ověření na únosnou mez.

3 Způsoby reprezentace dat s ohledem na možnosti kvantifikace jejich kvality

Vstupní data jsou reprezentována ve svém vlastním formátu. Tento formát sám o sobě zřídka umožňuje ohodnocení kvality jednotlivých údajů, proto je nutné formát rozšířit o některé další atributy. Nové atributy by měly být schopny pojmut hodnocení všech kategorií chyb.

Další otázkou je, jakou formou se hodnocení bude provádět. Nejjednodušším způsobem je ohodnocení každého údaje hodnotou **true** nebo **false**. Takový způsob může mít v určitých případech své opodstatnění, ale současně se ztrácí případný potenciál chybného údaje. Lze totiž předpokládat, že údaj je sice chybný, ale přesto je poměrně blízký správné hodnotě (pokud bude vyhodnocován například údaj z nepřesného monitorovacího zařízení, víme, že je chybný, ale je možné jej brát jako použitelný v určité toleranci).

Nabízí se tedy další známé možnosti, jak data reprezentovat, přičemž všechny jsou silnější než použití hodnot **true**, **false** (které je pouze jejich speciálním případem):

- pomocí intervalu, ve kterém je údaj platný,
- pomocí fuzzy množin,
- pomocí pravděpodobnosti, uváděné u každého údaje,
- pomocí kombinace výše zmíněných přístupů.

Intervalová aritmetika se používá k zachycení chyb, které vznikají buď vzhledem k nepřesnostem v měření nebo vzhledem k existenci několika alternativních postupů k odhadu parametrů (viz [2]). Nevýhodou je, že pokud jsou známy pravděpodobnostní distribuce vstupů, aplikací intervalové aritmetiky přicházíme o tyto informace.

Teorie fuzzy množin je metodou, která pokrývá analýzu nepřesností takových systémů, kde nepřesnosti vznikají spíše díky jisté vágnosti, nepřesnosti a nejasnosti než díky nahodilosti (viz [3]). Teorie je vhodnější pro kvalitativní rozhodování a klasifikaci údajů do fuzzy množin než pro kvantitativní předpoklady. Takto pojatý formální popis není vhodný pro zachycení chyb při nelineárních strukturálních analýzách.

Pravděpodobnostní analýza se používá zejména v případech, kdy je známá pravděpodobnostní distribuce analyzované veličiny (viz [1]). Nejistota je zde charakterizována jako pravděpodobnosti asociované s událostmi.

Ve chvíli kdy data na vstupu takto reprezentována nejsou (v některých případech mohou být, ale ve většině případů budou data na vstupu bez tohoto rozšíření) bude nutno doplnit hodnoty chybějících atributů. Samotné doplnění těchto hodnot je již určitým zásahem do kvality dat, kdy z implicitní kvality dat (například nulové) se po aplikaci nějaké metriky stane kvalita dat.

4 Metriky kvality dat

Na základě vhodné reprezentace dat je možno přistoupit k měření některé kategorie jejich kvality aplikací příslušné *metriky*. Metriky jsou na reprezentaci přímo závislé, ale jejich závislost by se měla projevovat spíše na typu příslušné chyby. Proto bude nutno dobře zmapovat vztahy mezi reprezentací dat, typem chyby a použitou metrikou.

Samotné mapování metrik by mělo být jedním z hlavních přínosů dalšího výzkumu. Současně se předpokládá vytvoření nových metrik, použitelných obecně nebo v některých speciálních případech (v případových studiích).

5 Možnosti zásahu do dat

Jak již bylo výše zmíněno, nad daty je možno provádět dva druhy operací: buď se provádí přímo *zásah do údaje* (doplnění, nahrazení, vyřazení) nebo se zásah provede do podpůrné struktury, dané reprezentací dat (*ohodnocení kvality*).

Další výzkum by měl metodicky sjednotit tyto akce, stanovit pravidla pro vedení verzí údajů, případně dále podrobněji členit akce typu *ohodnocení kvality* a *nahrazení údaje*.

Tyto možnosti zásahu do dat jsou posledním stavebním kamenem, nutným pro vytvoření algoritmů zásahů do dat.

6 Algoritmy pro některé možnosti zásahu do dat

Algoritmy pro zásahy do dat za účelem zvýšení jejich kvality budou finálním výstupem celého výzkumu. Zde bude nutno definovat (nalézt definice) nové pojmy.

Nejdůležitějším pojmem je zde *model*. Volně řečeno se jedná o funkci, která na základě mnoha různých vstupů vytvoří *hodnotu*, která se pak nějakým způsobem bude dostávat do údajů. Způsob vstupu hodnoty dané modelem do údajů bude dán algoritmy.

Algoritmy budou využívat výstupy metrik a modelu a jejich úkolem bude v datech určit: množiny vyřazených údajů, množiny doplněných údajů a množiny nahrazených údajů spolu s novými hodnotami.

Kromě toho budou hledány algoritmy pro ohodnocení dat, pokud možno nezávislé na algoritmech zásahů do dat (aby nedošlo k „potvrzování sebe sama“).

7 Případové studie

7.1 Odpadové hospodářství

Jako data posloužila databáze hlášení o produkci a nakládání s odpady v Jihomoravském kraji (a související databáze – Registr ekonomických subjektů, demografické údaje atd.). Analyzované informace slouží k vyhodnocení indikátorů odpadového hospodářství.

Zvýšení kvality dat se zde projevilo ve vyšší kvalitě informací (indikátory byly přesněji vyhodnoceny a budou tedy lépe srovnatelné v časových řadách případně s jinými kraji).

7.2 Evidence vozidel

Jako data budou sloužit databáze odcizených vozidel ČR, inzeráty s prodejem automobilů, znalostní weby s informacemi o vozidlech, případně další data z oblasti evidence vozidel. Analyzované informace budou sloužit k poskytnutí podrobnějších statistických výstupů nebo dotazů, které jinak nejsou ve výše zmíněných databázích k dispozici (například výpis odcizených vozidel podle roku výroby, atd.).

Zde bude zvýšení kvality dat využito k propojení všech databází, které by jinak vzhledem k jejich značné nekonzistenci nebylo reálné. Výstupní informace budou tedy informacemi, které by bez využití tohoto postupu nebyly vůbec k dispozici, protože by nemohly obsahovat požadované atributy.

7.3 Evidence nemovitostí

Jako data budou sloužit realitní inzeráty na různých realitních serverech i jiných nesespecializovaných inzertních serverech, katastr nemovitostí, mapové podklady, případně další data z oblasti evidence nemovitostí. Analyzované informace budou sloužit k poskytnutí podrobnějších statistických výstupů nebo dotazů, které jinak nejsou ve výše zmíněných databázích k dispozici (například výpis nemovitostí na základě některých parametrů, rozpoznání stejných nemovitostí v nabídkách různých knacelářů, filtrování reklam stavebních společností, atd.).

Zde bude platit stejný princip jako u předchozí případové studie – díky zvýšené kvalitě dat bude možno propojit databáze a budou k dispozici nové informace.

8 Závěr

Kvalita dat je oblastí, která v současné době výrazně vystupuje do popředí. Současně ale chybí její pevnější teoretické zázemí. Výsledky výzkumu by měly posloužit jako pevný teoretický základ pro případné praktické využití.

Byly vytyčeny prakticky použitelné definice pojmů. Dále budou zmapovány možnosti reprezentace dat a metrik k měření kvality dat a navrženy způsoby vylepšení kvality dat. Tyto způsoby budou v rámci možností hodnoceny tak, aby bylo možno odhadnout jejich úspěšnost. Nevýhodou je ovšem to, že kvalitu dat není v mnoha případech možno přesně posoudit, takže případné kladné výsledky se (ne)projeví až s odstupem času.

Proto se jeví jako vhodné, aby byla teorie co nejvíce experimentálně ověřována. Jinak než na základě určité zpětné vazby totiž většinou nebude možno její závěry potvrdit, případně po přehodnocení doplnit a opravit. Závěrečná část příspěvku je tedy věnována stručnému přehledu realizovaných i připravovaných případových studií.

9 Literatura

- [1] Helton, J.C., Davis, F.J.: Illustration of Sampling-Based Methods for Uncertainty and Sensitivity Analysis. Risk Analysis, 22(3), 2002
- [2] Kearfott, R.B., Kreinovich, V.: Applications of Interval Computations. Kluwer, Boston, 1996.
- [3] Uncertainty in Engineering: <http://www.uncertainty-in-engineering.net/>, (stav k 20.8.2006).

Informační potřeby státní správy, nezávislých organizací, podniků, škol a veřejnosti v oblasti životního prostředí

Jiří Hřebíček, Danka Némethová, Jaroslav Ráček

Masarykova univerzita, Institut biostatistiky a analýz,
Kamenice 126/3, 625 00 Brno

hrebicek@iba.muni.cz, nemethova@iba.muni.cz, racek@iba.muni.cz

Abstrakt

Doposud není plně automatizována ani standardizována komunikace z informačních subsystémů Jednotného informačního systému životního prostředí (JISŽP) České republiky (ČR) s mezinárodními komunikačními systémy v environmentální oblasti ať už z důvodu rozdílné atributace, informačních a komunikačních technologií nebo i metodiky, atd. Proto je od roku 2005 na Masarykově universitě řešen tříletý projekt vědy a výzkumu (VaV) Ministerstva životního prostředí (MŽP) č. SM 10/99/05 Návrh komunikačního rámce s mezinárodními informačními systémy v ŽP [1], [2], který je koordinován odborem informatiky MŽP. Při jeho řešení spolupracuje řešitelský tým s Českou informační agenturou životního prostředí (CENIA). Je zaměřen na efektivní metody této komunikace reflektující potřeby strategického plánování, nadoborové informační podpory státní správy, informování veřejnosti, spolupráce s komerční sférou a mezinárodní reporting. V tomto příspěvku shrneme některé zajímavé výsledky získané v průběhu jeho řešení v letech 2005 až 2006.

Abstract

Communication and environmental information exchange of subsystems of the Corporate Environmental Information System (CEIS) of the Czech Republic are not computerized and standardized from the different reasons like attributes, ICT, methodology, etc. Therefore since 2005, the three years scientific research project SM 10/99/05 A Design of Communication Framework within International Environmental Information Systems of the Ministry of Environment (MoE) of the Czech Republic is solved by the Masaryk University and coordinated by the department of informatics of MoE. The project is solved in collaboration with Czech environmental information agency (CENIA). An Communication Framework will be defined as the overarching set of policies, standards and guidelines, cooperation with businesses and international reporting, which describe the way in which interested parties have agreed, or should agree, to communicate with each other. In this paper are summarized some interested results reached during 2005 – 2006 in solving this projects.

Klíčová slova

JISŽP, komunikační rámec, interoperabilita, environmentální informace, environmentální informační systém.

Keywords

CEIS, communication framework, interoperability, environmental information, environmental information systems.

1 Úvod

Cílem projektu VaV SM 10/99/05 *Návrh komunikačního rámce s mezinárodními informačními systémy v ŽP* je nalezení efektivních metod komunikace českých informačních systémů v rámci JISŽP s mezinárodními komunikačními systémy v environmentální oblasti. Výzkumný projekt je plánován na dobu 36 měsíců, s tím že výzkumné práce byly zahájeny začátkem roku 2005 a je rozčleněn do tří etap [3], [5]:

Identifikace potřeb komunikace - od 1. do 15. měsíce projektu

- Identifikace informačních potřeb veřejné správy: 1.-15. měsíc
- Identifikace informačních potřeb podniků: 1.-15. měsíc

- Identifikace informačních potřeb veřejnosti: 1.-15. měsíc
- Identifikace informačních potřeb škol a vědeckých pracovišť: 1.-15. měsíc
- Vyhodnocení uživatelských potřeb: 16.-18. měsíc

Analýza - od 15. do 27. měsíce projektu

- Analýza IS a jejich ICT: 15.-27. měsíc
- Analýza legislativního a organizačního rámce: 15.-27. měsíc
- Analýza standardizačního rámce: 15.-27. měsíc

Návrh - od 25. do 36. měsíce projektu

- Prvotní návrh komunikačního rámce: 25.-30. měsíc
- Oponentura prvotního návrhu: 31.-33. měsíc
- Návrh komunikačního rámce: 34.-36. měsíc

V první etapě řešení projektu proběhla analytická fáze projektu zahrnující specifikaci dlouhodobého rámce příslušných spoluprací [1], [2]. V této fázi výzkumu byly identifikovány informační potřeby jednotlivých cílových skupin (*veřejné správy, podniků, vědeckovýzkumných a vzdělávacích organizací, veřejnosti a nevládních organizací*) vzhledem ke komunikaci s relevantními informačními systémy v zahraničí. Vycházelo se ze současného stavu a možností daných evropskou legislativou, zejména směrnicí č. 2003/4/EC, *o veřejném přístupu k informacím o životním prostředí* a směrnicí č. 2003/98/EC, *o dalším užití informací veřejného sektoru*. Pozornost byla zaměřena zejména na systémy provozované v rámci EU.

V první polovině roku 2006 proběhla v rámci řešení projektu č. SM 10/99/05 formou dotazníkového průzkumu identifikace informačních potřeb *veřejné správy, podniků, vědeckovýzkumných a vzdělávacích organizací, veřejnosti a nevládních organizací* vzhledem ke komunikaci s relevantními environmentálními informačními systémy v zahraničí.

Byly vytvořeny dva základní typy dotazníků, z nichž jeden sloužil ke zjišťování současného stavu využívání informačních a komunikačních technologií a informačních potřeb v oblasti veřejné správy životního prostředí ze strany občanů ČR a druhý byl použit pro zjištění stavu u organizací působících na území ČR, tj. u podniků, vědeckovýzkumných a vzdělávacích organizací, nevládních organizací a veřejné správy. Celkově bylo elektronickou a papírovou formou distribuováno přes 6200 dotazníků. Z toho odpovědělo 1778 respondentů. Zastoupení jednotlivých skupin bylo mezi zodpovězenými dotazníky následující:

• Nevládní organizace	91
• Veřejná správa	578
• Podniky	389
• Vědeckovýzkumné a vzdělávací instituce	572
• Občané	148

Statistické zpracování a vyhodnocení sesbíraných dotazníků bylo ukončeno v červnu 2006. Výsledky byly transformovány do podoby samostatné zprávy, jejíž závěry jsou shrnuty v tomto příspěvku.

2 Struktura dotazníků

Dotazník pro organizace byl rozdělen do tří základních částí. První část sloužila ke zjišťování stavu obecného využívání Internetu. Jednalo se zejména o otázky zjišťující *míru, způsob a důvody používání Internetu, důvěru v Internet, formáty nejčastěji vyhledávaných informací a používané softwarové nástroje*. Druhá část dotazníku se již specializovala na *informace o životním prostředí*. Otázky z této části se týkaly *způsobu sběru informací o životním prostředí, oblastí v nichž jsou environmentální*

informace vyhledávány, sledovaných faktorů majících vliv na životní prostředí a povědomí o související legislativě. Třetí část zjišťovala znalost zahraničních informačních zdrojů o životním prostředí a zde obsažené otázky se týkaly například Evropské agentury životního prostředí (EEA) agentury nebo některých projektů a iniciativ jako jsou například EIONET, INSPIRE či GMES. K získaným odpovědím byly před vyhodnocením doplněny z databáze další statistické údaje o oslovených organizacích, jako je například velikost, typ, oblast působnosti nebo region.

Dotazník pro občany obsahoval nejdříve *sociodemografickou část*, která sloužila pro rozřazení respondentů a následné porovnání odpovědí dle různých skupin respondentů. Dále dotazník obsahoval část sloužící ke *zjišťování stavu obecného využívání Internetu* a část specializovanou na *informace o životním prostředí*. Obě tyto části se téměř shodovaly s odpovídajícími částmi dotazníku pro organizace. Na rozdíl od dotazníku pro organizace neobsahoval dotazník pro občany část zaměřenou na znalost zahraničních informačních zdrojů o životním prostředí.

3 Využití Internetu

K datu ukončení sběru dotazníků (30.4.2006) bylo k dispozici celkem 1778 dotazníků. Nejvyšší počet dotazníků byl získán z pracovišť státní správy (578; 32,5%) a škol (572; 32,2%), dále z podniků (389; 21,9%) a nevládních organizací (91; 5,1%). Od občanů bylo získáno 148 (8,3%) zodpověděných dotazníků.

Procento subjektů pracujících s Internetem je u pracovišť státní správy, podniků a škol vyrovnané (99,1-99,8%; Tabulka 1). U nevládních organizací je o málo nižší, ovšem rovněž velice vysoké procento organizací pracujících s Internetem (97,8%). Z dotazovaných občanů pracuje s Internetem pouze 79,7% občanů. Více než čtvrtina občanů, kteří používají Internet (29,7%), jej nepovažuje za důvěryhodný zdroj informací. Z dotazovaných subjektů jsou nejméně nedůvěřivé nevládní organizace (pouze 5,6%), nejvíce školy (14,5%; Tabulka 1). Celkem 157 subjektů uvedlo důvod nedůvěry vůči Internetu. Největší výhrady ze stran nevládních organizací, státní správy, podniků i škol byly vůči garanci informací (57,3%), menší vůči bezpečnosti (30,6%) a neobjektivitě (7,6%).

Tabulka 1: Práce s Internetem a nedůvěra vůči Internetu

	Práce s internetem		Nedůvěra vůči internetu	
	Počet	%	Počet	% z pracujících s internetem
nevládní organizace	89	97.8	5	5.6
státní správa	573	99.1	42	7.3
Podniky	387	99.5	31	8.0
Školy	571	99.8	83	14.5
Občané	118	79.7	35	29.7

Celkem 81,6% subjektů (nevládní organizace, státní správa, podniky, školy) vyhledává informace na Internetu zadáváním klíčových slov vyhledávači. Podíl subjektů vyhledávajících informace zadáváním klíčových slov je u nevládních organizací, pracovišť státní správy, podniků i škol podobný (81,6% až 82,8%). Mezi občany vyhledává zadáváním klíčových slov 87,3% respondentů.

Celkem 62,3% subjektů (nevládní org., státní správa, podniky, školy) a 66,1% občanů nepoužívá jiný Internetový prohlížeč a poštovního klienta než Microsoft Internet Explorer a Microsoft Outlook/Express. Kvůli bezpečnosti používá jiný prohlížeč a jiného poštovního klienta 11,4% subjektů a 7,6% občanů, kvůli kvalitě 7,1% subjektů a 14,4% občanů. Z důvodu jiného operačního systému nepoužívá Microsoft Internet Explorer a Outlook/Express pouze 1,7% subjektů, ovšem až 11,0% občanů.

4 Sledování informací o životním prostředí

Celkem 53,7% subjektů (nevládní org., státní správa, podniky, školy) vyhledává informace o životním prostředí. Nejvyšší zájem o informace o životním prostředí byl zjištěn u nevládních organizací (72,5%), nejmenší u škol (37,2%). Mezi občany se o životní prostředí zajímá 82,4% respondentů (Tabulka 2).

Pro získávání informací o životním prostředí využívají různé subjekty zejména Internet (89,7 – 98,5%), dále média (83,8 – 92,5%) a prostřednictvím osobních setkání (41,3 – 69,7%). Občané využívají jako zdroj informací hlavně Internet (82,0%), dále různé média (62,3%) a služby knihoven (42,6%). Osobní setkání jako zdroj informací o životním prostředí využívá pouze 14,8% občanů (Tabulka 2).

Většina subjektů vyhledává na Internetu textové dokumenty (84,8%). Méně jsou vyhledávány plány a mapy (43,9%), obrázky a fotografie (35,7%), resp. tabulky a grafy (31,1%).

Tabulka 2: Vyhledávání informací o životním prostředí

	Vyhledávání informací o ŽP		Zdroj informací o ŽP (% z těch, kteří vyhledávají informace o ŽP)					
	Počet	%	Internet	Média	Osobní setkání	Knihovny	Telefon	Jiné
nevládní organizace	66	72.5	98.5	83.3	69.7	47.0	22.7	12.1
státní správa	357	61.8	96.4	79.3	55.2	14.0	36.1	12.0
Podniky	240	61.7	95.8	79.6	55.0	25.0	28.8	19.6
Školy	213	37.2	89.7	92.5	41.3	33.8	6.6	7.0
Občané	122	82.4	82.0	62.3	14.8	42.6	9.8	6.6

V další části se věnujeme vyhodnocení vyhledávání informací v různých oblastech týkajících se životního prostředí (Tabulka 3).

Co se týče ovzduší, všechny subjekty se zajímají nejvíce o úroveň znečištění ovzduší (52,7%) a o meteorologickou situaci (51,1%). Nejvyšší zájem o tyto oblasti projevují pracoviště státní správy (52,1 a 56,3%). O chemické složení atmosféry se zajímá 20,2 % subjektů, nejvíce nevládní organizace (42,2%). O ozónovou vrstvu se zajímá 12,6% subjektů, nejvíce školy (23,9%).

V oblasti vod se všechny subjekty zajímají nejvíce o kvalitu vody (47,4%), dále o čištění vody (44,1%). O kvalitu vody se nejvíce zajímají školy (53,1%) a nevládní organizace (50,0%), o čištění vody pracoviště státní správy (48,5%) a podniky (46,3%). Nevládní organizace projevují vysoký zájem o spodní a povrchové vody (45,5%), o vodní zdroje (48,5%) a o využívání vodních zdrojů (42,4%).

O úhrnné srážky se nejvíce zajímají pracoviště státní správy (35,6%). V oblasti půdy se všechny subjekty zajímají nejvíce o ochranu půdy (52,4%). Nejvyšší zájem o tuto oblast projevují pracoviště státní správy (62,7%) a nevládní organizace (59,1%). O úroveň znečištění půd se nejvíce zajímají podniky (54,2%) a školy (50,7%), o využití půd nevládní organizace (43,9%).

Všechny subjekty se velice zajímají o chráněné krajinné oblasti a národní parky (60,8%). Největší zájem o CHKO a NP projevují nevládní organizace (až 83,3%). Státní správa, podniky i nevládní organizace se hodně zajímají také o katastrální mapy (50,0% - 70,6%).

Tabulka 3: Vyhledávání informací v různých oblastech týkajících se životního prostředí (v procentech)

		nevládní org.	státní správa	podniky	školy	průměr
OVZDUŠÍ	Úroveň znečištění	51.5	52.1	47.1	60.6	52.7
	meteorologická situace	42.4	56.3	43.3	54.0	51.1
	chemické složení atmosféry	42.4	11.5	25.8	21.6	20.2
	ozónova vrstva	12.1	9.0	7.9	23.9	12.6
VODA	Kvalita vody	50.0	45.1	45.0	53.1	47.4
	Čištění vody	37.9	48.5	46.3	36.2	44.1
	Spodní vody, povrchové vody	45.5	37.8	45.0	17.4	35.4
	vodní zdroje	48.5	34.5	30.4	32.4	33.9
	Úhrnné srážky	18.2	35.6	26.3	26.3	29.5
	využívání vodních zdrojů	42.4	23.5	22.5	27.2	25.6
PŮDA	ochrana půdy	59.1	62.7	39.6	47.4	52.4
	Úroveň znečištění	42.4	31.1	54.2	50.7	43.0
	využití půdy	43.9	31.1	16.7	30.0	27.9
	typy půd	13.6	16.5	19.6	20.2	18.0
KRAJINA	CHKO, NP	83.3	52.7	53.3	76.1	60.8
	katastrální mapy	50.0	70.6	54.2	29.1	54.5
BIODIVER- ZITA	chráněné oblasti	68.2	62.5	45.8	57.3	57.1
	ohrožené druhy	65.2	43.4	24.6	58.2	43.5
	geneticky pozměněné org.	37.9	10.4	26.3	23.9	20.1
ENERGIE	obnovitelné + neobn. zdroje	86.4	62.5	51.3	59.2	60.4
	energetické zdroje	63.6	41.2	40.0	55.9	46.1
	Spotřeba	40.9	21.8	30.4	34.3	28.7
	Výroba	51.5	16.8	37.1	30.5	28.3
	Distribuce	16.7	12.9	18.3	13.6	14.8
DOPRAVA	vliv jedn. druhů dopravy na ŽP	66.7	58.5	49.2	63.4	57.8
	dopravní trendy	56.1	37.8	47.5	42.7	43.0
ODPADY	nakládání s odpadem	72.7	66.1	72.9	62.4	67.6
	recyklace	69.7	57.7	59.6	56.3	58.8
	zpracování odpadu	65.2	54.9	60.4	52.6	56.6
	zneškodňování nebezp. odpadu	43.9	51.8	55.4	45.1	50.6
	sklárky	37.9	47.1	50.4	33.8	44.1
EMISE	Emisní limity	53.0	52.9	58.3	51.2	54.0
	znečišťující látky	51.5	47.3	53.8	55.4	51.4
	Plynné škodliviny	28.8	15.7	23.8	25.4	21.2
	pevné částice	33.3	17.6	22.5	17.8	20.2
	tepelné znečištění	21.2	14.3	12.5	15.5	14.6

Co se týče biodiverzity, všechny subjekty se zajímají nejvíce o chráněné oblasti a výskyt chráněných živočichů (57,1%) a o ohrožené druhy (43,5%). O geneticky pozměněné organizmy se zajímá 20,1% subjektů. Nejvyšší zájem o všechny aspekty biodiverzity projevují nevládní organizace.

Z hlediska energie se všechny subjekty zajímají nejvíce o obnovitelné a neobnovitelné zdroje energie (60,4%), dále o energetické zdroje (46,1% subjektů). Nejvyšší zájem o všechny aspekty energie projevují nevládní organizace, o obnovitelné a neobnovitelné zdroje energie se jich zajímá až 86,4%.

Z hlediska dopravy se všechny subjekty výrazně zajímají o vliv jednotlivých druhů dopravy na ŽP (57,8%). Největší zájem o tento aspekt dopravy projevují nevládní organizace (66,7%). Subjekty se hodně zajímají také o dopravní trendy (43,0%), největší zájem znovu projevují nevládní organizace (56,1%).

Co se týče odpadů, všechny subjekty se zajímají nejvíce o nakládání s odpadem (67,6%), dále o recyklaci (58,8%) a zpracování odpadu (56,6%). Nejvyšší zájem projevují nevládní organizace a podniky.

Z hlediska emisí se všechny subjekty zajímají nejvíce o emisní limity (54,0%) a o znečišťující látky (51,4%). Nejvyšší zájem o emisní limity byl zjištěn u podniků (58,3%), o znečišťující látky se nejvíce zajímají školy (55,4%). O plyné škodliviny, pevné částice a tepelné znečištění projevují nejvyšší zájem nevládní organizace.

5 Znalost zahraničních informačních zdrojů o životním prostředí.

Evropská agentura životního prostředí (EEA) je známá u 16,0 % subjektů. Zná ji až 25,3 % nevládních organizací, ovšem pouze 12,9% škol. Evropská monitorovací a pozorovací síť informací o životním prostředí (EIONET) je známá u 5,0 % subjektů. Nejvyšší znalost EIONETu byla zjištěna u nevládních organizací (8,8%), nejnižší u pracovišť státní správy (2,8%). Evropský projekt INSPIRE (Infrastruktura prostorových informací v Evropě) je známý u 1,8 % subjektů. Podobně jako u sítě EIONET, i projekt INSPIRE je nejznámější u nevládních organizací (5,5%) a nejméně znám u pracovišť státní správy (0,7%). Projekt GMES (Celosvětová iniciativa pro monitorování životního prostředí a bezpečnost) je známý u 9,0 % subjektů. Nejvyšší znalost projektu GMES byla zjištěna u nevládních organizací (11,0%), nejnižší u podniků (6,9%). Projekt GEOSS (Systém celosvětového pozorování země) je známý u 11,5 % subjektů. Zná jej až 15,4% nevládních organizací, ovšem pouze 9,5% pracovišť státní správy.

Tabulka 4: Znalost zahraničních informačních zdrojů o životním prostředí (v procentech)

	nevládní org.	státní správa	podniky	školy
Znalost agentury EEA	25.3	16.3	18.0	12.9
Znalost sítě EIONET	8.8	2.8	7.2	5.1
Znalost projektu INSPIRE	5.5	0.7	2.6	1.7
Znalost projektu GMES	11.0	7.8	6.9	11.4
Znalost projektu GEOSS	15.4	9.5	11.8	12.8

6 Závěr

Statistické vyhodnocení výsledků první etapy řešení projektu SM 10/99/05 ukázalo jaký je současný stav informačních potřeb o životním prostředí v ČR u výše uvedených cílových skupin (nevládní organizace, státní správa, podniky, školy) a občanů. Potvrdilo současné evropské (ale i národní) trendy ve využívání ICT ve veřejné správě (eGovernment) [4], [6]. Řešitelé se těmito technologiemi budou zabývat v další etapě řešení projektu SM 10/99/05 ve vyšší míře, než jaká jim byla v původním návrhu projektu vyčleněna.

7 Literatura

- [1] Hřebíček J., Dušek L., Svoboda B.: Projekt VaV SM/10/99/05. Návrh komunikačního rámce s mezinárodními informačními systémy v ŽP. Roční zpráva o průběhu řešení projektu v roce 2005. CBA LF a PřF, Masarykova univerzita, Brno, 2005.
- [2] Hřebíček J., Ráček J.: Projekt VaV SM/10/99/05. Návrh komunikačního rámce s mezinárodními informačními systémy v ŽP. Průběžná zpráva o řešení projektu v prvním pololetí roku 2006. CBA LF a PřF, Masarykova univerzita, Brno, 2006.
- [3] Hřebíček J.: Environmental Communication Framework in e-Government of the Czech Republic. In proceedings of 19th International Conference Informatics for Environment Protection, EnviroInfo 2005. Masarykova univerzita, Brno, 2005.

- [4] Hřebíček, J., Pitner, T., Ráček, J.: Development of National Environmental Reporting System. In proceedings of 19th International Conference Informatics for Environment Protection, EnviroInfo 2005. Masarykova univerzita, Brno, 2005.
- [5] Hřebíček, J., Zálišová, I.: Interoperability Framework of eGovernment in Environmental Information Exchange of the Czech Republic, Proceedings of eGovenment Days. Praha, 2006.
- [6] Hřebíček, J., Zálišová, I.: Investigation of eParticipation in Environmental Protection and Sustainable Development of the Czech Republic. 2006.

Příspěvek vznikl za podpory grantu Ministerstva životního prostředí České republiky č. SM/10/99/05 „Návrh komunikačního rámce s mezinárodními informačními systémy v ŽP“.

Výpočetní knihovna pro systém hodnocení ekologického stavu povrchových vod

Karel Kisza, Matěj Štefaník

Masarykova univerzita, Institut biostatistiky a analýz,
Divize environmentální informatiky a modelování
Jana Uhra 10, 602 00 Brno
stefanik@iba.muni.cz, kiska@iba.muni.cz

Abstrakt

Projekt ARROW (Assessment and Reference Reports of Water monitoring) je systémovým řešením pro sledování a hodnocení ekologického stavu vod. Odráží v sobě požadavky WFD a současně i poptávku po komplexním řešení problematiky od sledování jednotlivých biologických a hydromorfologických složek až po sběr a hodnocení dat. Nedílnou součástí je také kalibrace a nastavení referenčních modelů pro hodnocení. Projekt navazuje na dlouhodobý vývoj problematiky v České republice a je současně i výsledkem paralelních projektů řešených pod záštitou Ministerstva životního prostředí v rámci přípravy programů monitoringu. V tomto příspěvku je prezentována metodologie, technologie a způsob vývoje výpočetní knihovny, která pomáhá řešit jak kalibraci a hodnocení referenčních lokalit tak také hodnocení lokalit nereferenčních.

Abstract

The ARROW (Assessment and Reference Reports of Water monitoring) project is the project of implementation of WFD in the monitoring of surface waters of the Czech Republic and covers all aspects of this problem from data sampling to informatics solution (focusing on the ecostat). The project incorporates both WFD demands and complex solution of sampling, biological, analytical and ICT problems. The integral part of the project is also the definition and calibration of reference network of sites. The project is based on long term development of this field in the Czech Republic and is related to parallel project of Ministry of the Environment of the Czech Republic connected to monitoring programs. Methodology, technologies and the way of developement of the Computational library which helps to solve calibration and evaluation of reference sites as well as evaluation of new (non-reference) sites are being discussed in this paper.

Klíčová slova

ARROW, XML, Java, hodnocení stavu, kalibrace, hodnotící metriky.

Keywords

ARROW, XML, Java, evaluation, calibration, evaluation metrics.

1 Úvod

Česká republika je povinná v roce 2006 zavést systém sledování a hodnocení ekologického stavu povrchových vod v souladu s požadavky jednotné implementační strategie Rámcové směrnice vodní politiky Evropských společenství (2000/60/EC).

S touto povinností souvisí i potřeba kvalitní metodiky vyhodnocení dat monitoringu a biomonitoringu sledovaných za tímto účelem a tedy faktické naplnění požadovaných výstupů hodnocení kvality povrchových vod.

Problém je komplikován faktem, že implementovaný systém hodnocení musí být schopen hodnotit data sbíraná v souladu s požadavky jednotné implementační strategie Rámcové směrnice vodní politiky Evropských společenství (2000/60/EC).

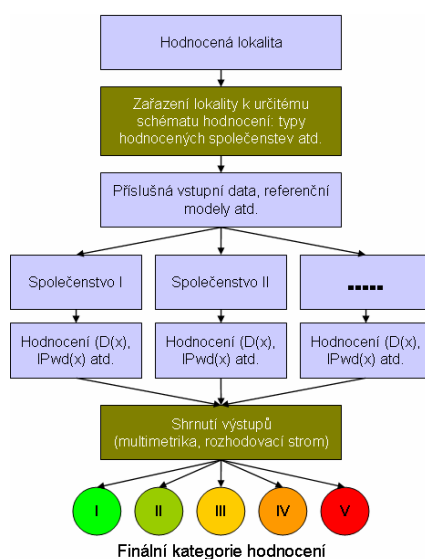
Je téměř jasné, že spojení takto různorodých pohledů na hodnocení stavu povrchových vod bude vyžadovat multimetrický přístup kombinující data z biologických, hydromorfologických a chemicko-fyzikálních pohledů, přičemž každý z těchto pohledů ponese svoji vlastní metodiku hodnocení. Jako nejperspektivnější metodou hodnocení jednotlivých společenstev se jeví srovnání lokalit

biomonitorovací síť s referenčními („přírozenými“) daty a určení míry narušení přírodního stavu lokalit srovnáním skutečného a očekávaného („přírodního“) stavu (jde o princip využívaný např. v rámci systému RIVPACS).

2 Obecné schéma hodnocení

Celý průběh hodnocení lze považovat za velmi komplexní proces, kde vlastní matematické metody tvoří pouze menší část celého problému. Vlastnímu zahájení analýzy musí předcházet definice „schémat hodnocení lokalit“, které respektuje vhodnost nebo vůbec využitelnost různých typů společenstev pro různé typy toků.

Teprve po této fázi, kdy je pro hodnocenou lokalitu určena sada biologických společenstev a dalších dat, je možné přistoupit k vlastní analýze dat. Analýza probíhá pro různá biologická společenstva a typy dat odděleně a po jejím skončení jsou dílčí výsledky různých částí analýzy shrnuty do formy jednoduchého kategoriálního hodnocení stavu, viz. následující obrázek:



Obr. 1: Kategorie Hodnocení

3 Technická realizace

Řešitelský tým navrhnul vyvíjet hodnoticí systém formou samostatné výpočetní knihovny. Hodnoticí systém je implementován v jazyce Java a s databází bude komunikuje skrze standardizované XML rozhraní, které je v podstatě obrazem databázových tabulek, které obsahují data nezbytná pro výpočet. Tímto řešením datové základny odpadá problém se synchronizací zrcadlené databáze a tím i případnou nekonzistencí dat ze strany vývoje.

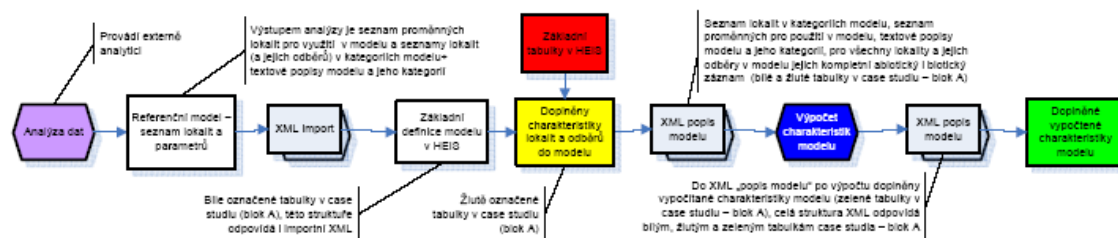
Během testovacího provozu se plánuje provoz knihovny jako webové služby (využívající protokol HTTP) poskytované na straně vývoje a využívané zadavatelem pomocí webového klienta. Po skončení projektu se knihovna bude implementovat na straně zadavatele, kde se bude spouštět lokálně a podle potřeby si zadavatel implementuje vlastní webové služby.

Díky standardizovanému rozhraní mohou s knihovnou v budoucnosti komunikovat prostřednictvím těchto webových služeb i další subjekty, které ji mohou využívat i pro hodnocení dalších lokalit, které nejsou zaznamenány v databázi centrálního IS.

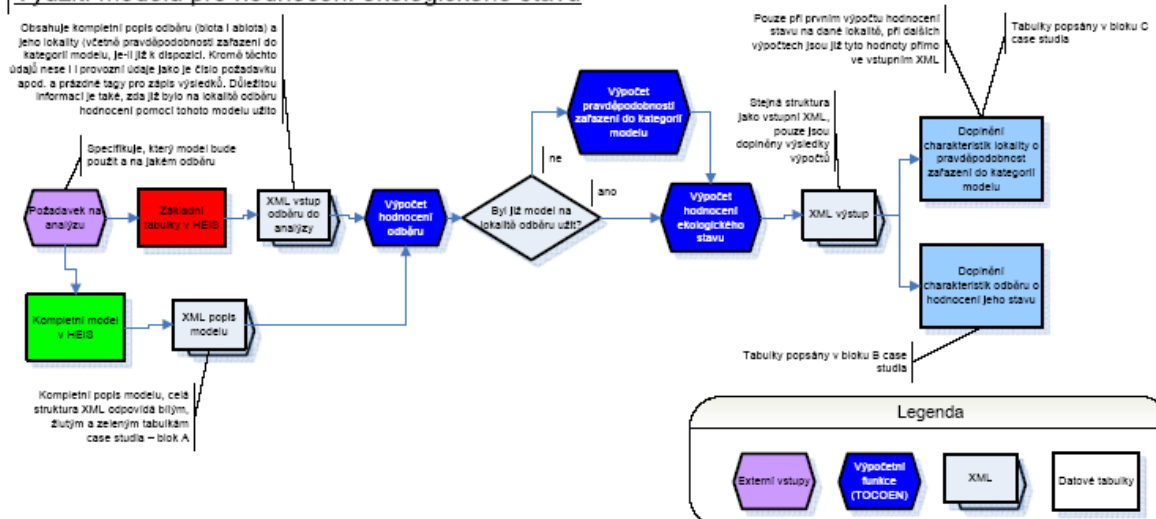
Provedené výpočty budou zasílány zpět centrálnímu IS opět formou XML souborů. Veškeré vstupní i výstupní XML soubory budou zároveň archivovány, aby mohly být v budoucnu použity jako datové podklady pro tvorbu reportů.

Následující obrázek popisuje procesy, které probíhají během výpočtu hodnocení. V první části se jedná o výpočet modelu a ve druhé jde o hodnocení ekologického stavu lokality:

Import modelu do systému a jeho příprava pro analýzu



Využití modelu pro hodnocení ekologického stavu



Obr. 2: Procesy řešené výpočetní knihovnou

4 Metody hodnocení dat biologických společenstev

Pro analýzu vlivu prostředí a pro popis struktury a chování biologických společenstev vznikla celá řada analytických metod sahající od jednoduchých rovnic (indexy) až po komplikované metody statistické analýzy (vícerozměrná analýza dat) vyvíjených z důvodů od čistě vědeckých až po empiricky odvozené a přímo prakticky použitelné výpočty.

Pro praktické nasazení biologických společenstev při hodnocení stavu prostředí je třeba tyto přístupy ovšem vyhodnotit co do:

- Informační hodnoty, kterou přinášejí pro hodnocení stavu povrchových vod
- Uchopitelnost a interpretovatelnost výstupů
- Náročnost na podkladová data
- Výpočetní a implementační náročnost

Lze říci, že celkové řešení hodnocení stavu bude využívat prvků ze všech dvou základních přístupů:

- Indexové hodnocení
- Vícerozměrné hodnocení

Indexové hodnocení

Nejjednodušším přístupem hodnocení je shrnutí biologických společenstev pomocí jednoduchých rovnic do jednoho čísla s určitou biologickou nebo praktickou interpretací. Indexové hodnocení je možno rozdělit na indexy:

- Popisující počet jedinců a/nebo druhů ve společenstvu bez jakékoliv vazby na vlastnosti konkrétních druhů
- Kombinující výskyt nebo abundance druhů s jejich detailnějšími charakteristikami (potravní nebo habitatové preference, saprobní valence atd.)

Mezi indexy první skupiny lze zařadit v podstatě pouze indexy diverzity. Tyto indexy postihují různé aspekty struktury biologických společenstev (počet druhů, poměry početností druhů) a odráží strukturální změny společenstva. Jejichž využitelnost pro komplexní hodnotící systém je spíše pomocná protože jde o poměrně málo citlivé ukazatele nespecifických změn ve společenstvu nerespektující výskyt konkrétních druhů. V metodice hodnocení stavu povrchových vod je počítáno s výpočtem následujících základních indexů diverzity, výpočet je univerzální pro všechny typy biologických společenstev:

- Počet druhů ve společenstvu
- Shannon index a Shannon evenness
- Simpson index
- Berger-Parker index
- Stochastické Q statistics

Vícerozměrné hodnocení

Nejkomplexnější přístup analýzy biologických společenstev využívá nikoliv jednoduchých číselných shrnutí (indexů) za celé společenstvo (tedy převedení na problém jednorozměrné statistiky), ale vícerozměrné analýzy společenstev, kdy každý taxon nebo abiotický parametr představuje jeden rozměr popisující hodnocenou lokalitu. Zároveň nejsou jednotlivé lokality/společenstva hodnoceny odděleně, ale vždy ve vzájemném stavu s ostatními lokalitami/společenstvy (jde vlastně o jejich vzájemnou pozici ve vícerozměrném prostoru) a můžeme tak přesně určit v čem se hodnocená lokalita liší od ostatních (např. znečištěná od neznečištěných). Výhodou přístupu je také využití veškeré variability nasbíraných dat a možnost predikcí na úrovni jednotlivých hodnocených lokalit.

Problematika nasazení XML

Při vývoji aplikace byla třeba rozhodnout, které technologie použít. Nasazení XML jako formátu pro přenos dat bylo odsouhlaseno všemi zúčastněnými již v počátečních fázích vývoje. Hlavním důvodem pro volbu XML byla kvalitní podpora ze strany vývojového prostředí (Java) a také fakt, že pomocí XML lze velmi jednoduše sjednotit strukturu dat. Dalším silným argumentem pro použití XML bylo jeho schopnost omezit variabilitu vstupních dat a tím také výrazně zjednodušit složitost následujícího výpočtu hodnotících metrik a modelu.

Dalším faktorem pro volbu XML formátu bylo vytvoření rozhraní aplikace tak, aby se vývoj modulu mohl rozdělit do dvou částí. První část byla realizována vývojovým týmem IBA, jedná se vývoj výpočetní knihovny. Druhá část je realizována bez naší podpory (obsahem druhého modulu je svázání hodnocených dat se strukturou databáze a správá výsledků spočítaných pomocí výpočetní knihovny). Data přenášená pomocí XML posloužila jako spojující prvek naší spolupráce a pomohla sjednotit a usměrnit vývoj modulu.

V pokročilejších fázích projektu začala velikost zdrojových XML dokumentů výrazně narůstat. (z velké části byl tento jev zapříčiněn používáním dat, která v původním návrhu nebyla respektována a uvažována).

Z tohoto důvodu musel být celkově přehodnocen přístup ke zpracování takto velkých souborů. Vzhledem k času potřebného ke zpracování jednoho dotazu, bylo použití DOM pro vyhledávání v XML již nepraktické nebo spíše nepoužitelné. Jedinou použitelnou alternativou bylo nasazení technologie SAX. Jak se ukázalo, toto rozhraní pro zpracování XML je nesrovnatelně rychlejší než DOM, což nám umožnilo zpracovávat rozsáhle soubory v řádu několika málo sekund.

Nečistota dat

I přes masivní nasazení XML a jeho restriktivních opatření nebylo možné zcela zamezit výskytu chyb ve vstupních datech. Nejčastějšími závadami se překvapivě stávala data, která ve zdrojovém XML souboru měla být obsažena (byla nutná k výpočtu metrik), ale v databázi (která sloužila jako zdroj informací) tato data nebyla uložena. Tyto nedostatky způsobovaly značné problémy, kdy například k vybranému taxonu nebylo možné dohledat některé z jeho vlastností. Tento problém byl ještě umocněn skutečností, že vlastnosti taxonů jsou organizovány do skupin v závislosti na jejich typu. To znamenalo, že v případě, kdy v některé skupině chyběla byť jenom jedna vlastnost, musely se i ostatní vlastnosti z této skupiny ignorovat.

V případech, kdy nepomohla aplikace XML Schema pro jasnou a jednoznačnou definici dat, bylo nutné řešit problém programově. Pro každou metriku, u které mohla nastat situace s nekompletními daty, se provedlo předzpracování počítaných dat a na data byly aplikovány filtry, které tyto neúplná nebo chybná data odstranily. V případě, kdy byla data správná proběhl výpočet standardním způsobem. Z důvodu objektivity spočítaných metrik bylo nezbytné, aby se pro každou funkci zaznamenávalo, kolik taxonů bylo při jejím výpočtu vyřazeno (pro každou funkci mohl být tento počet zcela jiný, protože v každé funkci se zpracovávají jiné vlastnosti taxonů).

Rychlost výpočtu

Pro podpoření rychlosti zpracování dat při výpočtu bylo nezbytné optimalizovat zdrojové kódy. Nejvýraznější prodlevy nastávaly při práci s XML. Naše práce se tedy zaměřila především na metody pracujícími se zdrojovými XML soubory. Nejvýraznějšího zrychlení jsme dosáhli nasazením technologie SAX místo dříve používaného DOM (zkrácení výpočtu z 15 minut na 30 vteřin). Ani toto pro nás nebyl uspokojivý výsledek a tak jsme výrazným způsobem optimalizovali ukládání zpracovávaných dat (aby se stejná data nemusela opakovaně načítat). Jednalo se ve své podstatě o jednoduchou formu cachování. Tímto prostředkem byla rychlost snížena přibližně na 5 vteřin. To je již dostatečně nízká hodnota a z tohoto důvodu již další optimalizace kódu nebyla potřeba.

5 Literatura

- [1] Jarkovský J, Dušek L, Némethová D, Kubošová K, Brabec K, Gelnarová E: Matematické metody hodnocení dat biologických společenstev využitelné pro hodnocení stavu povrchových vod. Masarykova univerzita, Brno, 2005.
- [2] Dokumentace SAX, <http://www.saxproject.org>
- [3] Dokumentace DOM, <http://java.sun.com/>

Příspěvek vznikl za podpory Projektu ARROW (Assessment of Reference Reports of Water Monitoring) Ministerstva životního prostředí ČR a výzkumného záměru INCHEMBIOL (Interakce mezi chemickými látkami, prostředím a biologickými systémy a jejich důsledky na globální, regionální a lokální úrovni, č. MSM0021622412) Ministerstva školství, mládeže a tělovýchovy ČR.

Ontologie v bioinformatice

Lukáš Kokrment

Masarykova univerzita, Institut biostatistiky a analýz,
Kamenice 126/3, 625 00 Brno
kokrment@iba.muni.cz

Abstrakt

V posledních letech jsme mohli být svědky nástupu bioinformatiky jako jednoho z nejdynamičtější se rozvíjejících vědních oborů. Bioinformatika se snaží o spojení jedné z klasických oblastí vědění, biologie, s informatikou, s oborem naopak velice mladým. Charakteristické pro tuto oblast je značné množství dat, které jsou generovány novými typy experimentů a které je nutné nějakým způsobem efektivně ukládat, zpracovávat a získávat z nich nové poznatky. Jedním z nástrojů, který lze využít k popisu biologických a biomedicínských dat jsou ontologie. Jejich původní využití v bioinformatice bylo pro anotaci dat a jejich formalizovaný popis. V průběhu času se však objevily další zajímavé možnosti jejich aplikace.

Abstract

During the last years we could see that bioinformatics became one of the most dynamically evolving fields of science. Bioinformatics connects one of the classical subjects of knowledge (biology) with one the youngest ones – informatics. What is typical for bioinformatics is the growing amount of data generated mainly by new types of biological experiments and the need for effective tools and methods how to save, process and extract new knowledge from this data. One of the tools which can be used for description of biological and biomedical data are ontologies. Original usage of ontologies in bioinformatics was for annotation and formalized description of the data. During the time new interesting applications of ontologies appeared.

Klíčová slova

Bioinformatika, ontologie, gene ontology, strojové učení.

Key words

Bioinformatics, ontologies, gene ontology, machine learning.

1 Úvod

V posledních letech jsme mohli být svědky nástupu bioinformatiky jako jednoho z nejdynamičtější se rozvíjejících vědních oborů. Bioinformatika se snaží o spojení jednoho z klasických vědních oborů, biologie, s informatikou, s oborem naopak velice mladým. Informační technologie umožňují zkoumat a modelovat biologické jevy a procesy, jejichž studium bylo v minulosti téměř nemyslitelné (zejména z důvodů výpočetní náročnosti, komplexnosti, apod.). Mezi výsledky, kterých bylo v posledních letech dosaženo, patří např. rozluštění lidského genomu.

K poznávání biologických procesů slouží množství poměrně nových technologií (např. microarrays [Nguyen et al. 2002]). Všechny tyto technologie a experimenty mají jedno společné – produkují obrovské množství dat, které je potřeba nějakým způsobem ukládat, zpracovávat a získávat z nich nové a relevantní informace a poznatky. Pro ilustraci můžeme uvést např. počet záznamů v databázi DNA sekvencí GenBank, který v roce 2006 přesáhnul 65 biliónů bází a 61 miliónů sekvencí [NCBI 2006].

Množství biologických informací, které jsou v současné době známy, je uloženo v mnoha databázích po celém světě. S počtem databází však vyvstává jeden důležitý problém. Různé databáze mohou používat jednak různou strukturu a uspořádání dat, a jednak mohou používat různé pojmenování či označení jak pro jednotlivé objekty, o nichž uchovávají informace, tak i při popisu těchto objektů či vztahů mezi nimi. My však v ideálním případě chceme transparentně využívat údaje uložené v jednotlivých databázích a využívat je např. pro získávání nových poznatků či při přípravě nových experimentů. Jedním z nástrojů, který by mohl pomoci odstranit tyto problémy jsou ontologie.

2 Ontologie

Pojem ontologie pochází původně z filozofie, kde označuje jednu z nejzákladnějších oblastí této vědní disciplíny. Ve filozofii se ontologií rozumí studium bytí či existence. Ontologie je považována za jednu z nejdůležitějších oblastí metafyziky a hledá nebo postuluje základní kategorie nebo vztahy týkající se bytí a existence a definuje typy a entity v těchto oblastech. Můžeme tedy říci, že ontologie ve filozofickém slova smyslu se zabývá hledáním a zkoumáním, které entity a typy entit vlastně existují.

V oblasti informačních technologií a computer science je ontologie chápána poněkud odlišně. Ontologií zde rozumíme „datový model“, který reprezentuje a popisuje určitou doménu nebo oblast zájmu. Ontologie je využívána k vyvozování závěrů o objektech patřících do dané domény a o vztazích mezi těmito objekty. Takto chápané ontologie se využívají např. v umělé inteligenci, sémantickém webu, softwarovém inženýrství atd. Ontologii zde tedy můžeme považovat za určitou formu reprezentace znalostí o okolním světě případně o jeho části. Představuje pokus o formulaci úplného a přesného konceptuálního schématu v dané doméně. Typicky se jedná o nějakou hierarchickou strukturu, která obsahuje všechny relevantní entity, vztahy mezi nimi a pravidla, která uvnitř dané domény platí [Floridi 2003].

Ontologií můžeme rozumět reprezentaci znalostí o doméně na několika úrovních [Gruber 1993] [Guarino 1995]. Nejjednodušší ontologií je pouze kontrolovaný slovník. Ontologie na této úrovni je vlastně pouze seznam termů, které lze použít k popisu domény respektive znalostí o doméně. Na vyšší úrovni složitosti je tzv. taxonomie, což je množina termů, které jsou uspořádány do určité hierarchie pomocí vztahů generalizace/specializace. Třetí stupeň složitosti představuje objektově orientované databázové schéma, pomocí něhož je definována hierarchie tříd (které mohou obsahovat atributy) a vztahů mezi těmito třídami. Nejvyšší úroveň složitosti v sobě zahrnuje systém reprezentace znalostí založený na logice prvního řádu (jsme pomocí něj schopni vyjádřit všechny předešlé vztahy a navíc i negaci a disjunkci).

Jednou z možností, jak překonat např. výše zmíněné problémy s používáním a porovnáváním biologických dat z různých zdrojů a databází je použití ontologií. Projektem, který si klade za cíl definovat ontologie pro použití k popisu jednotlivých genů a jejich produktů, za jejichž tvorbu jsou zodpovědné (proteiny, RNA) je projekt gene ontology [GO 2006]. Gene ontology zavádí jednotný jazyk pro popis či anotaci genů a jejich produktů, který lze použít na všechny organismy (od kvasinek přes myši až po člověka). Tento jazyk a systém popisu pak umožňuje vědcům odvozovat závěry o velkém množství genů bez nutnosti zkoumání každého genu jednotlivě.

O každém genovém produktu chceme zpravidla znát tři základní informace (nebo druhy informací). Zajímáme se o to, co konkrétní produkt způsobuje – co dělá, konkrétní aktivitu. Dále nás zajímá proč tuto aktivitu provádí a k čemu vede. Třetí typ informace, který chceme znát, bývá, kde tento produkt působí, tedy lokaci. Abychom byli z tohoto ohledu schopni popsat jednotlivé produkty, jsou definovány tři základní ontologie [Ashburner et al. 2000]:

- Molekulární funkce (*molecular function*) – slouží k popisu základní jednoduché úlohy, za kterou zodpovídá, nebo ne které se podílí daný genový produkt (např. vazání daného enzymu).
- Biologický proces (*biological process*) – představuje „širší“ biologický cíl, který je splněn sekvencí jednoduchých molekulárních funkcí (např. buněčné dělení).
- Buněčná komponenta (*cellular component*) – označuje subbuněčné struktury, lokace, makromolekulární komplexy a podobně (např. buněčné jádro).

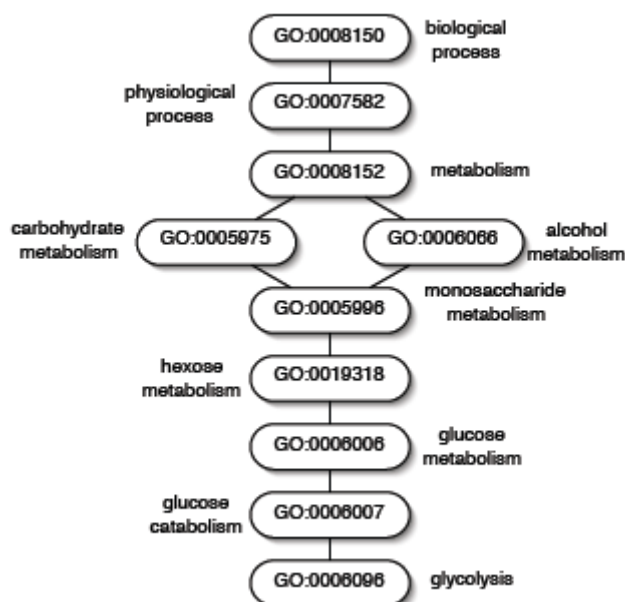
Ontologie mohou být chápány jako matematické grafy, přesněji řečeno orientované acyklické grafy, kde jednotlivé uzly představují koncepty v ontologii. Tyto uzly jsou pak spojeny hranami, které znázorňují vztahy mezi koncepty.

Gene ontology používá dva základní typy vztahů. Prvním typem je vztah či vazba *is-a*, která se dá chápat jako vztah generalizace/specializace. Druhým základním typem je vztah *part-of*, který lze chápat jako vztah celek/část.

Při tvorbě ontologií popisujících geny či jejich produkty pak musí být zachovány následující pravidla:

- Jednoznačnost – jednotlivé termíny mají stejný význam při každém jejich použití.
- Pozitivnost – termíny jako „ne-savec“ apod. nelze použít k označení tříd.
- Objektivnost – termíny jako „neznámý“ neoznačují biologické druhy.
- Jednoduchá dědičnost – žádná třída v hierarchii (grafu) by neměla mít více než jednoho *is-a* předka na bezprostředně vyšší úrovni.
- Jednoduchost definic – termíny použité v definicích jsou jednodušší než termíny definované.
- Základ v realitě – třídy, instance atd. mají svůj základ ve skutečném světě.

Jako příklad hierarchie může sloužit následující obrázek.



Obr. 1: Příklad hierarchie v gene ontology

Je důležité poznamenat, že gene ontology nepředstavuje přesnou a dokonalou formu popisu reality. Tak jak není zcela dokonalá a úplná naše znalost některých biologických procesů, druhů atd. nemůže být zcela přesný ani jejich popis pomocí ontologií. Gene ontology tedy dnes představuje vyvíjející se systém, který se mění na základě nových, aktuálních poznatků.

3 Využití ontologií v bioinformatice

Jak již bylo uvedeno v předchozím textu, jednou z původních motivací pro zavedení a používání ontologií byla potřeba určité standardizace biologických dat uložených v jednotlivých databázích a jejich popisu, potřeba práce s různými datovými zdroji a zpracovávání a porovnávání informací v těchto databázích obsažených. Ontologie tedy slouží jako nástroj, který nám může významně pomoci minimalizovat nekonzistenci a nekoherenci dat. Při používání tohoto nástroje se však objevily další zajímavé a možná na první pohled ne zcela zřejmé možnosti využití.

Gene ontology může být např. použita k ohodnocení podobnosti mezi jednotlivými geny či genovými produkty. V tomto případě můžeme hovořit o jisté formě sémantické podobnosti, která je určována za pomoci ontologií a pozice genů v hierarchii (pro ohodnocení sémantické podobnosti se používá jak

informační obsah, tak i struktura ontologie). Sémantickou podobnost mezi geny lze využít např. při vyhodnocování biologických experimentů a jejich výsledků. Ukazuje se totiž, že v praxi existuje v této oblasti poměrně velká spojitost mezi sémantickou podobností (definovanou pomocí ontologií, tedy jistým způsobem vlastně uměle) a podobností funkční (tedy tím, za co jsou jednotlivé geny či produkty odpovědné v reálném světě [Lord et al. 2003]).

Podobnost mezi jednotlivými genovými produkty lze definovat několika způsoby. Nejjednodušším způsobem, jak ohodnotit podobnost je jednoduché počítání hran mezi uzly v ontologii (reprezentované orientovaným acyklickým grafem). Výhodou této metriky je její jednoduchost a intuitivnost, na druhou stranu nebere v úvahu informaci spojenou s jednotlivými termy či uzly. Tento postup také předpokládá, že ontologie má určitým způsobem rovnoměrnou či homogenní strukturu.

Jako lepší (věrohodněji ohodnocující skutečný vztah, podobnost či vzdálenost mezi jednotlivými genovými produkty) se jeví metriky využívající nějakým způsobem teorii informace. Výhodou těchto složitějších technik je zejména fakt, že jsou schopny se vyrovnat s heterogenitou či nepravidelnostmi v ontologii. Metriky založené na teorii informace většinou pracují s informačním obsahem uzlu (information content), který se mění v závislosti na poloze uzlu v hierarchii (uzly vysoko v hierarchii mají malý informační obsah – představují obecné termy, pojmy, zatímco uzly níže v hierarchii mají informační obsah vyšší – představují konkrétnější termy). Informace, která je sdílená (společná) pro dva uzly v hierarchii je reprezentována společným předchůdcem těchto dvou uzlů. Čím více informace pak dva termy sdílejí, tím jsou si v ontologii podobnější. Podobnost genových produktů g_i , g_j , které jsou popsány množinou termů A_i , A_j , které zahrnují m respektive n termů můžeme ohodnotit např. podle vztahu

$$SIM(g_i, g_j) = \frac{1}{m+n} \times \left(\sum_k \max_p(sim(c_k, c_p)) + \sum_p \max_k(sim(c_k, c_p)) \right),$$

kde $sim(c_k, c_p)$ značí podobnost mezi jednotlivými termy [Azuaje/Wang/Bodenreider 2005]

Výše uvedené porovnání genů na základě jejich sémantické podobnosti lze pak využít v dalších úlohách. Jako příklad můžeme uvést některé typické úlohy z oblasti microarray experimentů. Dvě klasické úlohy při analýze microarray dat jsou clusterování a klasifikace. Při clusterování se používají zpravidla metody strojového učení bez učitele (pro svou činnost používají pouze informace získané ze vstupních objektů, nepoužívají žádné dodatečné informace) a naším cílem je hledat mezi vstupními objekty (v našem případě se může jednat o geny či jednotlivé případy) skupiny objektů, které jsou si nějakým způsobem podobné. Může se jednat např. o podtypy onemocnění apod. Při klasifikaci lze naopak s výhodou použít metod strojového učení s učitelem (tyto techniky se snaží zpravidla na základě dané množiny trénovacích dat odvodit nějakou funkci, kterou pak bude možné použít později při analýze nových případů) a naším cílem může být např. určit pro danou tkáň, jestli je zdravá nebo nemocná (rakovina) apod.

V [Kennedy et al. 2004] lze nalézt jeden z možných způsobů zapojení ontologií do procesu interpretace výsledků clusterování microarray dat. Proces vizualizace a analýzy clusterů má v tomto případě dva vstupy. Prvním je soubor genů vybraných z celkového datového souboru, které jsou nějakým způsobem zajímavé (prošly fází předzpracování – feature selection). Druhým vstupem jsou ontologie. Vybrané geny jsou pak shlukovány do skupin podle podobnosti definované pomocí ontologie. Každému genu, který projde výběrem, jsou přiděleny odpovídající termy ontologie. Mezi geny jsou pak hledány skupiny s podobnou funkcionalitou na základě vzdálenosti jednotlivých termů v ontologii. Biologický popis takto získaných clusterů je pak vytvořen opět na základě termů gene ontology, které jsou obsaženy v jednotlivých clusterech. Funkce vzdálenosti v tomto případě porovnává počet termů, které mají dva objekty společné a celkový počet termů asociovaných se všemi objekty. Do porovnávání se zahrnují také termy, které jsou ve struktuře ontologie výše (jsou obecnější), protože mohou nést pro porovnání nějakou relevantní informaci, jsou však váženy tak, aby do porovnání přispěly menší měrou než termy konkrétnější.

Zatímco pokus o zapojení ontologií do úlohy clusterování např. pro ohodnocení podobnosti mezi jednotlivými geny se jeví jako poměrně logický a intuitivní, poněkud méně zřejmá je již možnost využití ontologií pro úlohu klasifikace.

Možný způsob využití ontologií pro úlohu klasifikace dat popisujících genovou expresi je možné nalézt v [Midelfart 2002]. Ontologie jsou v tomto případě využity pro predikci funkce genů na základě údajů o jejich expresi (opět zpracováváme microarray data). V tomto případě jsou funkční třídy, mezi kterými chceme rozlišovat (do kterých chceme klasifikovat, automaticky zařazovat nové případy), uspořádány do ontologie. Pro generování rozhodovacích pravidel se využívá právě faktu, že množina cílových tříd není „plochá“, ale že tvoří nějakou hierarchickou strukturu. Použitím hierarchické struktury jsme schopni do rozhodovacího procesu zahrnout informaci o nadtřídách a podtřídách (obecnější/specializovanější třída). Výsledná predikce tedy nezávisí pouze na popisu tříd, ale také na jejich nadtřídách a vztazích mezi nadtřídami a podtřídami definovanými v hierarchii. Jednou z výhod tohoto postupu je, že snižuje možnost nesprávné predikce třídy a zvyšuje přesnost klasifikace. Nevýhodou je, že geny náležející do dané třídy jsou nyní pokryty nejen pravidly této třídy, ale i pravidly nadtříd. Konflikt při predikci třídy se pak řeší pomocí metody volby, kde třída dědí hlasy ze svých nadtříd. Výsledný klasifikátor je pak vlastně složený z několika sub-klasifikátorů, které lze využít zvlášť.

Samotný algoritmus najde sadu pravidel pro všechny uzly v ontologii s výjimkou kořene hierarchie. Pro každou třídu jsou geny rozděleny na pozitivní a negativní množinu (geny, které mají být a nemají být pokryty rozhodovacími pravidly). Pozitivní množina tedy obsahuje všechny geny, které jsou pokryty danou třídou nebo její podtřídou, negativní naopak. Každý gen může mít přiřazených více anotací (např. může se podílet na více biologických procesech), algoritmus tedy vytvoří zvláštní pravidlo pro každou možnou třídu. Pravidla jsou generována prohledáváním prostoru hypotéz zdola nahoru.

Predikce třídy nebo tříd, do kterých bude daný gen zařazen je pak provedena pomocí volby. Každá třída s pravidlem, které pokrývá daný gen, obdrží jeden počáteční hlas. Další hlas je pak přidělen za každou třídu, která se nachází na cestě ke kořeni (v případě více možných cest se přidělí nejvyšší možný počet hlasů). Pokud má třída podtřidu, je preferována tato podtřída.

Další možnost využití ontologií ve strojovém učení lze nalézt v [Lottaz/Bentink/Spang 2002]. V tomto případě byly ontologie použity pro podporu predikce diagnózy pacienta opět na základě dat popisujících genovou expresi. Jako základní algoritmus pro predikci třídy byla použita metoda nearest shrunken centroids [Tibshirani et al. 2002]. Tato metoda používá k popisu jednotlivých tříd tzv. centroidy, které určitým způsobem charakterizují úroveň genové exprese v jednotlivých třídách. Tyto centroidy se pak v iterativním procesu zmenšují a posunují k celkovému centroidu (zpřesňujeme klasifikátor). Výsledná klasifikace pak spočívá v nalezení nejbližšího centroidu.

Ontologie je zde použita k popisu množiny vzorků (opět si tedy vytvoříme určitou hierarchii). Pro každý uzel ontologie je pak vytvořen jednoduchý prediktor klasifikace. Každý takovýto prediktor opět kombinuje klasifikátor, který bere v úvahu geny popsané daným uzlem a klasifikátor, který bere v úvahu výstup prediktorů svázaných s potomky tohoto uzlu. Všechny klasifikátory jsou založeny na metodě nearest shrunken centroids. Fáze tréninku i fáze klasifikace pak začíná u listů v hierarchii a postupuje odspoda nahoru.

Výhodou použití ontologií spolu s metodami strojového učení s učitelem (tedy zejména pro úlohu klasifikace) je jednak zpřesnění dosažených výsledků a jednak fakt, že nejsme omezeni na fixní třídy, ale třídy mohou být definovány dynamicky na základě toho, jakou požadujeme přesnost predikce.

4 Závěr

Gene ontology patří do širšího spektra ontologií využívaných v bioinformatice. Jako jakýsi zastřešující projekt můžeme uvést open biomedical ontologies [OBO 2006], který představuje snahu o vytvoření ontologií použitelných v různých biologických a biomedicínských oblastech. Ontologie definované pro jednotlivé oblasti jsou na sebe ortogonální tak, aby se vzájemně nepřekrývaly a bylo umožněno jejich společné využívání.

Ontologie, které se do oblasti bioinformatiky dostaly zejména z důvodu potřeby nějakého standardizovaného nástroje pro (formální) popis velkých množství dat, se kterými je třeba pracovat, se ukazují jako zajímavý prostředek s možnostmi dalšího využití při procesu získávání znalostí.

5 Literatura

- [1] Ashburner, M. et al.: Gene ontology: Tool for the Unification of Biology. *Nature Genetics* 25, 2000.
- [2] Azuaje, F., Wang, H., Bodenreider, O.: Ontology Driven Similarity Approaches to Supporting Gene Functional Assessment. *Proceedings of the ISMB'05*, 2005.
- [3] Azuaje, F. et al.: Gene ontology Driven Similarity for Gene Expression Correlation Analysis. *Proceedings of SOFG'04*, 2004.
- [4] Baclawski, K., Niu, T.: *Ontologies for Bioinformatics*. MIT Press, 2005.
- [5] Batistella, E. et al.: *Bioinformatics: A Growing Field for Ontologies*. Workshop on Ontologies and their Application, 2004.
- [6] Floridi, L. ed.: *Blackwell Guide to the Philosophy of Computing and Information*. Blackwell, Oxford, 2003.
- [7] GO: Gene ontology webpage <http://www.geneontology.org>, 2006.
- [8] Gruber, T. R.: Towards Principles for the Design of Ontologies Used for Knowledge Sharing. *International Journal of Human-Computer Studies*, vol. 43, issue 5-6, 1993.
- [9] Guarino, N.: Formal Ontology, Conceptual Analysis and Knowledge Representation. *International Journal of Human-Computer Studies*, 1995.
- [10] Kennedy, P. J., Simoff, S. J.: CONGO: Clustering on the Gene Ontology. *Proceedings of the 2nd Australasian Data Mining Workshop ADM'03*, 2003.
- [11] Kennedy, P. J. et al.: *Extracting and Explaining Biological Knowledge in Microarray Data*. Springer, Berlin/Heidelberg, 2004.
- [12] Lord, P. et al.: Investigating Semantic Similarity Measures Across the Gene Ontology: The Relationship Between Sequence and Annotation. *Bioinformatics* 19, 2003.
- [13] Lottaz, C., Bentink, S., Spang, R.: Gene Ontology Driven Classification of Gene Expression Profiles. *ECCB'02*, 2002.
- [14] Midelfart, H., Laegreid, A., Komorowski J.: Classification of Gene Expression Data in an Ontology. *ISMDA'02*, 2002.
- [15] NCBI: National Center for Biotechnology Information webpage <http://www.ncbi.nlm.nih.gov>, 2006.
- [16] Nguyen, D. V. et al.: DNA Microarray Experiments: Biological and Technological Aspects. *Biometrics* 58, 2002.
- [17] OBO: Open biomedical ontologies webpage <http://obo.sourceforge.net>, 2006.
- [18] Tibshirani, R., et al.: Multi-class Diagnosis of Cancer Using Shrunk Centroids of Gene Expression. *Proceedings Natl Acad Sci USA* 99, 2002.
- [19] Smith, B., Williams, J., Schulze-Kremer, S.: The Ontology of the Gene Ontology. *Proceedings of AMIA Symposium'03*, 2003.

Optimalizace SOA pomocí cache

Miroslav Kubásek

Masarykova univerzita v Brně, Fakulta informatiky
Botanická 68a, 602 00 Brno
xkubasek@fi.muni.cz

Abstrakt

Service-Oriented Architecture (SOA) je architektura, která umožňuje propojovat rozsáhlé aplikační celky bez ohledu na platformu a jazyk, v kterém jsou tyto jednotlivé části (webové služby) naprogramovány. Pro komunikaci mezi těmito webovými službami (WS) je jako standard používán protokol SOAP. Vzhledem k nárůstu komunikace mezi webovými službami a také vzhledem k tomu, že SOAP je založen na „ukecaném“ standardu XML vzniká velký problém s rychlostí takto používaných webových služeb a také s velkým nárůstem objemu přenesených dat. Ve svém příspěvku představuji cache, jako jedno z možných řešení, které omezí množství přenesených dat a zrychlí odezvu. Pro oblast webových služeb představuji řešení pomocí sémantické cache, popisují její základní části a její ukázkovou implementaci ve webovém portálu SVOD.

Abstract

Service Oriented Architecture (SOA) is an architectural style, which allows interaction of diverse applications regardless of their platform, implementation languages and locations by utilizing generic and reliable services that can be used as application building block. There are several factors that contribute to the performance in web services environment. Network transaction time, time to take to handle the message and time the service itself takes to execute are the main three often factors to culprit in ill performing web services. Current Web caching systems are not able to speed up web services because they are only intended for common web content. I'll introduce the semantic caching as a possible solution for caching SOAP messages. I'll also propose model of the Semantic Transparent Cache for Web Services and introduce implementation in web portal SVOD.

Klíčová slova

Webové Služby, Cache, SOA, SOAP, Výkonnost, XML, Xquery, XPath.

Keywords

Web Services, Cache, SOA, SOAP, Performance, XML, Xquery, XPath.

1 Úvod

Service-Oriented Architecture (SOA) je architektura založená na standardech webových služeb. SOA architekturu můžeme definovat jako systém volně provázaných asynchronně běžících webových služeb. Jejich propojováním získáme možnost tvorby rozsáhlých, komplexních aplikací. Hlavní výhodou takto koncipované architektury je značná flexibilita. Modulárnost řešení přináší i jednoduchou a centralizovanou správu, upgrade a také podporuje integraci heterogenních systémů. To umožňuje například flexibilní přizpůsobení se informačním systémům obchodních partnerů a snížení nákladů na integraci.

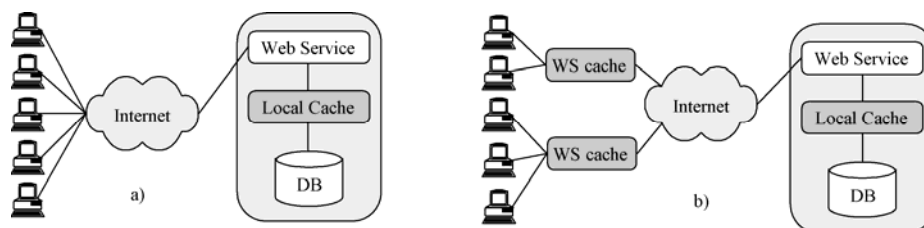
Vzhledem k využívání služeb podnikových informačních systémů (databázových a aplikačních serverů, systémů pro messaging a spolupráci) a aplikovanou modularitu řešení, zmenšuje se postupně rozsah vyvíjených částí produktů a modulů, zkracuje se doba vývoje a zmenšují se vývojářské týmy a pomocí modularity zvýšíme i flexibilitu vlastního systému. Moduly je možné v rámci řešení přeskupovat, případně nahrazovat novějšími. Když nastane potřeba přizpůsobit řešení novým požadavkům, nezasáhne to celou aplikaci, ale jen některé moduly, kterých se tyto změny přímo týkají.

Infrastruktura webových služeb je založena na třech základních technologiích:

- SOAP (Simple Object Access Protocol) – protokol pro posílání zpráv XML a je základem webových služeb. Ostatní standardy jako WSDL a UDDI vznikly až později po uvedení SOAPu a jen dále rozšiřují jeho možnosti a snadnost použití. SOAP umožňuje zaslání XML zprávy mezi dvěma aplikacemi a pracuje tedy na principu peer-to-peer.
- WSDL (Web Services Description Language) – slouží k popisu síťových služeb jako množiny koncových bodů zpracovávajících zprávy. Operace a zprávy jsou popisovány na abstraktní úrovni a teprve poté jsou svázány s konkrétním síťovým protokolem a datovým formátem.
- UDDI (Universal Description, Discovery and Integration) – nabízí mechanismy pro registrování, kategorizování a vyhledávání webových služeb. UDDI funguje jako velký adresář, který obsahuje informace o subjektech (firmách) a jimi poskytovaných službách.

2 Cachování webových služeb

SOA je založena na integraci již existujících systémů a aplikací, ve kterých se cachování běžně využívá. Nejčastější je použití cache při přístupu do databáze (SRBD), dále pak při cachování jednotlivých fragmentů kódů, objektů případně jiných částí aplikace[3]. Tyto řešení můžeme nazvat „lokální cache“, protože je toto dočasné uchovávání objektů řešeno v rámci jednoho systému, aplikace, případně serveru (viz. Obrázek 1).



Obr. 1: Architektura cache pro webové služby

Koncept globální cache vychází z principů cachování obsahu webových stránek. Tento způsob cachování však není pro webové služby úplně vhodný. Je to hlavně z toho důvodu, že webová cache umožňuje ukládat do paměti pouze GET požadavky HTTP protokolu, kdežto webové služby pro svou komunikaci využívají POST.

Webové služby mají mnohem dynamičtější charakter než běžné statické, i dynamicky generované webové stránky. Z tohoto důvodu se jako zajímavé řešení nabízí sémantický způsob cachování webových služeb. Tento princip cachování se již osvědčil právě při cachování odpovědí z databází a také při cachování výstupů při dotazech nad XML databázemi[5]. Vzhledem k tomu, že technologie webových služeb jsou relativně novou záležitostí, výzkum v možnostech cachování je teprve v začátcích a nabízejí se pouze velmi obecné řešení a návrhy možných řešení[6][7][8]. Sémantický způsob cachování webových služeb je nový koncept v této oblasti a bude proto dále podrobněji popsán.

3 Sémantické cachování

Sémantické cachování je technika která se používá velmi hojně v cachování dotazů v databázích. Základní objekt který se cachuje je oblast (region) a celá cache se skládá ze seznamu těchto oblastí. Tyto oblasti jsou pak adresovány ukazatelem (indexem), který v sobě již nese informaci o rozsahu, který uložený region obsahuje. Přístup se pak děje k jednotlivým těmto oblastem a mechanismus pro odstraňování nepoužívaných částí paměti (replacement policy) také pracuje s jednotlivými těmito oblastmi[1].

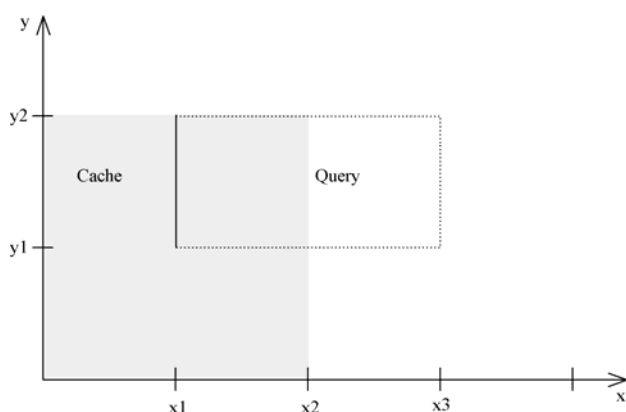
Princip sémantické cache je zřejmý z obrázku 2. Zde máme oblast cache C která představuje oblast, kterou máme uloženou v paměti cache. C můžeme popsat následujícím způsobem:

$$C = (x \geq x_1) \text{ AND } (x < x_2) \text{ AND } (y > y_1) \text{ AND } (y < y_2)$$

Query Q, nebo-li dotaz, pak znázorňuje oblast kterou požaduje uživatel. Tento dotaz pak můžeme rozdělit na dvě části. První je P (probe) nebo-li oblast, o které víme že ji máme uloženou již v cache. Druhá oblast pak je R (request) nebo-li oblast, kterou v cache nemám a musíme ji získat ze zdroje dat (v tomto případě databáze). Tyto oblasti pak jsou popsány následujícím způsobem:

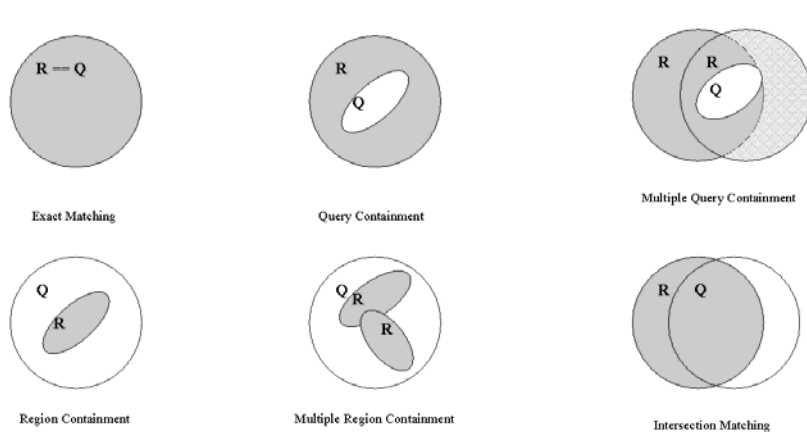
$$P(Q,C) = (x \geq x_1) \text{ AND } (x \leq x_2) \text{ AND } (y > y_1) \text{ AND } (y < y_2)$$

$$R(Q,C) = (x > x_2) \text{ AND } (x < x_3) \text{ AND } (y > y_1) \text{ AND } (y < y_2)$$



Obr. 2: Ukázka sémantické cache

Při sémantickém cachování můžeme narazit na několik kombinací, které se mohou vyskytnout při dotazu, zda-li je požadovaná oblast již uložená v cache či nikoliv (Obrázek 3). První skupinu tvoří případy, kdy oblasti uložené v paměti cache dokáží zcela odpovědět požadovaný dotaz. Nejjednodušší je přímá shoda, kdy dotaz odpovídá přesně oblasti kterou máme v paměti cache. Další možností je, že dotaz je podmnožinou nějaké oblasti uložené v paměti cache, v tomto případě požadovanou podoblast vyfiltrujeme a zašleme zpět jako odpověď. Třetí možností pak je, že požadovaný dotaz dokáže odpovědět více oblastí z paměti cache. V tomto případě vybereme jednu oblast (případně oblasti) které pokrývají dotaz a požadované podoblasti vyfiltrujeme, případně spojíme dohromady a zašleme jako odpověď klientovi.



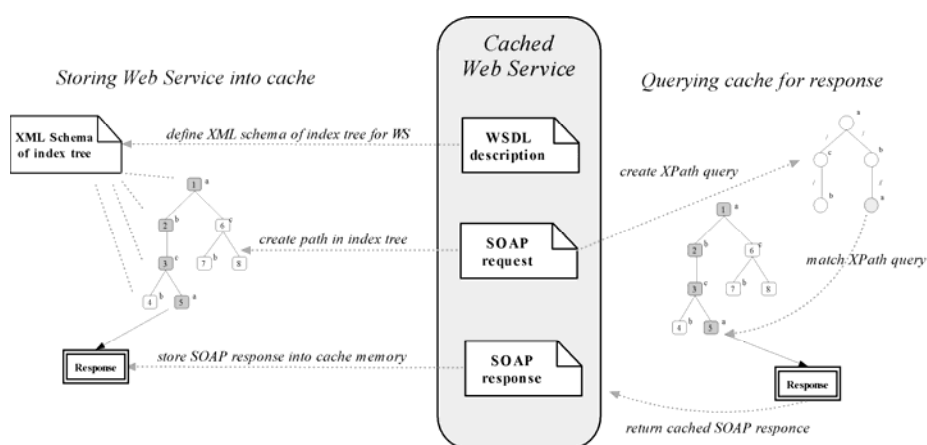
Obr. 3: porovnání oblastí v sémantické cache

Druhou skupinu pak tvoří případy, kdy požadovaný dotaz nemůže být vůbec (nebo ne zcela) zodpovězen z cache a je tedy nutná iterace se zdrojem dat. Záleží na způsobu získávání těchto dat, pokud tyto data nejsou výpočetně náročná, můžeme po zdroji dat požadovat zodpovězení přímo celého dotazu Q a tuto odpověď uložit do cache (a případně odstranit oblasti, které jsou podmnožinou požadovaného dotazu). V druhém případě pak po zdroji dat budeme požadovat pouze chybějící oblast R, tuto pak uložíme do cache a klientovi zašleme odpověď kterou získáme jako složení oblasti R a P.

4 Sémantické cachování webových služeb

V případě cachování webových služeb můžeme paměť cache popsat jako množinu SOAP odpovědí které jsou propojeny se vzory P. Tyto vzory (indexy) si můžeme představit jako „serializaci“ vstupních parametrů volané webové služby (viz obrázek č.4), pro jejichž zápis použijeme „query pattern tree“. Jako příklad nám poslouží dotaz na jednu webovou službu na portálu SVOD[2]:

```
/get_incmor[diag="C50,D05"][stage="I+II"][period_from>=1980][period_to>=1990]
```



Obr. 4: Schema sémantické cache pro webové služby

```

Initialize
 $L \leftarrow 0$ 
for each request to a message  $q$  do
  if  $q$  can be answered from cache then
    if  $q$  is answered from one XML message  $m$  then
      if  $m \notin Q_m$  then
        bring  $m$  into  $Q_m$ 
      end if
       $H(q) \leftarrow L + n * latency(q) * \frac{size(q)}{size(m)}$ 
    else
      for each message  $m_i$  answered  $q$  do
        if  $m_i \notin Q_m$  then
          bring  $m_i$  into  $Q_m$ 
        end if
         $H(m_i) \leftarrow L + n * \frac{size(q)}{size(m_i)}$ 
      end for
    end if
  else
    while there is not enough free space in the cache do
      if  $\frac{size(Q_1)}{size(Q_m)} > P$  then
        remove  $m$  from bottom of  $Q_1$ 
      else
         $L \leftarrow \min(H(m) \mid m \text{ is in the } Q_m)$ 
        remove  $m$  such that  $H(m)$  value
      end if
    end while
    bring  $q$  on top of  $Q_1$ 
  end if
end for

 $Q_1, Q_m$  - cache memory queues
 $P$  - proportion of queues
 $latency(m)$  = downloading latency,  $size(m)$  = message size,  $n$  = reference count

```

Obr. 5: Modifikovaný GDS algoritmus

Tento pattern nám představuje dotaz na webovou službu „get_incmor“ a kde nás zajímají diagnózy „C50“ a „D05“, stádium „I+II“ v letech 1980 až 1990. Na základě znalostí o sémantice této webové služby (tato sémantika je součástí WSDL souboru, který kromě základních informací o webové službě obsahuje také podrobnosti o jednotlivých parametrech) můžeme tento pattern přepsat na jiný, který nám umožní najít i případné částečné shody.

```

/get_incmor[diag="C50" or diag="D05" or diag="C50,D05"][stage="I" or stage="II" or
stage="I+II"][period_from >= 1980][period_to >= 1990]

```

Z tohoto patternu pak již můžeme jednoduše sestavit Xquery dotaz, který nám vrátí seznam indexů, které odkazují na SOAP odpovědi a umožní nám jak částečně, tak i úplně odpovědět dotaz na webovou službu.

```

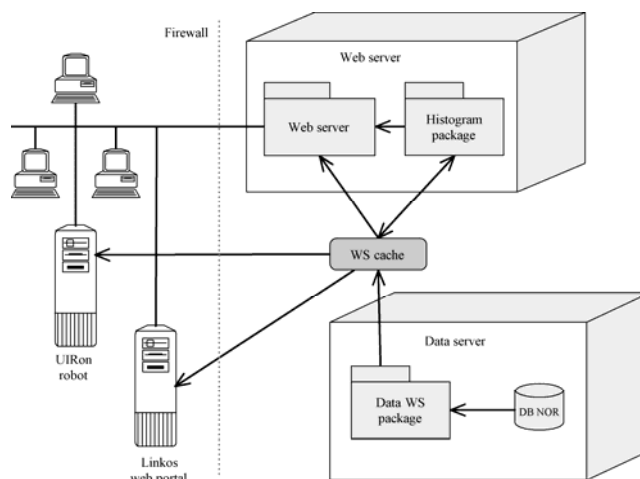
let $a := /cache
for $p in $a/get_incmor
where (
  fn:starts-with($p/diag,"C50") or fn:starts-with($p/diag,"D05") or fn:starts-
  with($p/diag,"C50,D05") )
and ((fn:contains($p/stage,"I") or fn:contains($p/stage,"II") or
  fn:contains($p/stage,"I+II"))
and $p/period_from <= 1980 and $p/period_to >= 1990
)
return <message>{string($p/@id)}</message>

```

Pro umístění SOAP odpovědí do paměti cache byl navržen algoritmus, který je založen na dvou paměťových frontách Q1 a Qm. První je organizována jako jednoduchá fronta FIFO a druhá část paměti je organizována pomocí modifikace algoritmu Greedy Dual Size[4] (viz obrázek. 5)

5 Semantická cache v praxi

Web portál SVOD[2] je systém podporující manažerské rozhodování a vědecké analýzy nad populačními onkologickými daty. Databáze webového portálu SVOD zpřístupňuje data Národního onkologického registru z let 1977–2003, což představuje přes 1 600 000 záznamů. Vzhledem k tomu, že jsou na portálu pomocí technologií webových služeb prováděny složité statistické výpočty, bylo zapotřebí optimalizovat komunikaci s těmito webovými službami. Na obrázku 6. je zakresleno schéma celého portálu z pohledu komunikace mezi webovými službami.

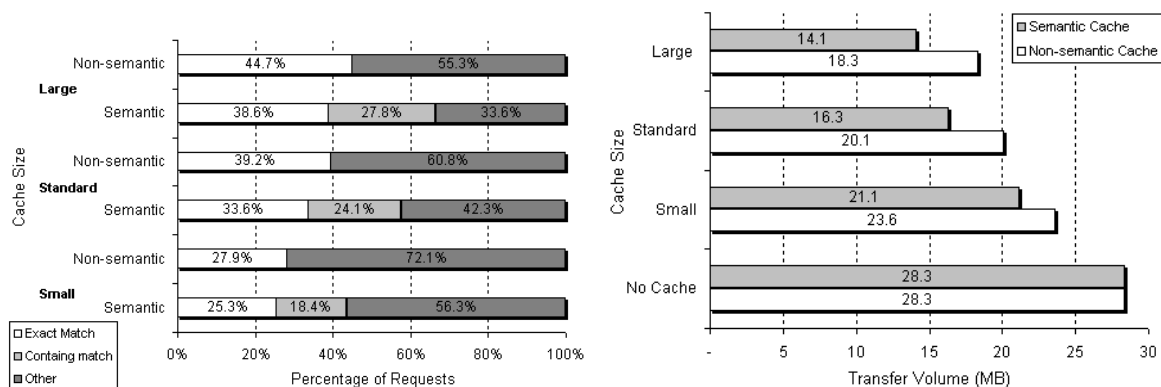


Obr. 6: Webové služby na web portálu SVOD

Sémantická cache byla umístěna na serveru (Pentium4, 3GHz, 1GB, 100 Mbit/s, OS Linux, Apache, MySQL) a implementována pomocí následujících technologií: Java 2 SDK v.1.4.1, XQEngine v. 0.69 jako Xquery a Xpath implementace, Xerces v. 1.4.4 jako XML parser a Axis v. 1.4 jako implementace SOAP protokolu.

6 Závěr

V rámci experimentů byla provedena měření, která prokázala že průměrná úspěšnost cache bez sémantické znalosti je přibližně 40% (počet SOAP odpovědí zaslaných z cache ku celkovému počtu dotazů na webové služby) a při použití sémantického rozšíření této cache je její úspěšnost přibližně 60%. Tato úspěšnost, jak vyplývá z naměřených hodnot, závisí také velmi na velikosti použité cache paměti (viz. Obrázek 7).



Obr. 7: Naměřené hodnoty cache na portálu SVOD

7 Literatura

- [1] Kubásek, M. a Pavlovič, J. a Pitner, T.: , Digital Library for PDA Facilities, In Proceedings of the IADIS International Conference Mobile Learning 2005. IADIS Press, Qawra, 2005.
- [2] Dušek L. et al.: The National Web Portal for Cancer Epidemiology in the Czech Republic. In proceedings of 19th International Conference Informatics for Environment Protection, EnviroInfo 2005. Masarykova univerzita, Brno, 2005.
- [3] Yagoub, K. a Florescu, D. a Issarny, V. a Valduriez, P.: Caching Strategies for Data-Intensive Web Sites. In Proceedings of the International Conference on Very Large Data Bases (VLDB), Cairo, 2000.
- [4] Cao, P. a Irani, S.: Cost-aware WWW proxy caching algorithms. In Proceedings of the USENIX Symposium on Internet Technologies and Systems, 1997.
- [5] Chen, L. a A. Rundensteiner, E. a Wang, S.: XCache - A Semantic Caching System for XML Queries. SIGMOD Conference, 2002.
- [6] Goodman, B.: Accelerate your Web services with caching, IBM Advanced Internet Technology, 2002.
- [7] Tatemura, J. a Hsiung, W. a Li, W.: Acceleration of Web Service Workflow Execution through Edge Computing, In proceedings of the Twelfth International World Wide Web Conference, WWW2003, Hungary, 2003.
- [8] Ramasubramanian, V. a B. Terry, D. a Li, W.: , Caching of XML Web Services for Disconnected Operation, Microsoft Research, USA, 2002.
- [9] Prtál SVOD, <http://www.svod.cz>, 2006.

Metodický rámec stanovení kvality informačního systému pomocí vrstveného modelu

Jan Ministr

VŠB-Technická univerzita v Ostravě, Ekonomická fakulta
Sokolská třída 33, 701 21 Ostrava
jan.ministr@vsb.cz

Abstrakt:

Článek se zabývá stanovením kvality informačního systému na základě dekompozice jeho požadovaných vlastností do hierarchicky uspořádaných vrstev. Počet identifikovaných vrstev celkové kvality informačního systému tvoří tzv. dimenzi kvality. Kvalita jednotlivé vrstvy je stanovena samostatně. Hodnoty kvality jednotlivých vrstev tvoří vektor, který popisuje kvalitu celého informačního systému. Následně jsou hodnoty vektoru agregovány do jediné výsledné hodnoty. Článek popisuje možné způsoby agregace včetně jejich kladů a záporů.

Abstract:

The article deals with the quality of information system determine on the base of its required characteristic decomposed in to hierarchy of sorted layers. The number of identified layers of total quality of information system creates in this way called quality dimension. The quality of individual layer is determined separately. Value of quality the single layers makes the vector which describes total quality of information system. The vector values are consecutively aggregated in to one final value. The article describes possible ways of aggregations including theirs positives and negatives.

Klíčová slova

Vrstva modelu, kvalita softwaru, vektor kvality, dimenze kvality, agregační funkce, COBIT.

Keywords:

Model layers, software quality, quality vector, duality dimension, aggregation function, COBIT.

1 Úvod

Myšlenka dekompozice systému do vrstev není nová. Stejně tak ani není nová myšlenka stanovení kvality na základě více typů kritérií, což lze například vyzorovat v normách řady ISO 9000 nebo v americké metodice MBNQA. Novým je ovšem přístup stanovení kvality informačního systému, který kombinuje oba tyto principy. Dále popsán způsob stanovení kvality informačního systému vychází z disjunktních, hierarchicky uspořádaných vrstev systému, které představují identifikované a dále vyhodnocované vlastnosti systému, u nichž lze určit hranice jednotlivých vrstev. Popsaný postup stanovení kvality informačního systému neodporuje již zavedeným normám.

V roce 2002 prezentoval autor tohoto článku spolu s kolegy z Fakulty informatiky Masarykovy univerzity v Brně na semináři „Informační technologie pro praxi“ příspěvek [3] zabývající se problematikou stanovení výkonnosti administrativních procesů veřejné správy. Proto lze na tenhle článek pohlížet jako na volné pokračování toho, co na semináři zaznělo již v roce 2002. Tentokrát se článek zabývá kvalitou informačního systému, kdy je na kvalitu pohlíženo nikoli z úzkého úhlu vymezeného jednotlivými procesy, ale v širším kontextu celého informačního systému.

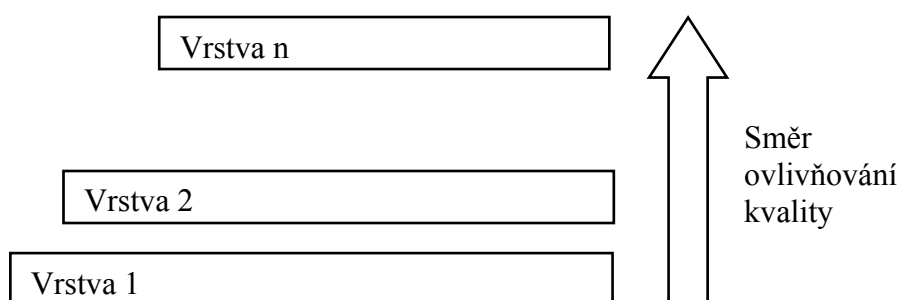
2 Vrstvy modelu kvality informačního systému

Dekompozice informačních systémů do hierarchicky uspořádaných vrstev vede k rozdělení rozsáhlého a složitého informačního systému do jednodušších, logicky uspořádaných částí, za účelem lepšího pochopení fungování informačního systému a následného optimálního využívání jeho vlastností. Jednotlivé části informačního systému jsou uspořádány hierarchicky, takže má smysl hovořit o vrstvách, kde objekty vyšších vrstev využívají služeb, které jim poskytují nižší vrstvy

prostřednictvím nejbližší nižší sousední vrstvy. Obecné členění informačního systému do n vrstev znázorňuje obrázek 1.

Počet vrstev je úměrný požadavkům uživatele na:

- technickou část informačního systému,
- programovou část informačního systému,
- specifické vlastnosti informačního systému (doba odezvy, bezpečnost apod.).



Obr. 1: Obecný vrstvený model kvality informačního systému

V praxi počet vrstev závisí na podmínkách konkrétní organizace, tj. způsobu implementace informačního systému (outsourcing, vlastní, kombinovaná) a specifických požadavcích jeho provozu. Lze předpokládat, že na začátku implementace složitého a rozsáhlého informačního systému bude vyšší počet vrstev v modelu kvality. Naopak po jeho delším provozu se počet sledovaných vrstev sníží. U jednodušších informačních systémů bude docházet ke slučování vrstev což zjednoduší i monitorování jeho kvality.

Nezávisle na počtu vrstev v konkrétním případě, je daný model kvality aplikovatelný na všechny architektury a standardy týkající se stanovení kvality informačních systémů (COBIT a podobně).

3 Dimenze kvality informačního systému

Možným způsobem stanovení kvality celého informačního systému organizace je:

- identifikace jednotlivých vrstev, vlastností, které jsou průběžně monitorovány,
- určení kvality jednotlivých vrstev,
- agregace získaných hodnot jednotlivých vrstev do celkového ukazatele.

Jednotlivé vrstvy jsou vždy posuzovány samostatně bez ohledu na to, jaká je kvalita zbývajících vrstev systému. Principy, techniky a postupy stanovení kvality jednotlivých vrstev se mohou v konkrétních případech značně lišit a jejich popis přesahuje rozsah tohoto příspěvku [3], [7].

V n -vrstvě modelu kvality je pak výsledkem takového postupu n na sobě nezávislých hodnot $q_1, \dots, q_n$, kde hodnota q_i vyjadřuje míru kvality i -té vrstvy systému. Celková kvalita systému je pak charakterizována *vektorem kvality systému*

$$Q = [q_1, \dots, q_n] \quad (1)$$

Hodnota dolní a horní meze, kterou mohou nezávislé hodnoty q_i (míry kvality i -té vrstvy) nabývat, obvykle leží v intervalu $\langle 0, 1 \rangle$, kde znamená:

- **0** minimální kvalitu i -té vrstvy,
- **1** maximální kvalitu i -té vrstvy.

Počet vrstev v modelu kvality, tj. číslo n , pak udává dimenzi tohoto vektoru. V tomto smyslu tedy lze hovořit o *n-dimenzionální kvalitě*.

4 Celková kvalita systému

Jednotlivé vrstvy systému jsou uspořádány tak, že nižší vrstvy poskytují elementární služby vyšším vrstvám prostřednictvím sousední vyšší vrstvy. Kvalita služeb, které poskytují nižší vrstvy systému, se tak přenáší i do kvality služeb vyšších vrstev, jak znázorňuje šipka na obrázku 1. Tato závislost musí být proto zohledněna ve způsobu agregace, kdy se z hodnot složek vektoru kvality systému stanoví výsledná hodnota kvality celého informačního systému.

Další skutečností, kterou je třeba při stanovení celkové kvality systému brát v úvahu, je, zda si jsou z pohledu stanovení kvality systému všechny vrstvy rovnocenné, nebo zda je třeba míru kvality některých vrstev ve výsledku zohlednit více než hodnoty vrstev ostatních.

4.1 Agregční funkce

Za celkovou míru kvality informačního systému nelze prohlásit pouze hodnotu, které dosahuje pouze nejvyšší vrstva (i když její hodnota na nižších vrstvách modelu zpravidla závislá), protože by se do celkového výsledku nepromítly hodnoty kvality nižších vrstev, což by byla chyba. Při konstrukci konkrétní *agregační funkce*, která je pro stanovení výsledné kvality systému použita, by se mělo vždy jednat o funkci n proměnných $q_1, \dots, q_n$, kde n je dimenze kvality a $Q = [q_1, \dots, q_n]$ je celkový vektor kvality informačního systému. Příkladem jednoduché agregační funkce může být součin nebo součet prvků vektoru celkové kvality informačního systému.

V případě součinu má agregační funkce tvar

$$f(Q) = q_1 \cdot \dots \cdot q_n, \quad (2)$$

V případě součtu má agregační funkce tvar

$$f(Q) = q_1 + \dots + q_n. \quad (3)$$

Obě výše uvedené funkce představují dva z nejjednodušších možných způsobů agregace. Jiným možným způsobem je například „geometrická“ velikosti vektoru, tedy druhá odmocnina ze součtu druhých mocnin jednotlivých složek.

$$f(Q) = (q_1^2 + \dots + q_n^2)^{1/2} \quad (4)$$

4.2 Zohlednění priority vrstev

Uživatel informačního systému přiřazuje určitým vrstvám modelu kvality prioritu z pohledu využití a provozu informačního systému. V takovém případě se nabízejí dva základní přístupy, jak tyto priority zohlednit.

První možností je zohlednit význam jednotlivých vrstev kvality informačního systému již při stanovení kvality jednotlivých vrstev, tj. při výpočtu hodnot vektoru kvality i -té vrstvy. Jinak řečeno to znamená, že v případě posuzování kvality vrstev, které se na celkové kvalitě systému podílejí více, jsou použita „sofistikovanější“ měřítka, než při posuzování vrstev, které jsou méně významné pro uživatele. I v tomto případě je zapotřebí zajistit, aby jednotlivé hodnoty složek vektoru kvality nepřekračovaly předem stanovené meze, přičemž tyto meze by měly být pro všechny složky vektoru kvality stejné tedy například již zmiňovaný interval $<0, 1>$.

Druhý přístup spočívá v zavedení vah v rámci výpočtu funkce $f(Q)$, kdy se každá složka vektoru Q násobí nezáporným koeficientem, který vyjadřuje podíl odpovídající vrstvy na celkové kvalitě systému. V případě použití již zmiňovaného součinu, respektive součtu, jako agregační funkce, by byly vztahy (2) a (3) modifikovány do formy

$$f(Q) = a_1 q_1 \cdot \dots \cdot a_n q_n, \quad (5)$$

respektive do tvaru

$$f(Q) = a_1 q_1 + \dots + a_n q_n, \quad (6)$$

kde $a_1, \dots, a_n$ jsou použité váhy představující podíl jednotlivých vrstev na celkové výsledné kvalitě systému.

5 Normování kvality

Použití agregačních funkcí při stanovení celkové kvality informačního systému má následující problémy:

- obory hodnot různých funkcí se mohou vzájemně lišit. Dolní mez oboru hodnot je obvykle u většiny funkcí 0, horní mez je však proměnlivá a bývá jí zpravidla nějaké kladné číslo, v závislosti na požití konkrétní funkce. Většinou se ale požaduje, aby výsledná kvalita nabývala právě hodnot z intervalu $\langle 0, 1 \rangle$, což ale některé funkce nesplňují.
- obory hodnot některých typů funkcí jsou závislé na dimenzi kvality. Příkladem takové funkce je součet (3), kde obor hodnot je vždy interval $\langle 0, n \rangle$. Takovou funkci je pak obtížné použít pro porovnání dvou různých informačních systémů v případě, že tyto jsou dekomponovány do odlišného množství vrstev.

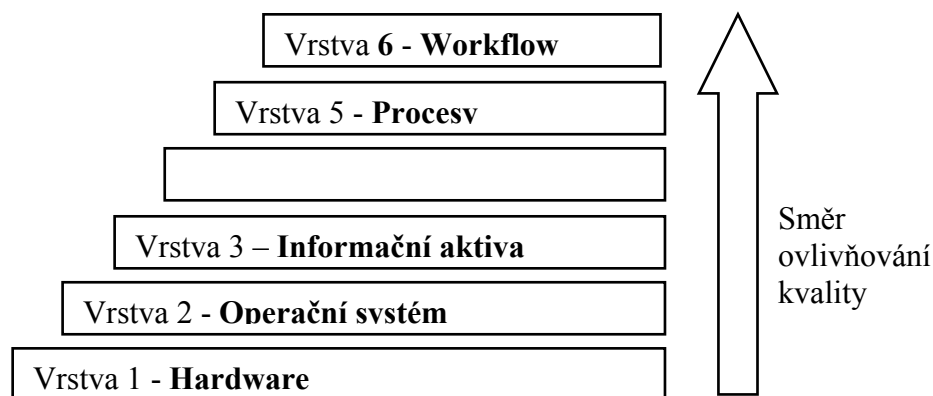
To jsou hlavní důvody, proč je často zapotřebí výsledné hodnoty agregační funkce $f(Q)$ následně normovat, tj. zpravidla násobit vhodnou konstantou, která způsobí, že výsledná hodnota padne právě do intervalu $\langle 0, 1 \rangle$. Mezi funkce, které není zapotřebí dodatečně normovat patří za předpokladu, že všechna q_i jsou z intervalu $\langle 0, 1 \rangle$, součin (2). Ostatní v tomto článku uváděné funkce je zapotřebí normovat. Konkrétně je v případě použití předpisu (3) zapotřebí výsledek vynásobit číslem $1/n$, výsledek vztahu (5) násobit číslem $1/(a_1 \cdot \dots \cdot a_n)$ a výsledek vztahu (6) násobit $1/(a_1 + \dots + a_n)$, což ve všech případech zajistí požadovaný obor hodnot právě v intervalu $\langle 0, 1 \rangle$.

V případě použití agregačních funkcí pro vzájemné srovnávání různých systémů nebo pro porovnání kvality jednoho systému v čase, je zapotřebí používat stále stejnou agregační funkci. Důvodem je skutečnost, že grafy jednotlivých agregačních funkcí, byť jsou tyto normovány na shodný obor hodnot, mají odlišný průběh, což by v případě jejich kombinace vedlo ke zkreslení výsledku.

Pokud se porovnávají dva odlišné systémy, které jsou dekomponovány do různého počtu vrstev, je volba vhodné agregační funkce ještě důležitější. Například zde uvedené funkce (2), resp. (5), jsou polynomy n -tého řádu, kde n je počet vrstev. Použití takovéto funkce pro srovnání systémů s odlišnou dimenzí kvality vede opět ke zkreslení. Oproti tomu funkce (3), resp. (6), jsou bez ohledu na dimenzi stále lineární, tedy pro tyto účely vhodnější.

6 Příklad dekompozice vrstev modelu kvality informačního systému

Možné členění informačního systému do vrstev znázorňuje obrázek 2. V tomto případě je celý systém požadavků na kvalitu informačního systému dekomponován do šesti vrstev. Nejnižší vrstvu tvoří požadavky související hardwarem systému. Druhou vrstvu představují požadavky na operační systém (v případě jednoduchých informačních systémů, lze tyto vrstvy sloučit). Třetí vrstvu představují požadavky na informační aktiva (data a aplikace) používaná v rámci realizace jednotlivých procesů. Čtvrtou vrstvu tvoří požadavky na lidské zdroje (uživatelé a obsluha informačního systému). Pátou vrstvu tvoří požadavky na implementované procesy a nejvyšší, šestou vrstvu tvoří požadavky na workflow systém řízení procesů.



Obr. 2: Šestivrstvý model workflow systému.

Výše popsaný šestivrstvý model představuje základní dekompozici požadavků na kvalitu informačního systému, nicméně v praxi se může počet vrstev lišit na základě konkrétních požadavků uživatelů informačního systému.

7 Závěr

V předchozím textu bylo několikrát uvedeno, že požadované vlastnosti informačního systému mají být dekomponovány do částí (vrstev), které jsou disjunktní a hierarchicky uspořádané. Dodržení disjunktnosti a hierarchického uspořádání není nutnou podmínkou pro následné stanovení kvality pomocí agregační funkce, nicméně funkce musí zohledňovat, v jakém vzájemném vztahu jsou jednotlivé požadavky na kvalitu informačního systému. V případě, že by informační systém nebyl dekomponován do disjunktních a hierarchicky uspořádaných vrstev, vedlo by to v jednotlivých případech ke vzniku odlišných agregačních funkcí, což by znemožňovalo vzájemné srovnávání různých informačních systémů.

Možnosti budoucího vývoje této metody lze spatřovat především ve třech oblastech. První oblastí je pevné stanovení vrstev, tj. požadavků na kvalitu dané části nebo etapy životního cyklu informačního systému, což by mělo být předmětem širší odborné diskuse. Stanovení pevného počtu vrstev, například na základě procesně orientované metody COBIT, by celou situaci značně zjednodušilo.

Druhou oblastí je stanovení metrik, kterými by se určovala kvalita jednotlivých vrstev. Přitom především v případě nižších vrstev, jako je hardwarová vrstva či vrstva operačního systému, lze využít současná měřítka kvality pro oblast ICT. U vyšších vrstev, jako je například vrstva procesů, je nutné vycházet z existujících a vznikajících standardů, případně navrhnout zcela nové, uživateli srozumitelné metriky.

Třetí oblast vývoje této metody vychází z předpokladu, že budou pevně stanoveny jednotlivé vrstvy požadavků na kvalitu informačního systému alespoň pro určité oblasti implementace (veřejná správa, výroba, zdravotnictví, služby apod.) a bude jednotná metodika pro stanovení kvality informačních systémů v těchto oblastech. Potom by měl být vývoj této metody spočívat v identifikaci a požití jediné a srozumitelné „univerzální“ agregační funkce pro stanovení kvality celého informačního systému, čímž by byla celá metodika završena.

8 Literatura

- [1] Fiala, J.; Ministr, J.; Ráček, J.: Optimalizace procesů a architektura systému workflow ve státní správě a samosprávě. In: Systems Integration 2002. The Czech Society for Systems Integration, Praha, 2002.
- [2] Hollingsworth, D.: The Workflow Reference Model, Issue 3.0, Workflow Management Coalition, 1995.

- [3] Ministr, J.; Ráček, J.; Fiala, J.: Měření a zvyšování výkonnosti administrativních procesů veřejné správy. In: Informační technologie pro praxi, Tanger, Ostrava, 2002.
- [4] Fiala, J.; - Ministr, J.: Průvodce analýzou a modelováním procesů, VŠB-TU, Ostrava, 2003.
- [5] Ráček, J.; Ministr, J.; Fiala, J.: Workflow Model and Process Interchange Formats of Public Environmental Administration. In: Information Technologies in Environmental Engineering. ICSC-NAISO Academic Press, Canada, 2003.
- [6] Ráček, J.: Workflow of Environmental Administrative Processes in the Czech Republic. In Environmental Communication in the Information Society. International Society for Environmental Protection, Vienna, 2002.
- [7] Učeň, P. a kolektiv: Metriky v informatice. Grada Publishing, Praha, 2001.
- [8] Krajčík, V.; Ministr, J.: Analýza procesů a řízení projektu veřejného IS – Informační místa pro podnikatele (InMP). Systémová integrace - 04/2005, ČSSI, Praha, 2005.

Příspěvek vznikl za podpory projektu FRVŠ 1757/2006 „Vytvoření předmětu Řízení podnikových procesů“ a výzkumného záměru MŠMT č. MSM0021622418 „Dynamická geovizualizace v krizovém managementu“.

Implementace EMS v národních parcích USA

Petra Mísařová

Mendelova zemědělská a lesnická univerzita v Brně, Provozně ekonomická fakulta
Zemědělská 1, 613 00 Brno
petra.misarova@email.cz

Abstrakt

Příspěvek se zabývá implementací EMS s využitím ICT ve společnosti Xanterra Parks & Resorts, USA. Tato společnost je vlastníkem koncese na provozování největšího počtu národních a státních parků v USA. Společnost poskytuje veškeré služby v těchto parcích. Má dokončenou certifikaci EMS podle ISO 14 001 a zavedený systém environmentálního řízení ve všech parcích, které spravuje.

Abstract

This paper deals with the implementation of EMS using ICT in Xanterra Parks & Resorts company, USA. This company is the largest national and state park concessioner in the United States. Xanterra offer all service in these parks. Xanterra has completed certification of EMS with respect to ISO 14001 and implemented environmental management systems at all national park.

Klíčová slova

EMS, ISO 14 001, ICT, systém environmentálního řízení, recyklace, environmentální cíle, environmentální výsledky společnosti.

Keywords

ISO 14 001, environmental management system, recycling, environmental goals, environmental results of the company.

1 Úvod

Společnost Fred Harvey Company je dnes známá jako Xanterra Parks & Resorts (<http://www.xanterra.com/>). Tato společnost je v soukromém vlastnictví a skládá se ze dvou samostatných společností a to:

- Xanterra Parks & Resorts, Inc.,
- Xanterra South Rim, L.L.C.

Tato společnost je vlastníkem koncese největších státních a národních parků v USA. Koncese na provozování a spravování jednotlivých národních parků získávala společnost postupně. Např. v Grand Canyon South Rim má společnost koncesi platnou až do roku 2020. Xanterra po převzetí koncese pokračuje v nastavených standardech ubytování a restaurací.

Xanterra nabízí široký rozsah rekreačních možností a zážitků. A to od obchodních jednání na golfovém nebo tenisovém hřišti, až po ideálně strávený čas s rodinou formou turistiky, rybaření, kempování nebo jiných rodinných výletů.

Společnost zaměstnává 8 000 zaměstnanců pracujících ve 33 hotelích a jiných ubytovacích zařízeních s více než 5 000 pokoji pro hosty. Dále pracují v 50 vlastních prodejnách, 64 restauracích, 3 přístavech, na 5 golfových hřištích a v kempech o 1 800 kempovacích místech uvnitř národních a státních parků.

Každý rok navštíví národní a státní parky, které Xanterra spravuje, více než 17 milionů lidí. Mezi tyto navštěvované parky patří národní parky jako:

Yellowstone, North and South Rims Grand Canyon, Bryce Canyon, Zion, Crater Lake, Death Valley, Petrified Forest, Everglades a Mount Rushmore National Memorial, ale také zařízení v 7 Ohio State Parks, jeden v New York State Park a jedno středisko v soukromém vlastnictví – Silverado Resort v Napa Valley, California.

2 Implementace EMS podle ISO 14001 s využitím ICT

Marketingové výzkumy a pravděpodobně také styk s veřejností často vytvářel dojem, že společnost nebere dostatečný ohled na životní prostředí, přitom opak byl pravdou. Proto se Xanterra začala iniciativně zajímat o externí osvědčení a certifikace, které by potvrdily skutečnost, že společnost podniká kroky snižující negativní dopady na životní prostředí. Proto se rozhodla zavést systém environmentálního managementu (EMS) podle mezinárodní normy ISO 14 001:2004 s využitím informačních a komunikačních technologií (ICT), které jsou v USA na vysoké úrovni a běžným nástrojem v manažerském řízení společnosti.

Vedení společnosti využilo ICT k vytvoření komplexního EMS zahrnující certifikace podle ISO 14 001:2004 s monitoring a sběrem environmentálních dat systémem „ECOMETRIX“, který má zabudovaný „reporting“ systém pro automatizované vytváření zpráv o environmentálním profilu společnosti, které jsou generované do html formátu na web společnosti (<http://www.xanterra.com/Summary-of-Environmental-Performance-373.html>).

Environmentální primární data jsou sbírána prostřednictvím intranetu a ukládána do databáze Corrective Actions Database (CAD) od firmy Lynk Software (http://www.lynksoftware.com/corrective_action_database.htm), což umožňuje jednoduše provádět interní audit.

Environmentální zprávy na webu zajišťují komplexní obrázek o environmentálně přijatelném chování společnosti Xanterra, která má dokončenou certifikaci podle ISO 14 001:2004 a zavedeny systémy environmentálního řízení ve všech parcích, které spravuje.

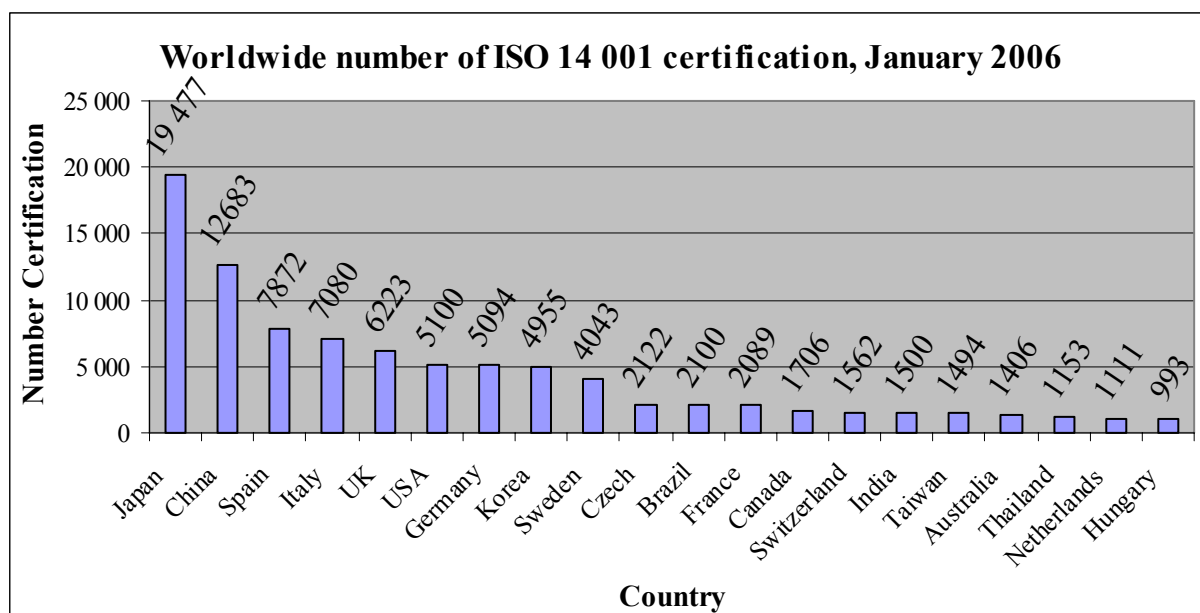
Tato společnost je tisíce (údaj z roku 2005, rok 2006 viz graf) společností v USA, které mají certifikovaný a zavedený EMS podle normy ISO 14 001:2004. Navíc z těchto 1 000 certifikovaných společností jen několik podniká v oblasti cestovním ruchu.

Certifikace EMS podle ISO 14 001 probíhala od roku 2002 v jednotlivých národních parcích v USA následovně:

- Mount Rushmore National Memorial, 2002,
- Zion National Park, 2003,
- Bryce Canyon National Park, 2003,
- North Rim of Grand Canyon National Park, 2003,
- Crater Lake National Park, 2004,
- Death Valley National Park (Furnace Creek Inn & Ranch Resort, Stovepipe Wells a Scotty's Castle), 2004,
- South Rim of Grand Canyon National Park, 2004,
- Yellowstone National Park, 2004,
- Petrified Forest National Park, 2004,
- Everglades National Park, 2005.

3 Srovnání implementace EMS v USA a ve světě

Na obr. 1 vidíme současný stav zavádění EMS dle ISO 14 001 ve světě.



Obr. 1: Počet certifikovaných firem s EMS podle ISO 14 001 (leden 2006)

Zdroj dat: sběr dat - Reinhard Peglau, c/o Federal Environmental Agency, Germany, zpracování - Corporate Risk Mangement Company, Japan

Na obr. 1 již nejsou uvedeny další země, kde je počet certifikovaných firem s EMS dle ISO 14 001 následující :

Finland: 973, Denmark: 837, Rumania: 752, Singapore: 658, Poland: 597, Malaysia: 566, Austria: 550, Belgium: 530, Portugal: 504, Argentina: 454, Norway: 452, Mexico: 422, Iran: 400, South Africa: 393, Hong Kong: 385, Indonesia: 369, Philippines: 361, Egypt: 354, Slovenia: 313, Ireland: 294, Israel: 270, Slovakia: 266, Greece: 254, Turkey: 240, Lithuania: 217, Russian Federation: 185, Colombia: 165, United Arab Emirates: 161, Chile and New Zealand: 155 each (každý), Estonia: 148, Croatia: 130, Latvia: 90, Belarus: 87, Serbia and Montenegro: 70, Vietnam: 63, Ukraine: 60, Peru: 55, Syrian Arab Republic: 53, Bulgaria and Costa Rica: 52 each (každý), Luxembourg: 50, Uruguay: 42, Jordan: 39, Pakistan: 38, Bosnia and Herzegovina: 34, Azerbaijan: 32, Tunisia: 30, Cyprus: 29, Zimbabwe: 28, Morocco: 26, Liechtenstein: 22, Sri Lanka: 21, Venezuela: 20, Saudi Arabia: 17, Ecuador and Macau: 15 each (každý), Bolivia and Ghana: 14 each (každý), Bahrain: 13, Mauritius: 11, Nigeria: 10, Qatar: 9, Kenya and Oman: 8 each (každý), Bangladesh, Kuwait, Lebanon and Trinidad & Tobago: 7 each (každý), Puerto Rico, Honduras, Kazakhstan, Papua New Guinea and Palestine: 6 each (každý), Iceland, Malta, Namibia, Nepal and Swaziland: 5 each (každý), Botswana, Brunei Darussalam, Guyana, Jamaica, Panama, Paraguay, Seychelles, Tanzania and Uganda: 4 each (každý), Algeria, Andorra, Barbados, Cuba, Dominican Republic, El Salvador, Guatemala, Malawi, Monaco, Myanmar, Senegal and Turkmenistan: 3 each (každý), Belize, Cameroon, Fiji, Greenland, Lao, Madagascar, Mozambique, Netherlands Antilles, Niger, San Marino and Zambia: 2 each (každý), Cambodia, FYR of Macedonia, Nicaragua, Saint Lucia, Sudan and Uzbekistan: 1 each (každý).

Jak vidíme z obr. 1 bylo v lednu roku 2006 v USA 5100 certifikovaných firem, což je oproti roku 2005 obrovský posun kupředu. Tento posun je sice nepřehlédnutelný, ale vzhledem k množství podnikatelských subjektů, které v USA podnikají, to není nijak velké číslo. Z obr. 1 vidíme, že na prvním místě je Japonsko s 19 477 certifikacemi a taková malá země jako je Česká republika je na 10 místě s 2122 certifikacemi.

4 Cíle společnosti Xanterra Parks & Resorts v EMS do roku 2015

Každý krok v EMS, v rámci managementu firmy, podporuje ochranu životního prostředí a prevenci proti znečišťování. Celkové chování firmy se zavazuje k neustálému zlepšování životního prostředí. Systém environmentálního řízení ve Xanterra Parks & Resorts se snaží skloubit cíle National Park Service (NPS) s pravidly federální Executive Order a dlouholetými pravidly Xantery v oblasti EMS, kde jsou přesně definovány všechny environmentální aspekty NPS. Ty jsou propojeny s každoročními cíly společnosti a se specifickými cíly společnosti, které jsou uváděny v rozsáhlé zprávě společnosti – „Environmental Vision 2015“.

Všechny dlouhodobé cíle společnosti jsou uvedeny v reportu s názvem „2015 Environmental Vision Goals“, kde environmentální cíle společnosti jsou kvantifikovány :

- Fosilní paliva – snížit využívání fosilních paliv o 30 % (začátek v roce 2000).
- Obnovitelná energie – zvýšit využívání obnovitelné energie tak, aby poskytovala 7 % z celkové spotřeby energie.
- Emise - snížit skleníkové plyny (CO₂), emise o 30 % (začátek v roce 2000).
- Pevný odpad – odvést ze skládek 50 % z celkového množství vyprodukovaného pevného odpadu.
- Ekologický provoz kuchyně – zvýšit nákupy ekologicky udržitelných potravin na 50 % ze všech spotřebovávaných potravin ve společnosti.
- Doprava – dosáhnout v rámci rozlehlé společnosti CAFE (corporate average fuel economy) standartu ve spotřebě paliva 35 mil z galonu (EPA stanovený kombinovaný pro město i mimo město) pro všechny osobní automobily (automobily do 10 míst).
- Nebezpečný odpad – nevytvářet žádný nebezpečný odpad.
- Voda – snížit spotřebu vody o 25 % (začátek v roce 2003).

5 Konkrétní výsledky

Jako výsledek implementace EMS s využitím ICT je možno uvést pro jednotlivé složky životního prostředí:

Emise:

Celkové výše emisí skleníkových plynů nadále pokračuje v klesající tendenci. Během pěti let poklesla výše emisí o 4,75 %. Data byla zkoumána za roky 2000 – 2004. Podrobné informace o důvodech tohoto úspěchu jsou uváděny v reportech společnosti.

Energie:

Ve společnosti jsou samostatně sledovány tyto energie: elektřina, propan, zemní plyn, topný olej, benzín a nafta. Jeden z nejvýznamnějších úspěchů společnosti ve snížení zdrojů je 8,6 procentní snížení elektrické energie. Toto vysoké číslo je důkaz velkého významu a efektivnosti programu, který byl implementován před 5–ti lety. Data jsou dostupná za roky 2000-2004. V roce 2005 a 2006 společnost pokračuje v uplatňování energeticky efektivního programu.

Obnovitelná energie:

Z environmentálního hlediska přináší obrovský prospěch solární, větrná a geotermální energie. Využíváním těchto energií se redukuje smog, kyselý déšť a znečištění ovzduší. Společnosti se podařilo od roku 2001 do roku 2006 zvýšit množství vyráběných kWh z původních 226 538 kWh na dnešních 3 576 085 kWh. Xanterra využívá větrnou, solární i geotermální energii.

Doprava:

Společnost využívá tzv. CAFE (Corporate Average Fuel Economy) k porovnávání údajů v rámci společnosti. V roce 2002, kdy se začal CAFE používat, byl CAFE 18,8 mil z galonu pohonné hmoty. Informace jsou z roku 2004, novější informace nejsou známy. Říkají, že CAFE se zvýšilo na 22,1 mil z galonu. Xantera má v dlouhodobých cílech stanovené číslo 35 mil z galonu. Aby společnost dosáhla tohoto čísla, zvyšuje ve svém vozovém parku automobily Toyota Priuses, které jezdí na kombinovaný pohon a ujedou více než 52 mil z galonu. Je více automobilů značky Toyota, který je schopen ujet více než 40 mil z galonu a těmi také společnost nahrazuje svá stará vozidla. Postupnou obnovou vozového parku společnosti, provozováním autobusové dopravy, která jezdí na propan, a provozováním vozítek na elektrický pohon má společnost dobré vyhlídky ke splnění svého dlouhodobého cíle 35 mil z galonu jako průměr za společnost.

Nakládání s odpady:

Za odpad vytvořený parky se bere vše – obalový odpad, odpad potravin, kovový odpad a další. Pevný odpad, který společnost sleduje, je vytvářen nejen společností Xanterra a jejími zaměstnanci, ale také hosty, kteří parky spravované Xanterrou navštěvují. Za sledované období, kterým jsou opět roky 2000 – 2004 se snížil vyprodukovaný odpad o 17,5 % a recyklace se zvýšila o 129 %. V roce 2004 Xanterra recyklovala 4,28 milionů liber hliníku, skla, plastu, papíru, lepenky, pneumatik, starého železa atd.

Nakládání s nebezpečným odpadem:

Jednou z hlavních priorit managementu v oblasti odpadu je eliminovat všechny nebezpečný odpad. Pokud by to nebylo možné, chce zajistit jeho recyklaci. Pod skupinu nebezpečného odpadu spadá nejen čistě nebezpečný odpad, ale také univerzální odpad jako počítače, tiskárny, telefony, baterie z fotoaparátů, aut a domácností, pesticidy, rtuťové výrobky, zářivky atd. V mnoha sledovaných oblastech došlo k výraznému snížení spotřeby a tvorby nebezpečného odpadu a zvýšilo se množství recyklovaného odpadu v této oblasti.

Voda:

Společnost se snaží o výrazné snížení spotřeby vody. Vzhledem k předmětu podnikání začala Xanterra s výměnou sprch a toalet v pokojích pro hosty. Nové toalety spotřebují pouze 1,6 galonu za spláchnutí a nové sprchy využívají 1,5 – 2 galony za minutu. Kohoutky používané nyní u umyvadel snižují průtok vody o 0,5 až 1 galon za minutu.

6 Závěr

Společnost Xanterra Parks & Resorts je vlastníkem koncese na provozování velkého množství národních i státních parků v USA. Vzhledem k tomu, že společnost podniká na chráněných územích USA, musí svou činnost přizpůsobit ochraně přírody. To znamená, že její chování musí směřovat k ochraně životního prostředí. V tomto směru má společnost o něco složitější pozici než ostatní podniky. Musí se chovat ekologicky ne jen ona sama se svými zaměstnanci, ale musí se snažit ovlivňovat chování hostů a dalších návštěvníků, kteří do národních parků každý rok přijíždějí.

Ze svého 4 měsíčního působení v této společnosti mohu konstatovat, že společnost jako celek důsledně implementuje EMS. Mne samotné se dotklo to, že pravidelně probíhají školení zaměstnanců, které jsem absolvovala. Dále jsou vydány pokyny, jak recyklovat a spotřebovávat jednotlivé zboží, jsou vydány návrhy chování pro hosty, aby pomohli společnosti v ochraně životního prostředí.

Na druhou stranu se nedaří vždy vše dodržovat v EMS. V mnoha provozech společnosti dochází k plýtvání a nerecyklování. Řadoví zaměstnanci nemají zájem sdílet environmentální politiku vrcholového managementu a považují to za práci navíc, kterou nejsou ochotni dělat. Jediným východiskem je důkladná a velmi častá kontrola nadřízených se zaměřením na dodržování environmentálních pravidel.

Z údajů a informací, které mám k dispozici vyplývá, že se společnosti postupně daří splňovat jednotlivé cíle, které si stanovila.

7 Literatura

[1] Sustainability Report 2005. Xanterra Parks & Resorts, Greenwood Village, Colorado, 2005.

[2] Citováno z <http://www.ecology.or.jp/isoworld/english/analy14k.htm>.

Servisně-orientovaná architektura systému na výběr dekontaminační strategie

Jan Pavlovič

Masarykova univerzita v Brně, Fakulta informatiky
Botanická 68a, 602 00 Brno
pavlovic@fi.muni.cz

Abstrakt

Příspěvek se zabývá problematikou návrhu a implementace systému pro podporu rozhodování v problematice výběru dekontaminačních technologií. Popisuje využití nových možností v oblasti servisně-orientovaných architektur pro vědecké výpočty v oblasti podpory rozhodování. Dále procesní modelování celého procesu rozhodování pro výběr dekontaminační strategie. A vizualizaci získaných výsledků.

Abstract

This paper discusses the design and implementation of system for decision support in area of remedial process. It describes new possibilities that bring service-oriented architectures in scientific computations and decision support discipline. The paper discusses process modelling of remedial strategy as well. And the visualization of computed results.

Klíčová slova

dekontaminace, ESB, JBI, podpora rozhodování, SOA, vizualizace.

Keywords

ESB, JBI, decision support, remedial process, SOA, visualization.

1 Motivace

Ještě před několika málo lety nebylo dostatečně známo, jaký může mít znečištění životního prostředí dopad na lidské zdraví a přírodu kolem nás. Spousta odpadu se ocitala volně v přírodě. Výsledkem byly třeba jen v USA tisíce neevidovaných a nezabezpečených vysoce toxických znečištěných území. Stejně tomu bylo i v České republice, kde jsou po odchodu sovětských vojsk mnohá území zamořena dodnes nezjištěnými vysoce toxickými látkami, jejichž dopad na environment není plně znám.

Likvidace takto znečištěných území je komplikovaný, extrémně drahý a energeticky náročný proces. V současné době je v USA evidováno na 500 znečišťujících látek a na druhé straně na 130 technologií, které umožňují jednotlivé kontaminanty z daného území odstranit (Mahutova, Pavlovič 2004).

Zjišťování vhodné kombinace a nastavení technologií pro vyčištění daného území představuje netriviální úlohu. Vše závisí na mnoha parametrech, což v praxi představuje obrovské množství kombinací.

V průběhu úspěšné spolupráce mezi PPRC (Pacific Northwest Pollution Prevention Resource Center) USA, EPA (Environmental Protection Agency) USA a Fakultou informatiky Masarykovy university na environmentálním kalkulátoru (Pavlovič 2004, 2005), jsem byl požádán o navržení teoretického postupu a technického řešení, jak vyřešit otázku multikriteriálního výběru optimálního sledu dekontaminačních technologií. A tím, s využitím moderních poznatků teoretické informatiky, napomoci v rozhodovacím procesu výběru technologií.

2 Úvod do dekontaminační problematiky

Ekologické problémy můžeme rozdělit do několika skupin podle závažnosti a rozsahu. Jsou to přírodní katastrofy, které nastávají s větší nebo menší pravidelností. Patří k nim povodně, zemětřesení nebo lesní požáry. Snad ještě závažnějšími fenomény jsou globální oteplování, způsobené zvýšenými koncentracemi oxidu uhličitého v atmosféře, nebo tvorba ozonové díry. S globálními přírodními problémy se vyrovnáme v průběhu mnoha generací. Pokrokem v tomto ohledu je bezpochyby přijetí

mezinárodních smluv, směrnic, tedy obecně legislativy, jejíž aplikace vede ke zlepšení prostředí, ve kterém žijeme.

Existují i konkrétnější skupiny problémů, které legislativa ošetřuje. Jsou to například ekologické zátěže na izolovaných územích, na kterých jsou zvýšené koncentrace kontaminantů. Cílem je odstranit ekologickou zátěž tím, že se sníží koncentrace nežádoucích látek na přijatelnou míru. K tomu slouží řada nástrojů, metodik a prostředků. Než ale začneme hledat vhodné řešení, musíme přesně popsat stav znečištěného území. Proto si musíme zodpovědět několik zásadních otázek.

- Jak rozsáhlé území je zasaženo?
- Jaký typ média je znečištěn? (např. půda, vodní zdroj, ovzduší)?
- Jaké kontaminanty a v jakých koncentracích se na zkoumaném území nacházejí?
- Jaký cíl (mezí hodnotu koncentrace) si stanovíme pro koncentrace kontaminantů?

Pokud si dokážeme odpovědět na tyto otázky, pak postoupíme k výběru metody, na základě které kontaminanty ze zkoumaného území odstraníme. K těmto metodám patří nápravné dekontaminační technologie. Jedná se o definovaný postup, kterým dojde k nápravě izolovaného území. Každá z technologií je specifická pro odstraňování určitého kontaminantu nebo skupiny kontaminantů a její účinnost se vztahuje k určitému typu média. Při výběru je třeba zodpovědět následující otázky.

- Vyhovuje technologie našemu druhu znečištění?
- Je nám technologie dostupná? (ať už dosažitelností nebo cenou)
- Snížíme pomocí vybrané dekontaminační technologie dostatečně koncentraci kontaminantu? (např. je zbytková koncentrace pod povolenou normou)
- Nezpůsobíme odstraněním zátěže z území další znečištění, které přinese samotné nasazení technologie?

Poslední otázka je zásadní. Nechceme vyřešit problém vybraného území a způsobit tím ještě závažnější zátěž okolnímu prostředí. Proto zkoumáme řadu faktorů, které zjišťují, nakolik je využití dekontaminační technologie náročné (např. energeticky, emisně, atd).

3 Využití principů servisně-orientované architektury pro vědecké výpočty

Velké množství vědeckých „on-demand“ výpočtů má v principu podobnou strukturu a řídí se podobnými pravidly. Na vstupu máme data, která reprezentují informaci, a nich pomocí výpočetního procesu získáme požadovanou znalost. Tento proces většinou vyžaduje značné hardwarové a softwarové zdroje. Z hlediska konceptu integrace jednotlivých výpočetních systémů v rámci našeho prostředí, je vhodné jednotlivé výpočty zapouzdřit další funkcionalitou, které nám integraci usnadní.

Pokud tedy chceme navrhnout vysoce škálovatelný systém, je vhodné využít servisně-orientovanou architekturu (SOA). SOA představuje funkční zapouzdření výpočtů a zjednodušuje komunikační rozhraní mezi aplikací a prezentační logikou. Takovýto typ architektury je možné snadněji integrovat do ostatních softwarových řešení a informačních systémů (Král, Zemlička 2005)

4 Enterprise Service Bus

V průběhu poslední doby se objevily některé zásadní technologické trendy, jakými jsou servisně-orientované architektury (SOA), Enterprise Application Integration (EAI), Business to Business (B2B) a webové služby. Tyto technologie se pokusily popsat výzvy, které přináší integrace business procesů a staly se středem pozornosti průmyslových analytiků i IT odborníků. Právě Enterprise Service Bus (ESB) v sobě kombinuje nejlepší vlastnosti z výše zmíněných principů a soudobých ICT technologií (Chappell 2004).

Princip ESB představuje nový pohled na integraci „loosely coupled“ služeb a vysoce distribuovaných sítí, které přesahují pojetí principů EAI. ESB je standardizovaná integrace platformy pro posílání asynchronních zpráv, webových služeb a transformaci dat. A nad to propojuje a koordinuje integraci velkého množství rozdílných enterprise aplikací s transakční integritou.

ESB přináší nový pohled na spolupráci webových služeb a servisně-orientovaných architektur v robustní celek. ESB umožňuje okamžité použití webových služeb a ostatních integračních technologií spolu s nejmodernějšími technologiemi současné doby (Berlin 2004).

ESB je platformě nezávislý koncept. V ideálním případě je možné ESB implementovat pomocí jakékoli technologie. Na druhou stranu je ESB spjato s technologií Java, jelikož se jedná o široce používanou platformu v mnoha ICT oblastech (Ibarra 2004).

5 Java Business Integration

Java Business Integration (JBI) iniciativa je snaha Java Community Process na vytvoření specifikace, která bude popisovat nezávislý způsob komponentové integrace.

JBI je považováno za „enterprise“ technologii. Hlavním cílem JBI není využití všech možností Java EE specifikace. JBI je obzvláště zaměřeno na interoperabilitu, která umožní jednotlivým tvůrcům komponent komunikovat mezi sebou (Sommers 2005).

JBI definuje integrační kontejner - prostředek, který umožňuje nekompatibilním SOA systémům navzájem komunikovat. JBI vzniklo jako reakce na výrazný nárůst systémů postavených na architektuře SOA, které potřebují navzájem interagovat, přičemž tato interakce je kvůli absenci podobné specifikace často řešena nestandardními způsoby.

JBI je rozšiřitelné pomocí komponent dvou typů: Binding Components (BC) a Service Engines (SE). Jádrem funkcionality integračního kontejneru je vytvořit běhové prostředí, ke kterému se budou moci jednotlivé SOA systémy moci připojit přes nějaké rozhraní a komunikovat s ním po svém. Toto rozhraní pak zprávu zkonvertuje do formátu standardizovaného pro JBI kontejner.

Pokud se k takovému kontejneru připojí více různých SOA systémů, můžou spolu komunikovat, jelikož JBI zkonvertuje příchozí zprávy do svého standardizovaného formátu a odchozí naopak do formátu specifického pro příjemce. Tato rozhraní se nazývají Binding Components (BC). Autorům systémů, které mají být integrovány tedy stačí naimplementovat BC pro svůj systém. S implementacemi JBI jsou typicky dodávány BC pro SOAP/HTTP, JMS

Mimo interakci s okolím může být zpráva v rámci JBI kontejneru dále zpracována. K tomuto účelu zavádí JBI specifikace komponenty typu Service Engine (SE). Příkladem může být XSLT transformace obsahu zprávy, CBR služba nebo právě WS-BPEL „orchestration engine“.

6 Analýza procesu dekontaminační strategie

Pro implementaci celé problematiky je nutné formalizovat proces dekontaminační strategie. Pod tímto pojmem si lze představit proces, do kterého vstupují definiční informace o kontaminaci zvoleného území a požadovaná kritéria a kvalitu dekontaminace daného území. Environmentální odborník si buď zvolí částečné omezení na proces dekontaminace nebo nechá systém, aby mu dekontaminační strategii, dle zadaných kritérií vypočítal. Proces dále obsahuje možnost, jak již existující strategii zlepšit změnou parametrů jednotlivých technologií, které realizují dekontaminaci. Technologie jsou popsány formou balančních rovnic. V průběhu výpočtu strategie jsou použity genetické algoritmy, které jsou schopny z obrovského množství možných kombinací vybrat velice efektivní řešení.

Celkově můžeme na proces dekontaminační strategie nahlížet jako na desetice $P=(A, T, D, U, S, a_s, a_f, \delta_D, \delta_U, \delta_S)$, kde:

A je konečná neprázdná množina činností,

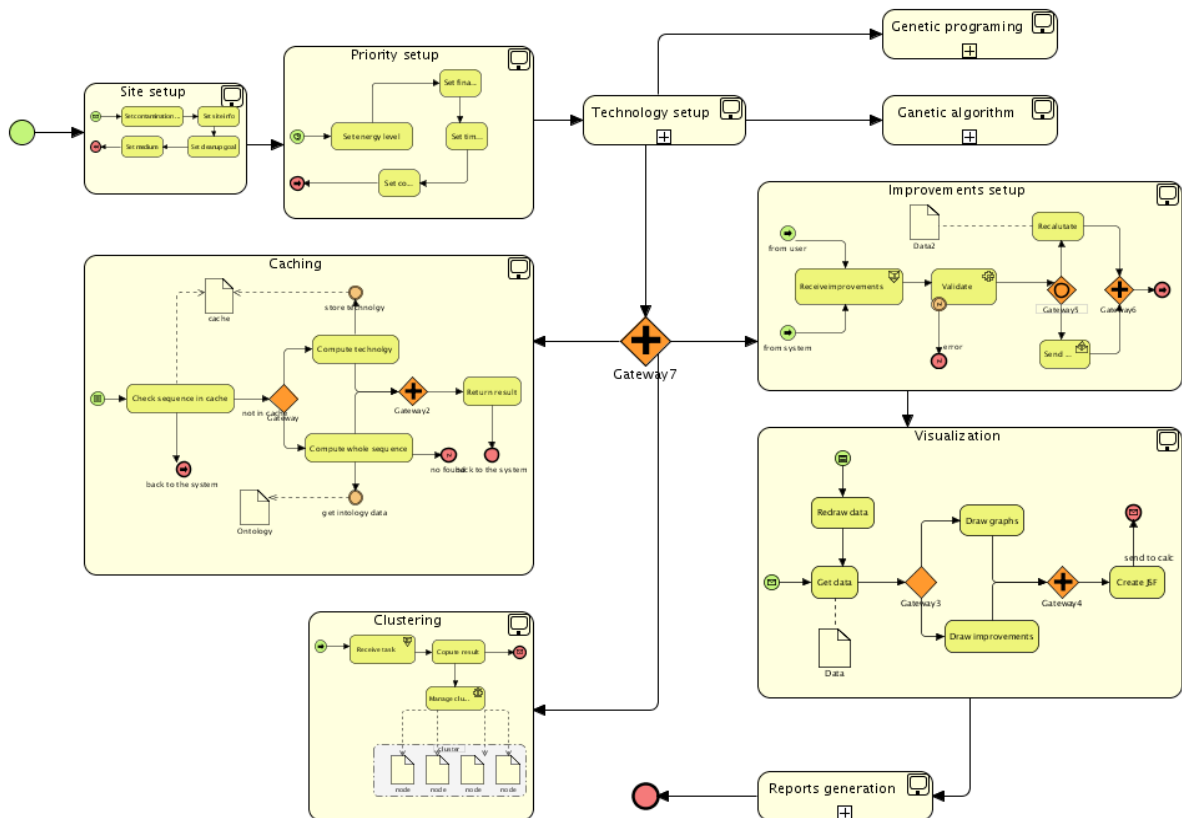
$T \subset A \times A \times C$ je množina přechodů mezi činnostmi a C je množina všech logických výrazů, D je množina dat,

U je množina aktorů,

S je množina zdrojů, kde $U \subseteq S$,
 $a_s \in A$ je počáteční činnost,
 $a_f \in A$ je koncová činnost,
 $\delta_D: A \rightarrow 2^D$, funkce přiřazující každé činnosti příslušná data,
 $\delta_U: A \rightarrow U \times \{\varepsilon\}$, funkce přiřazující každé činnosti aktéra,
 $\delta_S: A \rightarrow 2^S$ funkce přiřazující každé činnosti zdroje.

Podle této definice se dekontaminační proces skládá z činností, přechodů, dat, aktorů a zdrojů. Každý přechod je definován pomocí dvojice činností, kde první představuje činnost, ze které a druhá do které přechod vede. Součástí přechodu je dále logický výraz, jímž je přechod podmíněn. Dále je třeba každé činnosti přiřadit data, aktore a zdroje, jež se smí v jejím průběhu využívat. K tomu slouží funkce δ_D , δ_U , a δ_S .

Dekontaminační proces lze namodelovat v jazyce BPEL (Business Process Expression Language). Výhodou takto namodelovaného systému je možnost poloautomatizovaného převodu do aplikační logiky. Systém je pak možné rychleji implementovat a je také možné i funkcionalitu před vlastní implementací odsimulovat, co přináší značnou úsporu v případě redefinice dané funkcionality.



Obr. 1: Nástin realizace dekontaminační strategie

7 Architektura výpočetního systému

Jelikož je nutné navrženou architekturu celého systému koncipovat tak, aby šla integrovat do již existujících environmentálních systému, a byla dále snadno rozšiřitelná, je nutné vycházet ze základních principů servisně-orientované architektury. Zejména z principů ESB a JBI. Takto je možné systém rozdělit to několika autonomních jednotek a funkčně provázat.

Jádrem systému je ESB, která tvoří komunikační middleware systému. ESB dokáže realizovat i procesní logiku, v tomto případě je ESB použita jako rozhodovací jednotka, která přeposílá požadavky

na výpočet cachovací jednotce a tím snižuje celkovou režii systému. Na ESB jsou napojeny jednotlivé JBI komponenty. Jako implementace ESB specifikace je použita OpenESB (<https://open-esb.dev.java.net/>), která běží na Java EE 5 aplikačním serveru Glassfish 9.0.

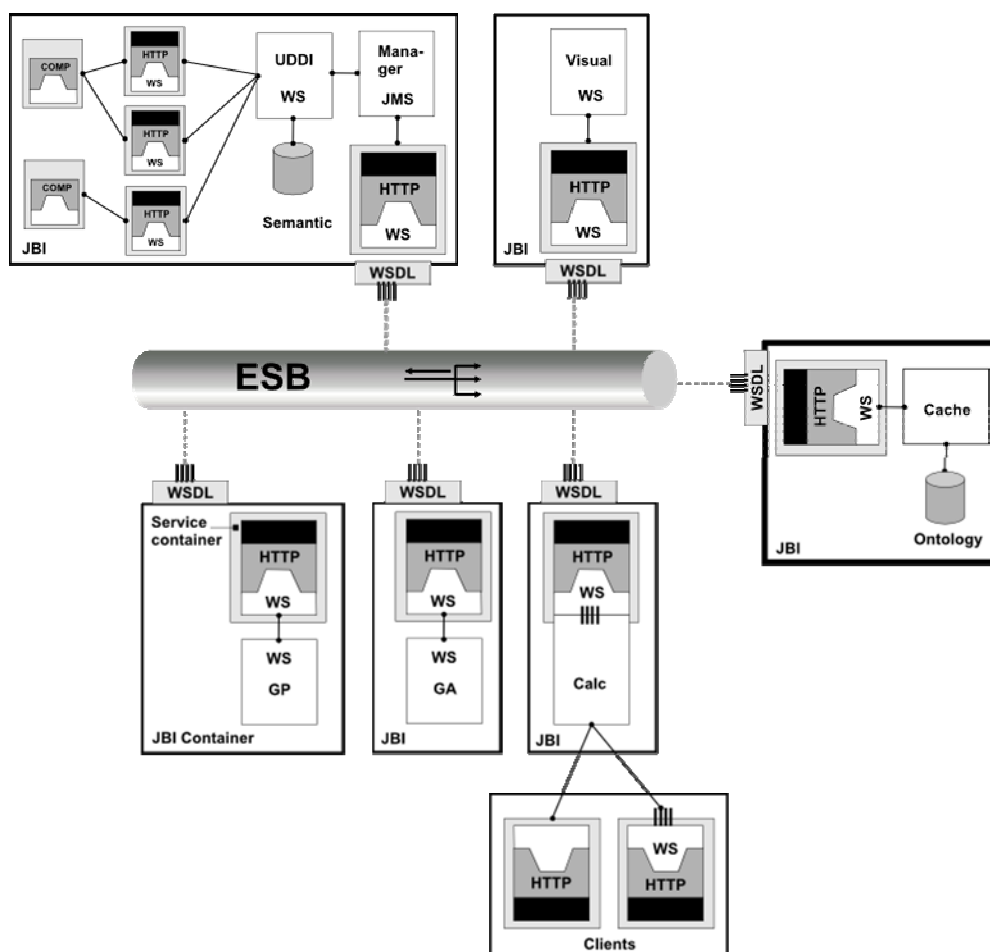
Vstupní informace od uživatele zpracovává JBI komponenta kalkulátoru, kde je implementována základní logika celého systému. Jelikož je funkčně tato komponenta zapouzdřena, je možné komunikaci z celým systémem provádět buď pomocí přímého vizuálního rozhraní (založením na Java Server Faces) nebo pomocí další jednotky, komunikující ze systémem přes webovou službu. Tímto je možné celý systém poměrně jednoduše integrovat do dalšího funkčního celku.

Výpočet dekontaminační strategie má na starosti komponenta genetického programování. Ta vybírá sekvence dekontaminačních technologií a sestavuje kandidáty na řešení podle vstupních kritérií. K parametrizaci dekontaminačních rovnic slouží jednotka genetického algoritmu (Pavlovič, Hřebíček 2006), která nastaví parametry, aby co nejlépe vyhovovali vstupním kritériím.

Jelikož jsou genetické algoritmy a genetické programování výpočetně náročné, je nutné v tomto případě použít pro výpočty paralelismus. Pro správu jednotlivých výpočetních uzlů, tak aby se jevíli pro systém jako jedna výpočetní služba. Pro tyto účely je v systému JBI komponenta JXTA (<http://www.jxta.org>), která spravuje celý výpočetní cluster. Jednotlivé výpočetní uzly lze do systému přidávat a odebírat. Navíc není systém, díky konceptu JXTA, citlivý na výpadek, kteréhokoli uzlu.

Dále je součástí systému JBI komponenta, pro cachování výpočtů, které řešeno na principu cachování webových služeb (Kubásek 2006). Cachování odlehčuje výpočetní komponentě a umožňuje tak dosahovat výsledků v lepším čase a pracovat s větší přesností. Jednotlivé služby jsou popsány bázi ontologie, kde je pro každou službu definován způsob cachování. Cachovací modul je na systému plně nezávislý a v případě potřeby může být odpojen.

Poslední stávající součástí systému je vizualizační komponenta. Ta přebírá od výpočetní jednotky výsledky a realizuje grafickou reprezentaci výsledných hodnot uživateli. Izolace vizualizace přináší možnost měnit styl grafické reprezentace bez větších zásahů do funkcionality systému. Jednotka vygeneruje JSF komponentu, která pomocí AJAX technologie komunikuje s uživatelem přes JBI komponentu kalkulátoru.



Obr. 2: Servisně-orientovaná architektura systému

8 Vizualizace

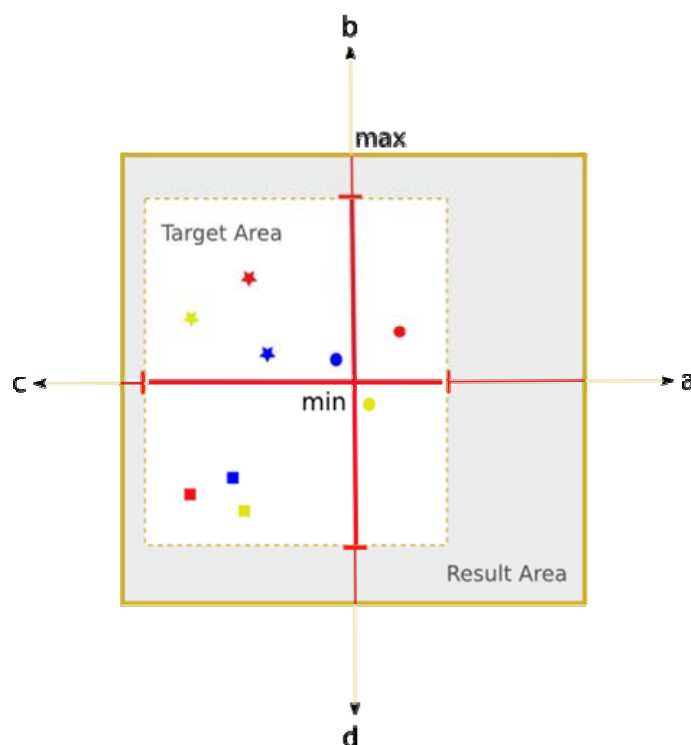
Jelikož je výsledkem výpočtů dekontaminační strategie obrovské množství kombinací, musí environmentální odborníci zadat několik počátečních kritérií, které výslednou množinu omezí. Nicméně i potom představuje výsledná množina velké množství kandidátů na finální řešení. Proto bylo nutné navrhnout novou metodiku abstraktní vizualizace výsledků, pomocí které by mohli experti lépe vybrat finální strategii a to s ohledem na počáteční kritéria.

V oblasti vizualizace již bylo popsáno několik metod (RadViz, FreeViz), které se svoji funkcionalitou blíží požadavkům na vizualizace dekontaminační strategie, ale nejsou zcela vhodné. Proto jsme s kolegy navrhli novou metodu. Díky této vizualizaci mohou experti porovnat několik dekontaminačních strategií v závislosti na zadaná kritéria. Dále je pak možné určit optimální kombinaci jednotlivých technologií v rámci dekontaminační strategie.

Obrázek 3. ukazuje schéma AxeViz (osově orientované) vizualizace pro 4-dimenzionální úlohu (Gregar, Pavlovič, Kokrment 2006). Tyto dimenze jsou definovány osami a, b, c a d. Nicméně je možný i jiný počet dimenzí. Všechny osy začínají ve stejném bodě (minimální možnou hodnotou všech dimenzí). Maximální hodnoty jsou ohraničeny oblastí výsledků (celková plocha dané problémové domény). Pro zjednodušení jsou tyto dimenze normalizovány (tudíž je možné je zobrazit v kružnici).

Uživatel omezí oblast výsledků nastavením jednotlivých dimenzí. V praxi není možné spočítat všechny výsledky. Jednalo by se o výpočetně náročný problém. Podle získaných výsledků, může uživatel zpřesňovat cílovou množinu změnami hranic. Celkově může uživatel modifikovat vizualizaci těmito způsoby:

- *Detailnější pohled v rámci stejných kritérií* – K tomu je možné použít funkci celkového zoomu. V rámci této operace nedojde k žádným novým výpočtům. Nicméně se uživateli naskytne detailnější pohled na shluky výsledků.
- *Změna kritérií vizualizace* – Posunutím hraničních oblastí výpočtů. Tato operace vyžaduje výpočet nových řešení ve vniklé oblasti.
- *Přidání nebo odebrání pozorovaných aspektů* – Což reprezentuje přidání nebo odebrání některých os. V tomto případě je nutné přepočítat veškeré výsledky.



Obr. 3: Osově orientovaná vizualizace

9 Závěr

Hlavním cílem této práce, bylo popsat celou problematiku dekontaminační strategie, navrhnout architekturu, která bude reflektovat požadavky moderních ICT technologií a umožní realizovat výpočet dekontaminační strategie. Jelikož se jedná o výpočetně náročnou problematiku, je nutné v návrhu architektury počítat s paralelismem výpočtů. Z hlediska realizace je potřeba koncipovat celý návrh tak, aby nevyžadoval příliš nákladnou realizaci a dalo se v maximální možné míře využít stávající infrastruktury. Použití genetických algoritmů a genetického programování se ukázalo jako vhodné řešení pro tuto problematiku (Jednička 2005). Genetické algoritmy jsou odolné vůči lokálním extrémům a dají se vhodně paralelizovat.

Celková koncepce servisně-orientované architektury umožňuje integrovat všechny funkční komponenty do jednoho celku a následně systém rozšiřovat a integrovat do již existujícího environmentálního systému.

10 Literatura

- [1] Berlin, D. (2004). The enterprise service bus - a ride you shouldn't miss, ZDNet Industry News, http://news.zdnet.com/2100-9595_22-5281487.html.
- [2] Chappell, D. (2004). Enterprise Service Bus: Theory in Practice, O'Reilly Media.
- [3] Gregar, T., Pavlovič, J., Kokrment, L. (2006). Visualization of Multidimensional Information from Scientific Computations. In Proceedings of I International Conference on Multidisciplinary Information Sciences and Technologies. Spain : American Society for Information Science and Technology.
- [4] Ibarra, F. (2004). The Enterprise Service Bus: Building Enterprise SOA, Dev2Dev, http://dev2dev.bea.com/pub/a/2004/12/soa_ibarra.html.
- [5] Jedlička, M. (2005). Využití genetických algoritmů při optimalizaci výběru sanačních technologií, diplomová práce, Fakulta informatiky Masarykova univerzita.
- [6] Král, J., Žemlička, M. (2005). Service Orientation in Environmental Information Systems, In Proceedings of the ENVIROINFO BRNO 2005 - Informatics for Environmental Protection, pages (12-23), Masarykova univerzita v Brně.
- [7] Kubásek, M. (2006). Caching Optimization in Service-Oriented Architecture, Ph.D. Thesis, Faculty of Informatics Masaryk University.
- [8] Mahutova, K., Pavlovič, J. (2004). Energy Management at Waste Clean up Sites, Avoiding Secondary Air Impacts to Human Health, First Biennial Central & Eastern European Environmental Health Conference.
- [9] Pavlovič, J. (2004). Webový portál pro energetický kalkulátor znečištěného území, In Proceedings of the Tvorba softwaru 2004, Vysoká škola báňská Technická univerzita Ostrava.
- [10] Pavlovič, J. (2005). Multidimensional Decision Support Tool for Remedial Technologies, In Proceedings of the ENVIROINFO BRNO 2005 - Informatics for Environmental Protection, pages (388-392), Masarykova univerzita v Brně.
- [11] Pavlovič, J., Hřebíček, J. (2006). Finding Optimal Solutions of Energetic Remedial Equations with Genetic Algorithms, Enviroinfo2006, Graz.
- [12] Sommers, F. (2005). Service-Oriented Java Business Integration: Put your Integration on Auto-Pilot with JSR 208, Artima Developer, <http://www.artima.com/lejava/articles/jbi.html>.

Komunikační systém NOP On-line

Miroslav Petráš, Jaroslav Ráček

Masarykova univerzita, Institut biostatistiky a analýz,
Divize environmentální informatiky a modelování
Jana Uhra 10, 602 00 Brno
petras@iba.muni.cz, racek@iba.muni.cz

Abstrakt

Příspěvek popisuje vývoj komunikačního portálu NOP On-line (Národní onkologický program on-line). Portál je vytvářen na principech třívrstvé architektury. Základem je datová vrstva obsahující databázi komplexních onkologických center, která skrze portál komunikují. Unikátní datovou strukturu představuje parametrická dokumentace pacienta. Dále jsou popsány jednotlivé komunikační funkce, jež jsou do portálu postupně přidávány v podobě samostatných modulů.

Abstract

This paper presents development of a web communication portal NOP On-line (National Oncological Program On-line). The portal is built on principles of three layer architecture. The base of system is data layer containing a database of complex oncological centers, which use portal for the communication. Parametrical documentation of patient presents a unique data structure of the portal. Target communication functions added into portal in the form of separated units are described in the paper.

Klíčová slova

Komunikační systém, medicínská informatika, webový portál, národní onkologický program.

Keywords

Communication system, medical informatics, web portal, national oncological program.

1 Úvod

Projekt NOP On-line je novou iniciativou v oblasti onkologické informatiky. Hlavní aktivitou projektu je provozování informačního portálu NOP On-line. Informační portál NOP On-line jsou odborné internetové stránky věnované onkologii.

Hlavním cílem projektu je vytvoření centrálního, odborně garantovaného a kontrolovaného informačního zdroje o průběhu naplňování Národního onkologického programu v České republice. Z hlediska uživatele se jedná o komplexní informační zdroj zaměřený primárně na jednotlivá komplexní onkologická centra (KOC), případně na jednotlivé komplexní onkologické skupiny (KOS), což jsou sdružení vždy několika onkologických pracovišť, ale i na odbornou a laickou veřejnost a na zahraniční uživatele sledující informace o stavu onkologické péče v České republice.

Projekt NOP On-line vznikl na základě schválení a pod garancí Výboru České onkologické společnosti. Projekt je řízen odbornou radou složenou z čelních představitelů všech zapojených center, členů Výboru ČOS a významných expertů pracujících v dané oblasti. Garančním pracovištěm odpovědným za vývoj tohoto informačního portálu je Institut biostatistiky a analýz, Lékařské a Přírodovědecké fakulty Masarykovy univerzity v Brně.

2 Funkce portálu

Funkce portálu NOP On-line lze rozdělit do tří základních skupin. První skupinu představují funkce využívané jednotlivými KOC/KOS. Jsou to:

- jednotná a standardizovaná prezentace základních informací o všech KOC v České republice,
- prostor pro individuální prezentaci jednotlivých KOC směrem k centrem zvolené cílové skupině (např. nabídka služeb pro zdravotnická zařízení v regionech),
- prostor pro přímou komunikaci s ostatními KOC a jinými zdravotnickými zařízeními v regionech až na úroveň konzultace individuálních pacientů při zajištění splnění všech legislativních norem pro tuto formu komunikace,
- možnost zapojit se do některého z existujících nebo zahájit nový projekt multicentrického sběru medicínských dat za účelem jejich vědeckého zpracování a publikace výsledků v odborném tisku,
- možnost využít nabídky statistického zpracování existujících dat KOC,
- možnost prezentovat anotace nebo výsledky vlastních vědeckovýzkumných projektů a oslovit další centra ke spolupráci,
- možnost aktivně se podílet na struktuře, obsahu a nabídce služeb portálu NOP On-line,
- zpřístupnění existujících významných datových zdrojů o stavu onkologie v ČR.

Druhou skupinu tvoří funkce portálu směrem k zahraničním uživatelům:

- přímý přístup k aktuálním a centralizovaným informacím o naplňování NOP a tedy o stavu prevence a péče o onkologicky nemocné v ČR,
- přímý přístup k významným datovým zdrojům (např. epidemiologické údaje za celou ČR),
- možnost přehledu nabídky služeb a výsledků práce jednotlivých KOC,
- přístup k anotaci a prezentacím výsledků významných projektů z oblasti prevence, diagnostiky a léčby rakoviny v ČR,
- možnost nalezení vhodných partnerů pro mezinárodní spolupráci na vědeckovýzkumných projektech.

Prezentaci projektu a jeho publicitu směrem k širší veřejnosti zajišťují následující funkce portálu:

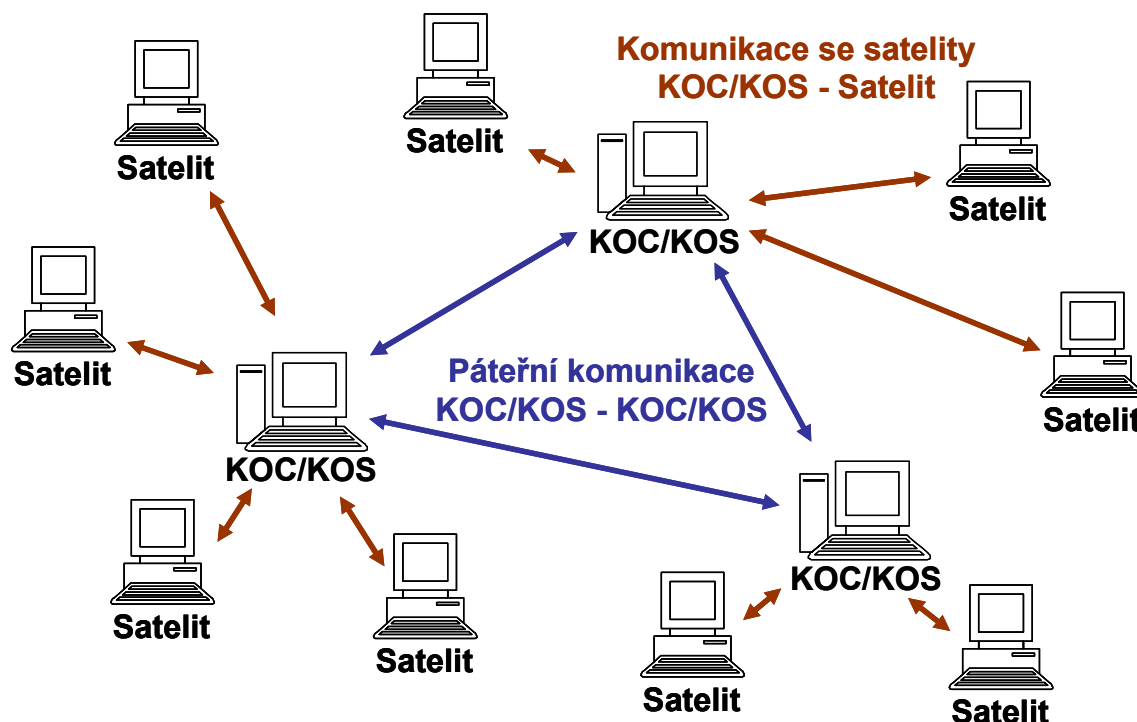
- přímý přístup k jednotným a přehledně strukturovaným informacím o stavu onkologie v ČR,
- přímý přístup k nabídce služeb jednotlivých KOC směrem k ostatním zdravotnickým zařízením i k jednotlivým pacientům nebo jejich rodinným příslušníkům.

3 Komunikační nástroj

Z informatického hlediska si zasluhuje pozornost zejména vývoj komunikačního nástroje KOC/KOS. Jedná se o funkce portálu, které jsou v jeho neveřejné části. Jejich hlavním úkolem je propojit jednotlivá portálem evidovaná onkologická pracoviště a umožnit jim komunikaci na základě předdefinovaných scénářů pro předem vytipované situace. Komunikace se přitom může odehrávat na úrovni jednotlivých zdravotnických zařízení, tj. mezi KOC, KOS a jejich satelity. Příkladem takové situace je předávání zdravotní dokumentace pacientů v okamžiku, kdy pacient přechází z jednoho KOC do jiného. Druhou úrovní, na níž se komunikace může odehrávat je výměna informací přímo mezi jednotlivými lékaři. Příkladem mohou být virtuální koncilia nebo úzce odborně zaměřené diskusní skupiny, kde mezi sebou komunikují přímo jednotliví lékaři.

Veškeré komunikační toky, jež portál nabízí, lze přitom rozdělit do dvou skupin. První skupina toků představuje tzv. *páteřní komunikaci*, kde mezi sebou komunikují KOC/KOS, které jsou z pohledu

hierarchie portálu na stejné úrovni. Druhou skupinu představuje komunikace se satelity, kdy příslušné KOC/KOS komunikují se svými satelitními zdravotnickými zařízeními. Zjednodušené schéma této komunikační sítě ukazuje obrázek č. 1.



Obr. 1: Topologie komunikačních sítě mezi KOC/KOS a jejich satelitními zdravotnickými zařízeními

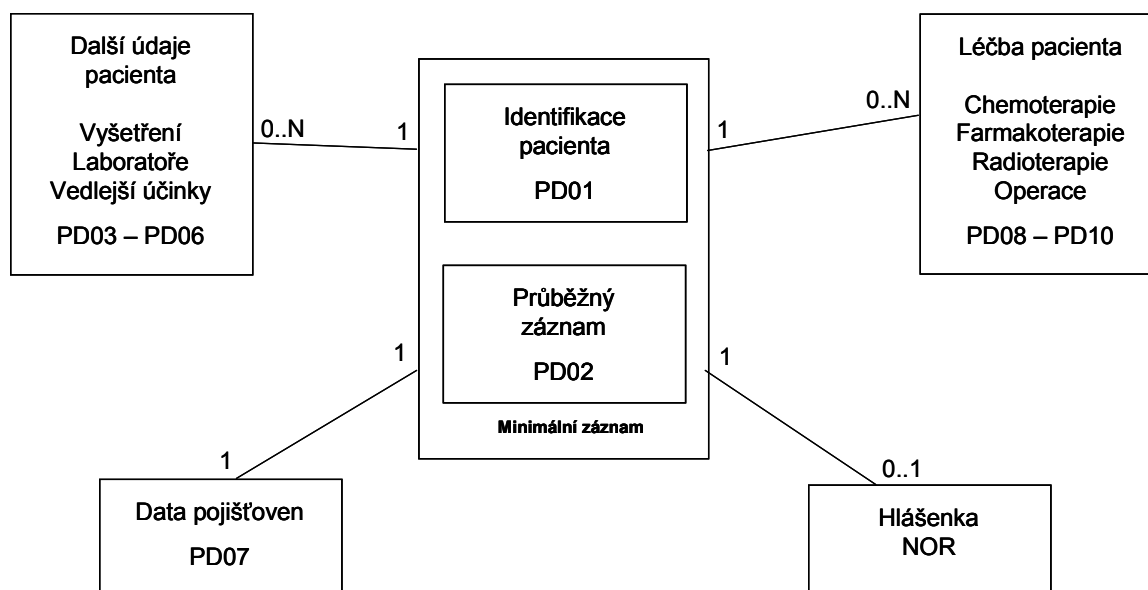
4 Základní databázové struktury

Komunikační nástroj portálu je budován na principech třívrstvé architektury, tj. obsahuje datovou, aplikační a prezenční vrstvu. Datová vrstva pokrývá databázové struktury, které tvoří základ systému. Tyto lze rozdělit do dvou částí, jimiž jsou databáze KOC/KOS a parametrická dokumentace pacientů.

Databáze KOC/KOS obsahuje informace o jednotlivých zdravotních zařízeních, lékařích, dalších zaměstnancích a jejich vzájemných vztazích. Jejím hlavním úkolem je uchovávat veškeré informace o všech komunikujících subjektech v systému, na jejichž základě se přidělují konkrétní uživatelské role a práva.

Unikátní datové struktury představuje parametrická dokumentace pacienta. Jedná se o strukturované informace obsahující veškeré potřebné údaje o pacientovi. Je to v podstatě alternativa ke starším zdravotním kartám, která však díky své databázové povaze umožňuje na rozdíl od původních (převážně textových) záznamů plné využití možností, které nabízejí současné ICT.

Parametrická dokumentace obsažená v komunikačním nástroji portálu NOP ON-line je v současnosti složena z 10-ti sad strukturovaných záznamů. Jádro a zároveň povinnou složku dokumentace tvoří záznamy PD01 „Minimální identifikační záznam“ a PD02 „Minimální průběžný záznam“. Oba záznamy společně zahrnují především vstupní identifikace pacienta a nemoci v základní sadě parametrů, zjistitelných při vstupu pacienta do zdravotnického zařízení. Tyto záznamy lze považovat za univerzálně platné a získatelné od všech pacientů. Jejich účel je částečně popisně statistický a částečně provozní. Součástí záznamů je rovněž popis stavu pacienta při první návštěvě zařízení. Většina parametrů zařazených do této sady se hodnotově nemění v čase. V modulární skladbě celé parametrické dokumentace patří tyto záznamy k povinným a stojícím nad detailními popisy léčby, výkonů, komplikací a samotných výsledků léčby. Základní logický rámec parametrické dokumentace ukazuje obrázek č.2.



Obr. 2: Logický model parametrické dokumentace pacienta

Z předchozího obrázku je patrné, že parametrická dokumentace obsahuje vedle výše zmíněných informací i data vyžadovaná zdravotními pojišťovnami. Dále lze do této struktury zakomponovat i data Národního onkologického registru (NOR).

5 Komunikační funkce KOC/KOS

V aplikační a prezentační vrstvě systém nabízí nástroje umožňující komunikaci, sdílení a předávání dat mezi jednotlivými uživateli (KOC/KOS, satelitní zdravotnická zařízení a jednotliví lékaři). Mezi funkce (nástroje) systému na úrovni komunikace se satelitními zdravotnickými zařízeními patří hlášení pacientů a elektronické dotazování. Na úrovni páteřní komunikace mezi jednotlivými KOC/KOS se jedná o virtuální konzilia, předávání dokumentace, sdílené registry a odborné diskusní skupiny.

Hlášení pacientů slouží k podávání informací do KOC/KOS o pacientech v satelitních zdravotnických zařízeních. Data o pacientech jsou ve formě parametrické dokumentace a jsou tedy kompatibilní i s daty NOR i s daty zdravotních pojišťoven. Anonymita pacienta je zajištěna použitím jednosměrného šifrování. V parametrické dokumentaci je pouze jednosměrný otisk rodného čísla pacienta a privátního klíče pacienta, který je uveden na zdravotní kartě. Ověření identity pacienta (nikoli její zpětný výpočet) lze provést pomocí klíče, který se předává důvěryhodným kanálem mimo systém. Komunikace probíhá na úrovni jednotlivých KOC/KOS a jejich satelitních zdravotnických zařízení.

Elektronické dotazování dává možnost vytvoření vlastního dotazníku z předem předdefinovaných částí, jeho následné rozeslání zvolené skupině uživatelů a správu odpovědí. Je možná i tvorba nových vlastních komponent dotazníků. Vytvořený dotazník lze uložit jako šablonu připravenou ke znovupoužití nebo jako část jiného dotazníku. Jako komponenty dotazníku mohou být použity i libovolné části parametrické dokumentace. Odpovědi jsou ukládány automaticky, k dispozici jsou nástroje pro jejich prohlížení a další zpracování. Komunikace probíhá na úrovni jednotlivých KOC/KOS, jejich satelitních zdravotnických zařízení i na úrovni konkrétních osob (lékařů).

Virtuální konzilia vytváří prostor pro diskuse vybrané skupiny přímo oslovených lékařů nad dokumentací konkrétního pacienta, přičemž identita pacienta se nesdílí. Celý proces vypadá tak, že data ve formátu parametrické dokumentace jsou spolu s připojeným dotazem rozeslána na další KOC/KOS. Odpovědi jsou ukládány automaticky, jsou k dispozici nástroje pro jejich prohlížení a další zpracování. V případě konzilií probíhá komunikace na úrovni konkrétních lékařů.

Předávání dokumentace konkrétního pacienta z jednoho KOC/KOS do jiného, případně i z jiných zdravotnických zařízení, je jednou z klíčových funkcí komunikačního nástroje. Na rozdíl od konzilií je

v tomto případě možné zjistit (předat) i pacientovu identitu. Ověření a následné přiřazení identity pacienta lze provést pomocí klíče, který se předává důvěryhodným kanálem mimo systém. Data jsou ve formátu parametrické dokumentace. Předávání dokumentace probíhá na úrovni jednotlivých KOC/KOS.

Sdílené registry představují sadu funkcí, které umožňují vytváření specializovaných registrů sdílených mezi více KOC/KOS. Libovolné KOC/KOS může definovat strukturu vlastního registru a nastavit přístupová práva ostatním. Data v registru mohou využívat parametrickou dokumentaci, resp. její části. Přístupová práva mohou být přidělena na úrovni celých KOC/KOS i na úrovni jednotlivých osob.

Odborné diskusní skupiny jsou určeny k diskusím na úzce specializovaná témata. V rámci diskuse mohou být zpřístupněna i další data, např. parametrická dokumentace, výsledky dotazování nebo obsah registrů. Každý KOC/KOS může tyto diskusní skupiny zakládat. Přístup do diskusí mohou mít i osoby z dalších KOC/KOS nebo i ze satelitních zdravotních zařízení. Přístupová práva jsou přidělována na úrovni jednotlivých osob.

6 Vývoj systému

V době vzniku tohoto příspěvku, tj. v září 2006 byla dokončena první etapa vývoje komunikačního portálu NOP On-line. Byly vytvořeny všechny základní struktury datové vrstvy a implementovány na serveru Institutu biostatistiky a analýz MU.

Současně s tím byl zahájen vývoj aplikační a prezentační vrstvy systému. Jednotlivé komunikační funkce byly namodelovány v podobě business procesů a následně převedeny dle metodiky SELECT Perspective do podoby sad UML diagramů. Tyto výstupy jsou podrobovány kritice odborné veřejnosti a poté budou v následující fázi sloužit jako podklady pro programování.

Systém je implementován jako webový portál nad databází Oracle9. Jako skriptovací jazyk bylo použito PHP4. Klíčové funkce sloužící zejména pro šifrování osobních údajů jsou programovány v jazyce Java.

7 Závěr

NOP On-line je ambiciózním projektem, který přináší do medicínské informatiky v České republice řadu nových prvků. Jedná se zejména o využívání parametrizovaných záznamů pro uchovávání údajů o pacientech a práci s nimi v prostředí WWW.

Velmi významným aspektem celého projektu je velmi vysoký důraz na bezpečnost osobních údajů pacienta, který vyplývá z legislativy. Z tohoto důvodu je velmi významnou částí vývoje celého komunikačního portálu implementace řady bezpečnostních mechanismů využívajících algoritmů symetrické i asymetrické kryptografie.

Vlastní funkcionalita portálu se bude v budoucnu neustále rozšiřovat na základě přání a připomínek uživatelů z jednotlivých KOC/KOS. Z tohoto důvodu je portál vyvíjen v režimu inkrementálního životního cyklu. Důraz, který je kladen na čistotu datové vrstvy, nad níž jsou jednotlivé aplikační přírůstky postupně budovány, by měl v budoucnu umožnit bezproblémové rozšiřování, případně modifikaci funkcionality.

8 Literatura

- [1] BPM, Business Process Modeling (BPM). White Paper, Select Business Solution Inc., 2003.
- [2] Král, J.: Informační systémy. Science, Brno, 1998.
- [3] Rumbaugh, J., et al.: Object Oriented modelling and Design, Prentice-Hall International, New Jersey, 1991.

- [4] Mužík, J., Dušek, L., Koptíková, J., Pavliš, P., Kubásek, M., Žaloudík, J., Vyzula, R.: SVOD - Information System for Inventive Analysis of Cancer Epidemiology. Proceedings of the 1st International Summer School on Computational Biology. Masarykova univerzita, Brno, 2005.
- [5] Rumbaugh, J., et al., Object Oriented modelling and Design, Prentice-Hall International, New Jersey, 1991.
- [6] Šebestová, M., Váňa, V., Sedláček, M.: Management služeb IT – Komentované vydání souboru ISO/IEC/DIS 20000:2004. Český normalizační institut, Praha, 2005.
- [7] Vorlíček, J.: Klinická onkologie - I. díl - obecná část. Masarykova universita, Brno, 1995.

Příspěvek vznikl v rámci projektu „NOP on-line“ podporovaného společnostmi Astra Zeneca pod garancí Výboru České onkologické společnosti.

Moderní metodiky, technologie a nástroje pro podporu výuky (E-learning 2.0)

Tomáš Pitner, Pavel Drášil

Masarykova univerzita, Fakulta informatiky,
Botanická 68a, 602 00 Brno
tomp@fi.muni.cz, xdrasil@fi.muni.cz

Abstrakt

Příspěvek ukazuje vývoj ve vnímání elektronické podpory výuky souhrnně označované jako E-learning 2.0. Zmiňuje nová metodická paradigmatata (zejména konektivismus) a jejich požadavky na elektronickou podporu. Naplnění těchto požadavků je ilustrováno na vzorové implementaci – platformě L2 – umožňující integrovat stávající výukové i obecně použitelné zdroje a vytvořit tak personalizované výukové prostředí (learning landscape) odpovídající principu konektivismu a kolaborativního učení.

Abstract

With the current shift in web technology called Web 2.0, many e-learning experts compare the traditional view of e-learning with the expected level of e-learning services and find common problems with present LMS platforms. Primarily, we will identify the most important properties of an E-learning 2.0 platform. Secondly, a prototype, Java-based open-source E-learning 2.0 platform L2 will be introduced and evaluated.

Klíčová slova

Web 2.0, e-learning 2.0, REST, adaptive XML inclusions, lightweight P2P personalization.

Keywords

Web 2.0, e-learning 2.0, REST, adaptive XML inclusions, lightweight P2P personalization.

1 Úvod

V průběhu několika posledních let jsme svědky zajímavého posunu ve způsobu využívání sítě Internet. Současný web přestává být pouhým nástrojem komunikace. Stává se platformou, kde se stírají role mezi tradičními poskytovateli informací a jejich konzumenty. Úspěšné technologie jako weblogging, wiki, či služby sdílení nejrozličnějších zdrojů (databáze odkazů, alba fotografií) jsou charakterizovány důrazem na širokou participaci, neformálnost, technologickou nekomplikovanost a především jednoduchost použití.

Vývoj je bezpochyby umožněn především snadnou dostupností internetových technologií a služeb a jejich zvládnutím širokými vrstvami i „netechnicky“ zaměřené populace. Trendy směřující k otevřenosti, participaci a slévání skupin poskytovatelů a konzumentů obsahu se odrážejí i v uživatelských očekáváních v oblasti podpory výuky.

Vzdělávání přestává být pouhým transferem „stabilních“ znalostí od učitele a jejich absorpcí studenty. V záplavě volně dostupných zdrojů, z nichž se může neformálně učit každý je podstatné, jak se v problematice orientovat, jak (si) určit výukový cíl, jaké zdroje k načerpání potřebných znalostí použít a jak přitom koordinovat práci s ostatními, protože výuka musí směřovat i k budování sociálních dovedností. Žádná z tradičních ani modernějších teorií učení a vzdělávání (konstruktivismus [9], behaviorismus, kognitivismus) nedokáže tyto trendy dostatečně popsat a metodicky podpořit.

Základem učení se stává schopnost budovat si *spojení* mezi jednotlivými pojmy, příslušnými elementy výukového obsahu, oblastmi použití a především osobami – kolegy, spolužáky nebo členy nějaké širší, formální či neformální komunity. Metodiky nahlížející na proces učení primárně jako na budování spojení se označují jako *konektivismus* [3].

V tomto příspěvku se proto pokusíme ukázat, jaké změny a zejména jaké výhody s sebou zmíněný trend přináší v oblasti e-learningu. Na závěr představíme i aktuálně vyvíjený prototyp e-learningového prostředí nové generace.

2 Web 2.0

Pozvolné změny v přístupu k využívání internetu se jeví jako natolik významné, že si zasloužily nové označení – Web 2.0. Obsah tohoto pojmu není snadné explicitně vymezit a vystihnout jednoduchou definicí, proto si pomáháme příklady vlastností, jež Web 2.0 charakterizují. Tim O'Reilly uvádí výčet těchto vlastností v [2]:

- jde o snadno rozšiřitelné služby (navíc stále se vyvíjející)
- spravující unikátní, obtížně obnovitelné datové zdroje, jejichž hodnota stoupá, pokud jsou využívány (síťový efekt)
- důvěřují uživatelům jako spoluvůrcům (každý je konzumentem i přispěvatelem)
- využívají „kolektivní inteligenci“ (konektivistická vize znalosti existující i mimo člověka – subjekt učení)
- pro své šíření využívají (a motivují) vlastní uživatele (většina peer-to-peer sítí a systémů)
- jsou využitelné na různých zařízeních (mobilní zařízení, pervazivní technologie)
- mají jednoduchá uživatelská i aplikační rozhraní (snadné použití, „marketingový kabát“ je důležitý, ale ne klíčový)

Lze také říci, že pod označení Web 2.0 spadá většina nejúspěšnějších služeb dnešního internetu – Google (s celou řadou služeb – *Gmail*, *Google Maps*, ...), *Wikipedii*, wiki, blogy, peer-to-peer síť *BitTorrent*, stejně jako služby *Flickr.com*, *del.icio.us* nebo *upcoming.org*.

3 Web 2.0 a e-learning

Vzhledem k inherentní provázanosti e-learningu a Internetu samozřejmě netrvalo příliš dlouho a koncept Webu 2.0 začal postupně pronikat i do elektronické podpory výuky. Někteří autoři pro tuto kombinaci dokonce začali používat logicky ospravedlnitelné, ale přesto kontroverzní označení *E-learning 2.0* [1].

3.1 Současný e-learning

E-learning je v podobě, ve které je dnes běžně provozován, typickou aplikací klasického webu („1.0“, „pre-2.0“). Jádrem celého procesu je zpravidla monolitický systém pro řízení výuky (LMS), který poskytuje nástroje pro správu kurzů, správu výukových materiálů, testování znalostí, různé formy komunikace mezi uživateli apod. Tyto funkce bývají někdy uspořádány tak, aby vhodně podporovaly některou z tradičních (před-e-learningových) teorií učení – většina tak či onak metodicky odráží behaviorismus a kognitivismus, existují však i elektronické nástroje podporující konstruktivistické přístupy.

Výukové materiály mohou být do systému dodávány ve formě standardizovaných (přenositelných) výukových objektů odpovídajících nejčastěji specifikaci SCORM [5]. Internet je využíván pouze k zajištění komunikace mezi vzdálenými počítači a umožňuje výměnu zdrojů, které by de facto mohly být vyměňovány i bez jeho pomoci.

3.2 Kritika

Konektivismus, tj. moderní teorie učení nadto přizpůsobená aktuálnímu stavu technologií, preferující osvojování nových znalostí formou propojování informací získávaných z různých (specializovaných) zdrojů, odpovídá způsobu vnímání současné generace daleko lépe [1]. Dnešní LMS tento trend bohužel nepodporují a díky tomu, do jaké role se od počátku staví, vlastně ani příliš podporovat nemohou.

3.3 E-learning 2.0

V souladu s idejemi *Webu 2.0* se dnes začíná jevit jako výhodnější zcela opačný přístup k budování systémů pro podporu výuky – konstruovat je jako snadno rozšiřitelnou sadu vzájemně kooperujících jednoduchých (jednouúčelových) služeb. Pro takto navržený systém pak není problém napojení na téměř libovolné internetové zdroje informací. Navíc je možné zcela transparentně (z pohledu uživatele i stávajících služeb) do systému doplňovat další funkcionalitu, například personalizační mechanismy.

Velmi zajímavá je i myšlenka aktivního zapojení studentů do přípravy a vedení kurzu. Každý z nás se již jistě někdy dostal do situace, kdy při čtení nějakého materiálu narazil na chybu. Jen málokdo už byl ale ochoten přerušit čtení, najít si kontakt na autora a komplikovaně mu například do mailu popisovat, kde našel jaký problém. O co jednodušší a hlavně účinnější je vyznačení nebo rovnou opravení chyby přímo v systému. I takováto zdánlivá maličkost může výrazně přispět ke kvalitě výuky. Oproti tradičním systémům sdílení výukového obsahu prostřednictvím LMS nebo CMS jsou jednoduché platformy na bázi wiki daleko praktičtější.

Reálné možnosti zapojení studentů pochopitelně sahají ještě mnohem dál.

4 Platforma L2

Pro ověření konceptu *E-learningu 2.0* se ve světě objevily experimentální systémy umožňující vytvářet prostředí založené na principech konektivismu – jejich typickým představitelem je Elgg [4].

Cílem našeho výzkumu a vývoje bylo konceptuálně navázat na snahy vývojářů Elgg, vybudovat funkčně příbuznou platformu, která bude mít:

- pevnější technologický základ - Java jako realizační platforma, *REST* (Representational State Transfer [6]) a *Restlet API* (<http://www.restlet.org>) jako architekturní báze
- silnější orientaci na integraci služeb – díky *Adaptive XML Inclusions* [7]
- jednoduché programové i uživatelské rozhraní
- personalizace vlastních i integrovaných zdrojů pomocí *Lightweight P2P Personalization* [8]

V souladu s popisovanými koncepty jde o sadu volně provázaných služeb, jejichž cílem je podpora a zjednodušení každodenních činností pedagogů i studentů takovou formou, která co nejvíce odpovídá jejich běžným návykům.

4.1 L2 jako základ „Learning Landscape“

Namísto fyzického shromažďování výukových materiálů na jediném místě podporuje platforma L2 jejich odkazování na původních zdrojích. Díky tomu je možné zcela transparentně přistupovat k datům nezávisle na jejich fyzickém umístění a přístupovém mechanismu. Prezentovány tak mohou být stejně dobře veřejné webové stránky jako dokumenty uložené v databázích a knihovnách nebo informace získané z RSS kanálů. Jedinou podmínkou je adresace zdroje pomocí URL.

Na jednotlivé objekty udržované platformou nebo integrované v ní se z pohledu REST nahlíží jako na *zdroje* (Resources). Zdroje používané určitou službou L2 (např. weblogem) mohou být distribuovány na libovolných instancích L2 nebo dokonce mimo L2. Stačí, jsou-li potřebné *reprezentace* zdroje přístupné na známém URL protokolem HTTP.

Kromě samotného procházení zdrojů poskytuje systém už v aktuální verzi svým uživatelům některé prvky interaktivního prostředí vhodného pro spolupráci – uživatelé si mohou zakládat osobní blogy, seznamy odkazů, mohou jednotlivé zdroje opatřovat soukromými či veřejnými poznámkami, psát a provozovat wiki stránky, ...



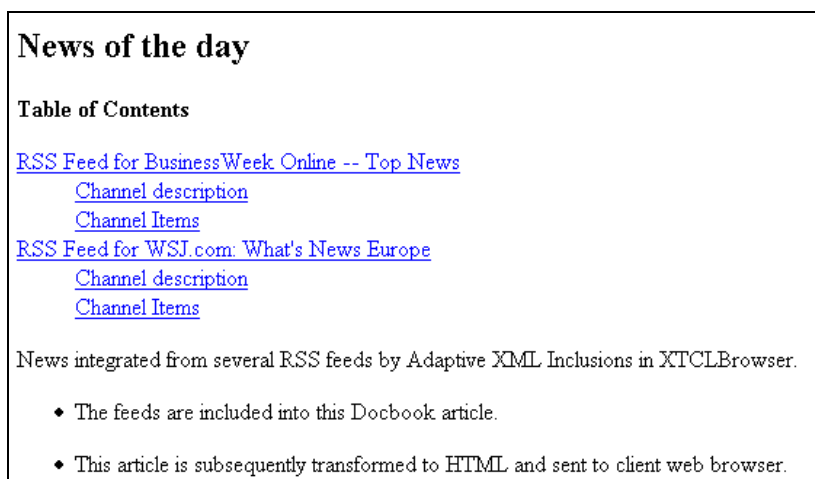
Obr. 1: Záznam do blogu na platformě L2

Takto postavený systém je velmi snadno napojitelný na jakékoliv služby dostupné na Internetu (vyhledávání apod.). Stejně tak není (z technického hlediska) problém ani v integraci s jinými prostředími, která by mohla poskytovanou funkcionalitu využívat.

4.2 Adaptační a personalizace zdrojů

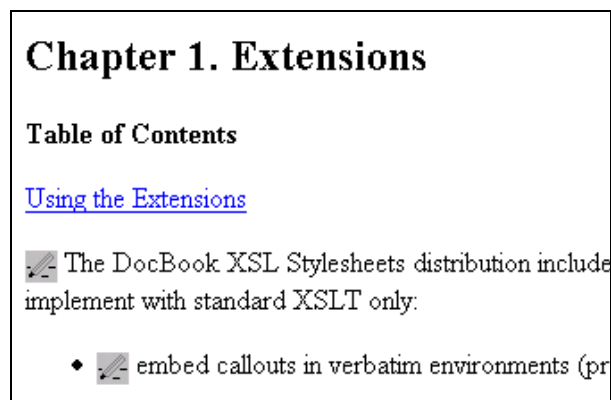
Podstatným znakem platformy L2 je také možnost personalizovat *jakýkoli* (tzn. i externí) zdroj integrovatelný do obsahu na platformě. Stačí se tedy např. z wiki stránky odkázat na RSS proud a platforma za nás zrealizuje oba podstatné úkony – *adaptaci* i *personalizaci*.

Za prvé tedy obsah proudu *adaptuje*, v tomto případě prezentuje RSS v HTML vyhovujícím každému prohlížeči, přičemž je dokonce schopná integrovat několik RSS proudů. Podobná adaptace se může dít nad jakýmkoli (XML) zdrojem – webovou stránkou, wiki obsahem, blogem apod. Následující obrázek ukazuje sdružení dvou RSS proudů do jednoho adaptovaného výstupu.



Obr. 2: Integrace a HTML prezentace RSS proudů v L2

Kromě toho umí L2 každý integrovatelný zdroj i *personalizovat*, tj. upravit do podoby specifické pro konkrétního uživatele či skupinu uživatelů. Personalizace je ve stávající verzi L2 chápána jako možnost vytvářet podle osobních či skupinových preferencí pohledy na integrované zdroje – typicky do nich virtuálně dopisovat poznámky či mazat (skrývat) nezajímavá místa.



Obr. 3: L2 dokáže personalizovat libovolný XML zdroj

4.3 Zamýšlená rozšíření

Platforma L2 je díky důslednému budování na principech REST webových služeb optimálně připravená na škálování a rozšiřování. Přímocárým rozšířením platformy L2 se jeví obohacení personalizační vrstvy o všechny standardně uvažované osy personalizace při zachování její ortogonalitě, tzn. její uplatnitelnosti na všechny zdroje integrovatelné do platformy – jimiž jsou libovolná XML data.

V souvislosti s reálnou použitelností aspoň některých nástrojů sémantického webu se nabízí zejména integrace základních služeb přístupu k ontologiím či mapám témat (Topic Maps) přímo do jádra L2. Tyto služby bude možné využít v adaptačním i personalizačním modulu a obohatit škálu dostupných technik adaptace a personalizace o transformace integrovaných zdrojů na „sémantické“ úrovni – nad rámec dosavadní úrovně syntaktické (tj. transformace značkování). Velmi zajímavou oblastí přímo odpovídající trendům E-learning 2.0 jsou rovněž vizualizační služby [10].

5 Závěr

Současné trendy v koncipování a používání webových aplikací označované jako Web 2.0 se odrážejí i v systémech podpory výuky. Proces koresponduje s metodickým posunem směrem ke konektivismu a bývá některými autory označován jako E-Learning 2.0. Hlavní soudobé metodické principy kladou důraz na kolaborativní charakter vzdělávání, nutnost propojovat a integrovat stávající zdroje a staví na širokém prostoru pro neformální komunikaci a aktivní participaci. Tento článek přiblížil volně dostupnou platformu L2, která – na rozdíl od tradičních systémů řízení výuky – dokáže uvedené principy elektronicky podpořit.

6 Literatura

- [1] Downes, S.: E-learning 2.. ACM eLearn Magazine, 2005.
- [2] O'Reilly, T.: What Is Web 2.0. Design Patterns and Business Models for the Next Generation of Software. 2005.
- [3] Siemens, G.: Connectivism: A learning theory for the digital age. 2004.
- [4] Werdmuller, B., Tosh, D.: Elgg Learning Landscape, <http://elgg.net>.
- [5] Drášil, P., Bažant, I., Šimák, B., Pitner, T.: Relevantní standardy v oblasti e-Learningu. CESNET, 2004. Technická zpráva.
- [6] Fielding, R. T.: Representational State Transfer (REST), kapitola z disertační práce, 2000.
- [7] Pitner, T., Adámek, P.: Improving Document Reusability with Adaptive XML Inclusions, XTech 2005 Conference, Amsterdam, 2005.
- [8] Pitner, T.: Lightweight Personalization of XML E-learning Materials, Berliner XML Tage, 2003.

- [9] Wikipedia. Constructivism (learning theory).
- [10] Gregar, T., Nevěčilová, Z., Rambousek, A., Pitner, T.: Vizualizace znalostí v e-learningu. In Sborník konference SCO 2006. Masarykova univerzita, Brno, 2006.

Příspěvek vznikl za podpory projektu Národního programu Informační společnost „E-learning v kontextu sémantického webu“, č. 1ET208050401.

Informační systém environmentálního reportingu

Jaroslav Ráček

Masarykova univerzita, Institut biostatistiky a analýz,
Divize environmentální informatiky a modelování
Jana Uhra 10, 602 00 Brno
racek@iba.muni.cz

Jiří Hřebíček

Masarykova univerzita, Institut biostatistiky a analýz,
Kamenice 126/3, 625 00 Brno
hrebicek@iba.muni.cz

Abstrakt

Příspěvek popisuje metodiku a výsledky výzkumného projektu Ministerstva životního prostředí, ve kterém byly analyzovány povinnosti České republiky v oblasti mezinárodního environmentálního reportingu a způsob řízení informací o životním prostředí v ČR. Významným výsledkem projektu je vytvořený informační systém reportingu, který je rovněž popsán v tomto příspěvku.

Abstract

This paper describes methods and results of a research project conducted by the Czech Ministry of Environment in which international environmental information management and reporting obligations of the Czech Republic have been analysed. A considerable output of this project is an information system for management of international environmental reporting described in this paper.

Klíčová slova

Environmentální reporting, informační systém reportingu, procesní analýza.

Keywords

Environmental reporting, reporting information system, process analyses.

1 Úvod

V letech 2003 až 2005 byl na Ministerstvu životního prostředí ČR zahájen projekt VaV 250-3-03 „Analýza a návrh prototypů environmentálních datových modelů a vnějšího rozhraní JISŽP kompatibilních s EU“, jehož řešitelem byla Masarykova univerzita v Brně (MU). Cílem projektu bylo vyvinutí centrálního výstupního datového modelu pro MŽP, který měl reflektovat potřeby strategického plánování, nadoborové informační podpory státní správy, informování veřejnosti, spolupráce s komerční sférou a mezinárodní reporting dle aktuálních požadavků na integrovaný reporting. Řešení projektu bylo rozděleno do čtyř základních etap (Analýza, Standardizace, Harmonizace a Integrace) naplánovaných na dobu 27 měsíců od října 2003 do prosince 2005.

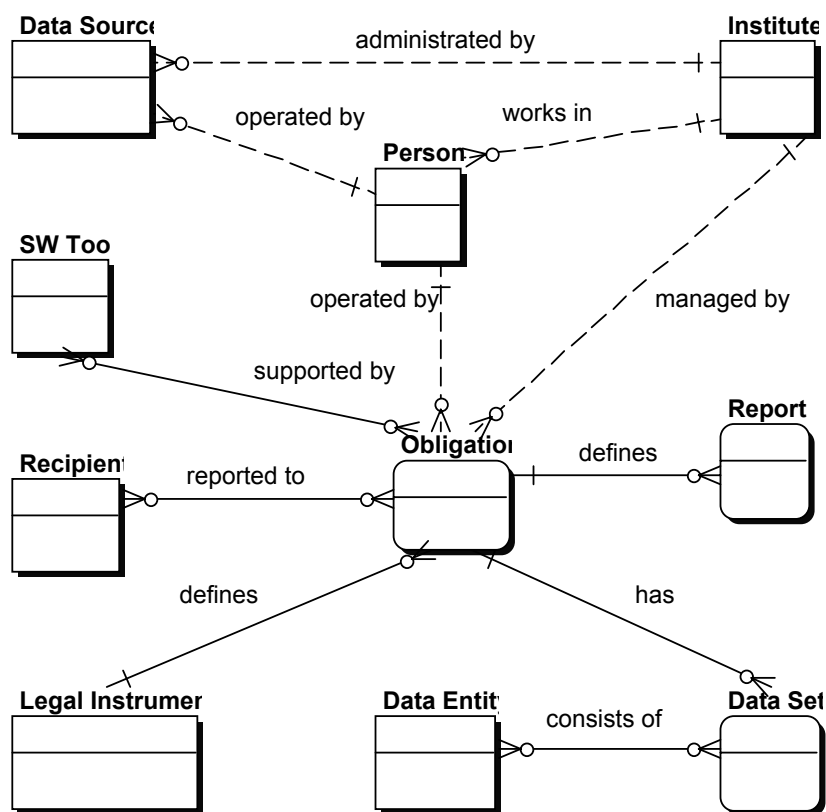
Jedním z výstupů projektu je i informační systému reportingu (ISR) pro sledování environmentálních reportingových povinností ČR vůči Evropské environmentální agentuře a dalším mezinárodním institucím. V první polovině roku 2006 proběhly úpravy ISR směřující k jeho nasazení na České informační agentuře životního prostředí (CENIA) a v červenci byl tento systém oficiálně předán a nasazen do plného provozu v CENIA.

2 Návrh systému

Na základě výsledků analýzy reportingových povinností, která proběhla v první etapě projektu, se řešitelský tým rozhodl řešit reportingový systém formou procesního skladu, který uchovává definice reportingových procesů a ve stanovený čas automaticky vytváří konkrétní instance reportů a jejich

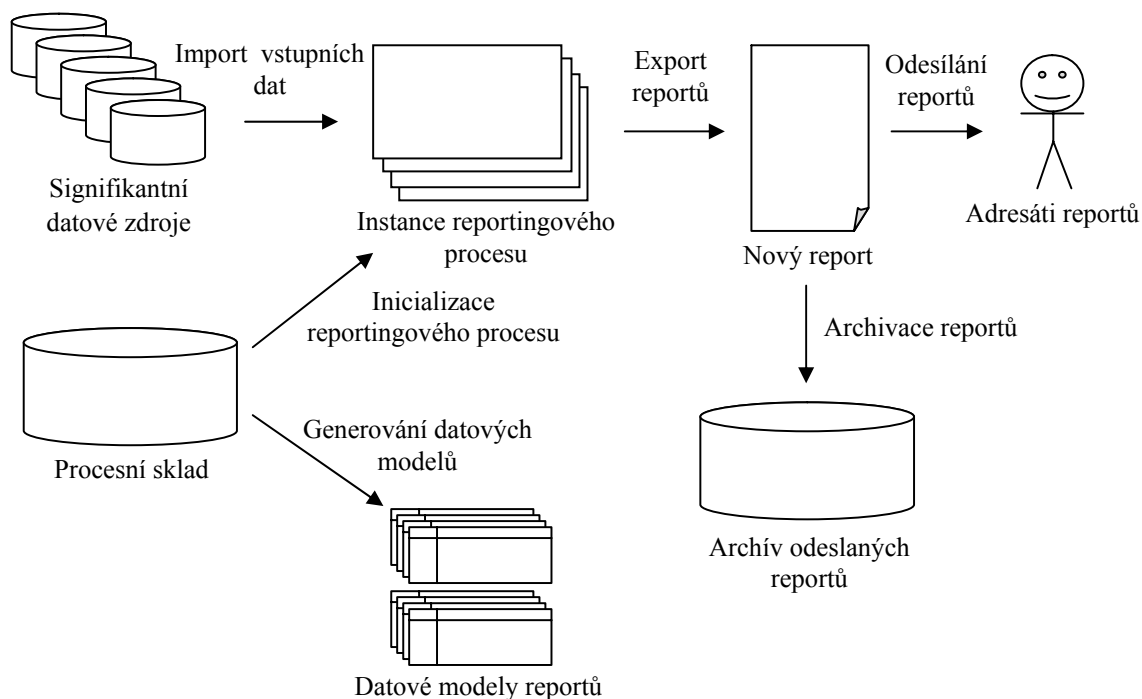
výkon automatizuje. Data reportů jsou přitom čerpána přímo ze současných signifikantních datových zdrojů pomocí konverzních můstků, které procesní portál nabízí pro konkrétní reportingové procesy. V centrálním systému se uchovávají definice reportingových procesů včetně konverzních můstků, pomocí nichž jsou reporty sestavovány, a data již zpracovaných reportů. Primární data, z nichž se tyto reporty vytvářejí, systém neuchovává. Z tohoto pohledu lze říci, že se jedná o procesně orientovaný metainformační systém.

Výhodou tohoto řešení je, že v případě budoucích změn týkajících se reportingu nebude třeba modifikovat vlastní systém a začlenit do jeho datového modelu nové datové položky reportů, které v něm doposud chyběly. Nedochází ani k duplicitě dat, která by se uchovávala v primárních datových zdrojích a zároveň by byla uložena (buď v odvozené formě) i v centrální reportingové databázi. Definice reportingových procesů, tj. způsoby plnění jednotlivých reportingových povinností, jsou uloženy v centrálním procesním skladu. Jejich součástí jsou informace o termínech reportů, datových formátech reportů, příjemcích, zodpovědných osobách, dostupných datových zdrojích, legislativě, konverzních můstcích a dalších aplikacích, které jsou pro tvorbu reportu v systému k dispozici. Logický datový model systému je uveden na obrázku č.1.



Obr. 1: Logický datový model ISR

Z těchto údajů jsou automaticky inicializovány instance reportingových procesů, které pomocí konverzních můstků čerpají zdrojová data z příslušných datových zdrojů. Z těchto dat jsou osobami, které systém automaticky vyzve, sestavovány reporty v požadovaných datových formátech. Ty jsou následně odesílány příjemci a zároveň archivovány v databázi vykonaných reportů. Řešení na bázi procesního skladu je schopné na budoucí změny reagovat zrušením platnosti stávající a přidáním nové definice reportingového procesu, která obsahuje konverzní můstky a definici konkrétního reportu. Popisy datových modelů jednotlivých reportů jsou součástí definic příslušných reportingových procesů, které systém uchovává ve formě metadat. Systém umožňuje dynamicky generovat aktuální datový model všech reportů nebo pouze zvolené podмноžiny reportů, např. pro požadované časové období, osobu nebo signifikantní datový zdroj. Princip navrhovaného systému je znázorněn na obrázku č.2.



Obr. 2: Architektura ISR

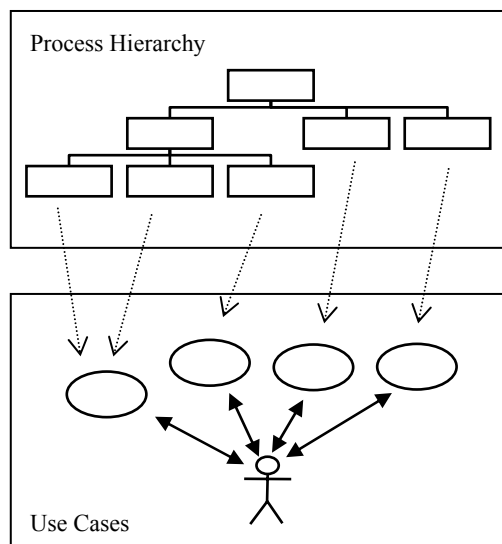
3 Použitá metodika

Během vývoje ISR, zejména ve fázi jeho analýzy a návrhu, byla použita modifikovaná metodika SELECT Prespective (www.selectbs.com). Tato metodika se skládá ze dvou základních fází, jimiž jsou:

- Procesní analýza
- Objektově orientovaná analýza a návrh

Úkolem první etapy je provést analýzu procesů sledované oblasti, nalézt klíčové procesy a jejich hierarchii. Cílem je zaznamenat procesní prostor ve formě PHD (Process Hierarchy Diagram) mající zpravidla stromovou strukturu. V PHD se na nejvyšší úrovni vyskytuje pouze jediný proces představující celý analyzovaný systém, nižší úrovně diagramu představují dekompozice procesů o úroveň výš. Následně je třeba pro jednotlivé procesy na nejnižší úrovni PHD vytvořit procesní mapy.

V okamžiku, kdy je hierarchie procesů zaznamenána v podobě PHD a pro každý proces na nejnižší úrovni je vytvořena právě jedna procesní mapa se provádí přechod od procesního modelu k objektově orientovanému modelu systému. Jednotlivé „elementární“ procesy jsou ztotožněny s případy užití systému. Obvykle jednomu procesu na PHD odpovídá jeden případ užití na use case diagramu celého systému. Ze vztahů procesů na PDH lze zároveň určit vzájemné vztahy jednotlivých případů užití (extend a include). Procesní mapy jsou následně použity jako základ scénářů nalezených případů užití. Převod diagramu hierarchie procesů na diagram případů užití znázorňuje obrázek 3.



Obr. 3: Transformace diagramu hierarchie procesů na use case diagram

Ve druhé fázi je provedena objektová analýza a návrh systému, při níž se vychází z nalezených případů užití. Obecně lze bez větších problémů postupovat podle libovolné „use case driven“ metodiky. V konkrétním případě tvorby reportingového systému postupoval řešitelský tým podle Rumbaughtovy OMT (Object Modeling Technique).

Výše zmiňovaná modifikace metodiky SELECT Perspective spočívala v tom, že na rozdíl od původní verze metodiky byl jazyk UML použit nejen ve fázi objektové analýzy systému, ale také v první etapě procesního modelování. Konkrétně se jedná o využití UML diagramů činností, které byly použity jako základní pohled při tvorbě procesních map. Vedle standardního použití metodikou OMT sloužily dále prostředky jazyka UML i pro tvorbu speciálních modelů. Konkrétně se jedná o využití stavových diagramů pro modelování stavu reportů během procesu reportingu. Vztahy mezi stavy procesů a procesními mapami byly modelovány pomocí časových diagramů, což je opět nástroj UML zavedený od verze 2.

4 Skupiny uživatelů

Používání ISR se mění dle toho, do jaké skupiny daný uživatel spadá. Na tomto místě uveden stručný výčet uživatelských rolí, které systém rozlišuje:

- *Uživatelé* – základní skupina uživatelů. Tato skupina může informace uvedené v systému jen číst, nemá žádné právo na editaci či vložení nějakých záznamů
- *Experti* – skupina osob plnící reportingové povinnosti. Má stejná práva jako uživatelé a navíc může pracovat se údaji o reportingových povinnostech.
- *Editoři* – speciální uživatelé, kteří mají oproti předchozím navíc právo pracovat s reporty, tzn. odesílat a vkládat nové, upravovat u povinností relevantní atributy apod.
- *Řešitelé* – tato skupina byla vytvořena speciálně pro členy řešitelského týmu z Masarykovy univerzity v Brně. Po nasazení systému do plného provozu byli tito uživatelé zrušeni.
- *Administrátoři* – skupina s nejvyššími právy. Administrátor může v systému dělat prakticky cokoliv, co systém dovoluje. Od vkládání nových záznamů, přes jejich schvalování až po mazání.

5 Závěr

Počátkem roku 2004 byl zahájen provoz prototypu tohoto systému. Systém byl provozován na serveru Centra biostatistiky a analýz MU (CBA MU), což bylo společné pracoviště Lékařské a Přírodovědecké fakulty Masarykovy univerzity v Brně. K systému měli přístup jeho vývojáři z MU a dále vybraní odborníci z resortu MŽP, kteří systém testovali a připomínkovali. Na základě jejich připomínek byl systém nadále rozšiřován a modifikován.

Vytvořený prototyp byl rovněž konzultován a koordinován s oficiálními zahraničními partnery, jako jsou EEA, OECD, UNEP, EUROSTAT, SAŽP, UBA Rakousko a Německo a s řešiteli twinningového projektu CENIA. Návrh prototypu byl dále posuzován z pohledu potřeb dalších aktivit, jimiž jsou například GMES, INSPIRE a realizované projekty 5. a 6. rámcového programu EU či další mezinárodní vědeckovýzkumné programy.

Navržený systém reflektuje požadavky i ve směru informování veřejnosti podle nových evropských legislativních požadavků. Přitom musí poskytovat informační základnu i pro strategické plánování ve spolupráci s komerční sférou. Integrovaný reporting a požadavky na něj kladené však zůstávají klíčovou oblastí. ISR rovněž respektuje informační pyramidu MDIAR definovanou EEA i národní a mezinárodní standardy.

Současně byla formalizována komunikační stránka s využitím odpovídajících nástrojů (XML, ODBC) a připraven rámec pro zabezpečení celého systému včetně návrhu skupin uživatelů a administrátorských nástrojů. Celý systém přitom směřuje k plné integraci s nejnovějšími komunikačními nástroji a je navržen s dostatečnou modularitou umožňující další vývoj. V létě 2006 byl systém předán CENIA a uveden do plného provozu. Zároveň s tím byl ukončen provoz prototypu na serveru MU.

6 Literatura

- [1] Hřebíček J., Pitner T., Ráček J.: Analysis of Environmental Information Management in Czech Republic. In International Symposium on Environmental Software Systems (ISESS'05). International Federation for Information Processing, 2005.
- [2] Hřebíček J., Pitner T., Ráček J.: Analýza a návrh environmentálních datových modelů a vnějších rozhraní JISŽP kompatibilních s EU - Závěrečná zpráva o průběhu řešení projektu v roce 2005. Masarykova univerzita, Brno, 2005.
- [3] Pick, T., Saarenmaa, H., Lounamaa, K.: Management of Environmental Information in the European Information and Observation Network (EIONET), in Proc. Int. Conf. On Dublin Core and Metadata Applications 2001, National Institute of Informatics, Tokyo, 2001.
- [4] Reporting Obligations Database (ROD), <http://rod.eionet.eu.int>.
- [5] Rumbaugh, J.: Object Oriented modelling and Design, Prentice-Hall International, New Jersey, 1991.
- [6] Rumbaugh, J., Jacobson, I., Booch, G.: The Unified Modelling Language Reference Manual, 2nd ed., Addison-Wesley, Boston, 2004.

Příspěvek vznikl za podpory grantu Ministerstva životního prostředí České republiky č. SM/250/3/03 „Analýza a návrh environmentálních datových modelů a vnějších rozhraní JISŽP kompatibilních s EU“.

Odhady nákladů workflow projektů metodou FPA

Jana Sedláčková

Vysoké učení technické v Brně, Fakulta informačních technologií
Božetěchova 2, 602 00 Brno
jana.sedlackova@gmail.com

František Huňka

Ostravská univerzita, Přírodovědecká fakulta
30. dubna 22, 701 03 Ostrava
frantisek.hunka@osu.cz

Abstrakt

Příspěvek se zabývá problematikou stanovení odhadu nákladů pro vyvíjené workflow projekty. Představuje návrh na modifikaci jedné z klíčových metod, jež je určena pro odhady nákladů klasických softwarových projektů – metody FPA, pro workflow projekty. Je zde popsána problematika stanovení počtu neupravených funkčních bodů, z nichž se následně počítají upravené funkční body a náklady na vývoj celého projektu.

Abstract

This paper deals with the cost estimation methods of workflow development projects. It presents suggestion of the modification of one main method – FPA for the workflow projects, because this method was determined for costs estimation of the classical software projects. In this paper there is described estimation question of unadjusted function point count, of these are computed adjusted function points subsequently and cost for the whole project development.

Klíčová slova

Funkční body, workflow systém, upravené funkční body, neupravené funkční body, IN, OUT, ENQ, FILE a FILEE.

Keywords

Function points, workflow system, adjusted function points, unadjusted function points, IN, OUT, ENQ, FILE a FILEE.

1 Úvod

Procesní přístup k řízení institucí a podniků se stává stále běžnějším. Tento způsob vytlačuje starší techniku funkčního pohledu, jež dělila práci mezi funkční jednotky vytvořené na základě příslušných dovedností. Takto vzniklé jednotky vytvářely jen svoji konkrétní činnost nezávisle na celkovém výsledku.

Procesní způsob řízení je orientován na výsledek práce, tedy na konkrétní produkt. Práce není vykonávána separátně v jednotlivých funkčních jednotkách, jako tomu bylo u funkčního přístupu, ale dílčími jednotkami protéká. Systém je řízen základními potřebami zákazníka obvykle formou interakcí a rozhraní, což představuje řízení produktů.

Podobně jako u vývoje informačních systémů i workflow projekty sebou přinášejí nemalá finanční rizika spojená s jejich vývojem. Vyžadují proto důkladné plánování a velkou míru zodpovědnosti při rozhodování. Je třeba co nejpřesněji analyzovat náklady, které sebou vyvíjený produkt přináší. Běžně používané metody pro stanovení odhadu nákladů pro vyvíjený softwarový projekt jsou však určeny klasickým softwarovým systémům, nikoliv však čím dál častěji používaným workflow projektům.

Pro stanovení odhadu nákladů softwarových projektů se často používá metoda funkčních neboli FPA, jež je založena na popisu funkcionality výsledného produktu. Metodu je možné použít pro výpočet očekávaných odhadů workflow projektů, avšak s jistými modifikacemi.

2 Odhad metodou funkčních bodů

Metoda FPA se zabývá aplikační oblastí a neměří zdrojový kód (viz [1]). Při samotném výpočtu odhadu nákladu workflow procesu je třeba proto nejprve spočítat tzv. neupravené funkční body, tedy identifikovat základní funkční typy workflow procesu a jejich příslušné složitosti. Počet neupravených funkčních bodů *NFP* je pak dán součtem všech váhových hodnot (viz [2]). Z neupravených funkčních bodů se dále počítají tzv. upravené funkční body, jež jsou přizpůsobeny danému konkrétnímu workflow procesu a prostředí, ve kterém se vyvíjí. Stanoví se vlivy 14ti faktorů obecných charakteristik systému včetně příslušné hodnoty vlivu na aplikaci. Tímto způsobem se vypočte faktor přizpůsobení a vynásobí hodnotou neupravených funkčních bodů. Tento výsledek dá počet upravených funkčních bodů *FP*, na jehož základě se stanoví pracnost a cena jednoho funkčního bodu a dále i celkové náklady pro vývoj workflow procesu (viz [3]).

3 Výpočet neupravených funkčních bodů

Podle zdroje [1], [3] se neupravené funkční body vztahují k transakčním funkcím, mezi které patří externí vstupy (*IN*), externí výstupy (*OUT*) a externí dotazy (*ENQ*), nebo k datovým funkcím, což jsou vnitřní logické soubory (*FILE*) a soubory vnějšího rozhraní (*FILEE*).

3.1 Funkční body vztažené k transakčním funkcím

IN – externí vstupy

Za externí vstupy se započítají každá unikátní uživatelská data nebo zadání uživatelských povelů, které vstoupí přes externí rozhraní do aplikace a přidá, mění, ruší nebo jinak pozmění data v externím logickém souboru. Započte se rovněž řídicí informace, která vstoupí přes aplikační hranici a zajistí soulad s funkcí specifikovanou uživatelem. Externí vstupy se u workflow procesu rozpoznají z činností procesu. Například se může jednat o formuláře, respektive jejich čtení, což odpovídá jednomu externímu vstupu. Nezapočítávají se však činnosti, které jsou automatizované, ale jen ty, na jejichž vstupu je účastník.

OUT – externí výstupy

Za externí výstup se považují všechna unikátní uživatelská data nebo data řídicí, která opouští externí hranici daného systému. Externí výstup je považován za unikátní, pokud má odlišná data nebo pokud jiná aplikace vyžaduje odlišnou logiku zpracování oproti jiným externím výstupům. U workflow systému se mezi externí výstupy započítávají i výstupy, jež pochází z automatických činností. Příkladem takového výstupu může být například tisk faktury.

ENQ – externí dotazy

Jako vnější dotaz se započte každá unikátní vstupně/výstupní kombinace, kde vstup je příčinou a generuje výstup. Vnější dotaz je považován za unikátní, pokud se od ostatních dotazů odlišuje typem výstupních datových elementů nebo pokud vyžaduje odlišnou logiku zpracování v porovnání s ostatními externími dotazy.

3.2 Funkční body vztažené k datovým funkcím

FILE – vnitřní logické soubory

Za vnitřní logický soubor se považuje každá velká logická skupina uživatelských dat nebo informací použitých pro řízení aplikace. Započítá se každý logický soubor, nebo v případě DB každé logické seskupení dat z pohledu uživatele, které je vytvořeno, používáno nebo udržováno aplikací. Spíše než fyzické soubory je započteno každé logické seskupení dat tak, jak je viděno z pohledu uživatele. Nezapočtou se však soubory, jež nejsou přístupné uživateli prostřednictvím vnějšího výstupu nebo dotazu a které nejsou nezávisle udržovány. Příklad výskytu interního souboru je třeba uvnitř formuláře.

FILEE – soubory vnějšího rozhraní

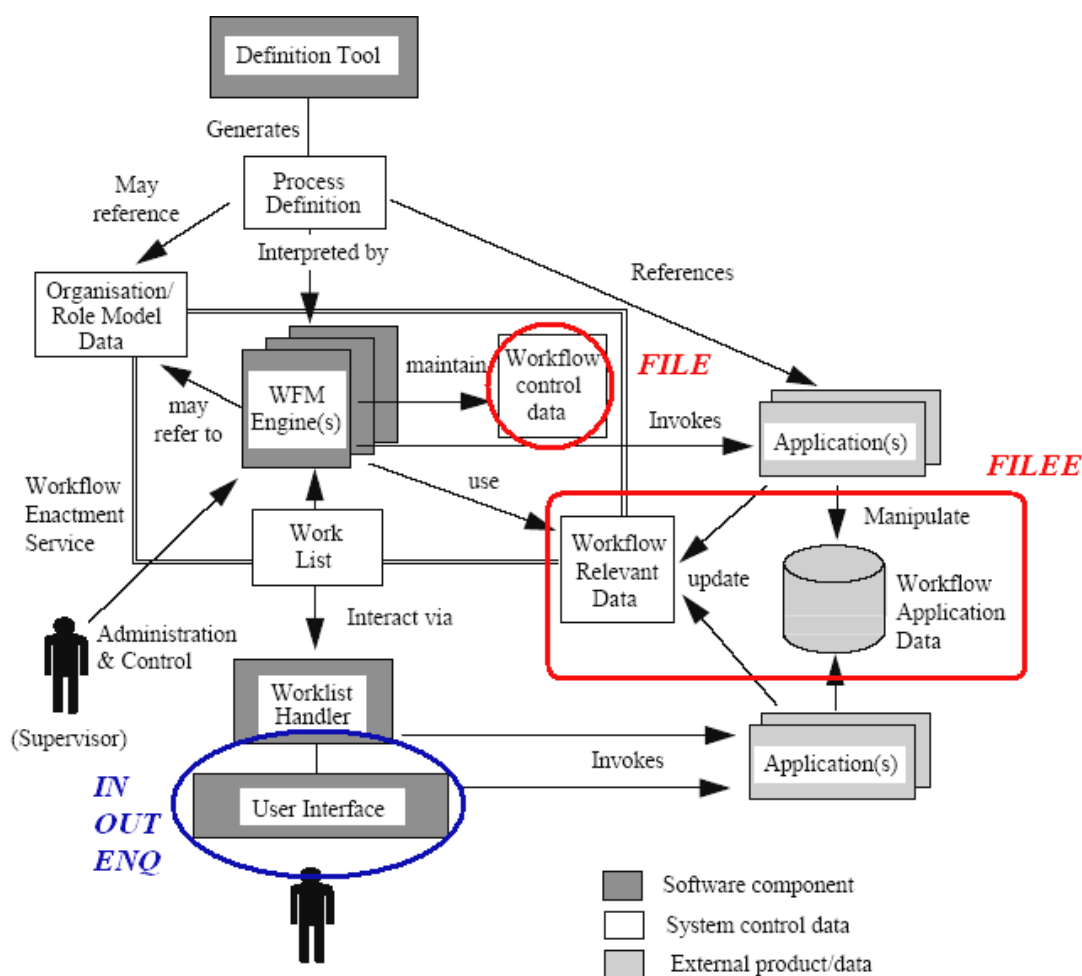
Jako soubor vnějšího rozhraní se označí každá velká logická skupina uživatelských dat nebo řídicí informace používané aplikací. Informace však musí být udržována jinou aplikací. Započte se i každý logický soubor nebo logická skupina dat z pohledu uživatele, dále každá velká logická skupina uživatelských dat nebo řídicí informace, která je extrahována aplikací z jiné aplikace ve formě souboru externího rozhraní.

3.3 Identifikace transakčních a datových funkcí ve workflow procesu

Hlavní problematikou stanovení neupravených funkčních bodů je identifikovat příslušné externí vstupy, externí výstupy, externí dotazy, vnitřní logické soubory a soubory vnějšího rozhraní ve workflow procesu (viz [4]).

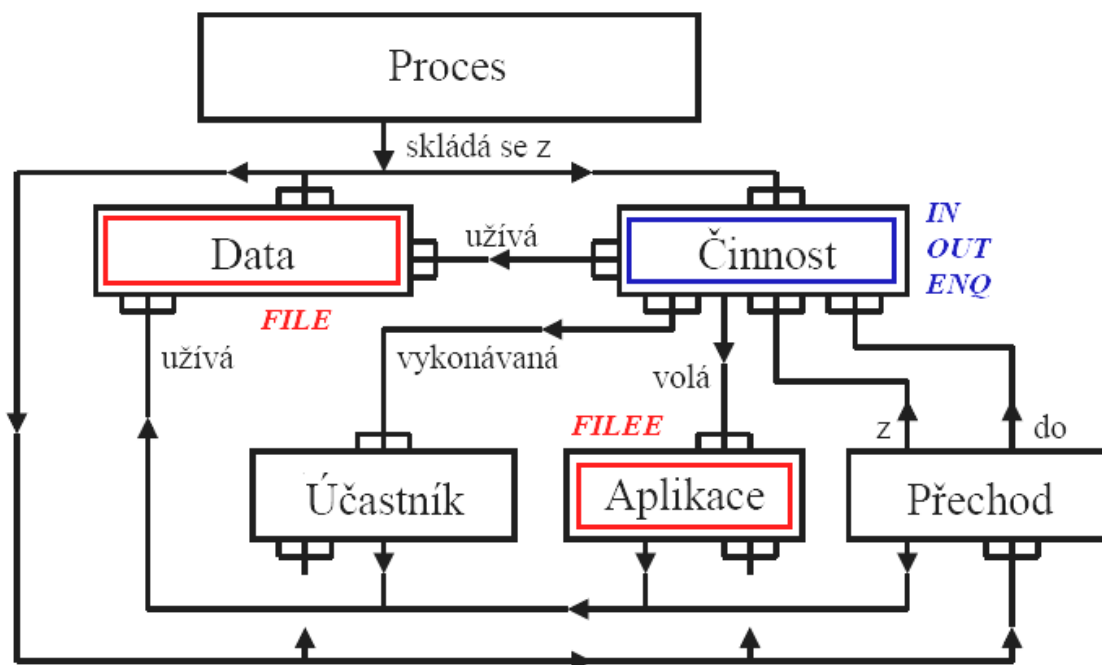
Následující schéma představuje samotnou architekturu workflow systému (viz [5], [6]). Je zde možné vymezit oblasti, ve kterých je třeba hledat externí vstupy *IN*, výstupy *OUT* a dotazy *ENQ*, rovněž vnitřní logické soubory *FILE* a soubory vnějšího rozhraní *FILEE*.

Externí vstupy *IN*, externí výstupy *OUT* a externí dotazy *ENQ* se tedy v základní struktuře workflow systému nacházejí v oblasti uživatelského rozhraní. Oblast pro identifikaci vnitřních logických souborů *FILE* je vymezena v prostoru výskytu workflow kontrolních dat (Workflow Control Data). Oblast, v níž je možné rozpoznat soubory vnějšího rozhraní *FILEE*, se nachází v prostoru mezi jednotlivými aplikacemi, kde proudí příslušná workflow data (Workflow Relevant Data) a workflow aplikační data (Workflow Application Data).



Obr. 1: Základní struktura workflow systému

Druhé schéma zobrazuje meta-model workflow procesu (viz [6]) rovněž s vyznačenými místy pro identifikace *IN*, *OUT*, *ENQ*, *FILE* a *FILEE*. Oblast výskytu externích vstupů *IN*, externích výstupů *OUT* a externích dotazů *ENQ* se nachází v entitě workflow činnost. V případě *IN* a *ENQ* se jedná pouze o činnosti manuální, tedy ne automatizované. U *OUT* se berou v potaz činnosti manuální i automatické. Oblast pro identifikaci vnitřních logických souborů *FILE* se nachází v entitě workflow data. Oblast výskytu souborů vnějšího rozhraní *FILEE* se nachází uvnitř entity workflow aplikace. Konkrétně je vymezují parametry příslušné workflow aplikace.



Obr. 2: Meta-model workflow procesu

Všechny specifikované *IN*, *OUT*, *ENQ*, *FILE* a *FILEE* aplikace je potřeba rozřadit do skupin podle své složitosti. Počet prvků každé skupiny se vynásobí příslušnou váhou a následně se vše sečte. Výsledný součet dává počet neupravených funkčních bodů (*NFP*) (viz [3]).

4 Výpočet upravených funkčních bodů

Z neupravených funkčních bodů *NFP* jsou dále vypočteny upravené funkční body *FP* v závislosti na prostředí, ve kterém je daný proces realizován. Výpočet bude dále probíhat standardním způsobem jako při odhadu nákladů klasických softwarových projektů (viz [3]). Z upravených funkčních bodů je možné dále stanovit například dobu vývoje procesu, vynaložené úsilí a počet osob, které je potřeba na daný projekt alokovat.

5 Závěr

Odhadování nákladů potřebných pro realizaci softwarového projektu je považováno za běžnou a nezbytnou činnost, kterou se zabývá každý manažer pověřený vedením vývoje softwarového projektu. Je proto potřeba stanovit tyto odhady i pro nově nastupující workflow systémy. Protože se jedná o poměrně nový přístup, nejsou doposud pro odhad nákladů workflow systémů vyvinuty žádné speciální metodiky, které by autorům tohoto příspěvku byly známy. Je proto třeba pro tento účel upravit některé z metod určených pro odhad nákladů klasických softwarových projektů.

6 Literatura

- [1] International Function Points User Group: Function Points Counting Practices Manual, 1999.
- [2] Král, J.: Informační systémy. Science, Brno, 1998.
- [3] Sedláčková, J.: Odhady nákladů SW projektů. Masarykova univerzita v Brně, Fakulta informatiky, Brno, 2005. Diplomová práce.
- [4] Sedláčková, J.: Odhady času a úsilí workflow projektů. Masarykova univerzita v Brně, Fakulta informatiky, Brno, 2006. Rigorózní práce.
- [5] Ráček, J.; Ministr, J.; Fiala, J.: Workflow Model and Process Interchange Formats of Public Environmental Administration. In: Information Technologie in Environmental Engineering. ICSC-NAISO Academic Press, Canada, 2003.
- [6] Hollingsworth D.: The Workflow Reference Model, Issue 3.0, Workflow Management Coalition, 1995.

Analýza a návrh matematického modelu na podporu rozhodování ve veřejných projektech

Jana Soukopová

Masarykova univerzita v Brně, Ekonomicko-správní fakulta
Lipová 41a, 602 00 Brno
soukopova@econ.muni.cz

Abstrakt

Jedním z aktuálních problémů současnosti je zefektivnění alokace veřejných výdajů. Článek je zaměřen na systemizaci poznatků z oblasti hodnocení veřejných projektů. Tato oblast je velmi rozsáhlá a proto je článek zaměřen na oblast rozhodovací analýzy a modely hodnocení a výběru veřejných projektů. V článku bude definován matematický model rozhodování ve veřejných projektech, který je dále možné použít pro elektronické zpracování dat při hodnocení veřejných projektů.

Abstract

One of actual problems is efficiency in public expense allocation. This article is aimed on systematisation of knowledge in the field of public project's valuation. Because this field is very extensive, this work is particularly aimed on field of decision analysis and models of public projects valuating and assesment. In this article will be defined mathematical model of public projects decision, which is possible to use further to electronic data processing in public project valuation.

Klíčová slova

Veřejné projekty, analýza veřejných projektů, metody hodnocení veřejných projektů, matematický model rozhodovací situace.

Keywords

Public projects, analysis of public projects, valuation methods for public projects, mathematical decision model.

1 Úvod

Jedním z hlavních problémů fungování veřejného sektoru, je zajištění efektivnosti veřejných výdajů. Panuje široce sdílené přesvědčení, že hospodaření s veřejnými prostředky svádí k nesrovnatelně méně hospodárnému a v konečném důsledku méně efektivnímu použití, než když se jedná o nakládání s vlastními prostředky soukromých subjektů. Příčin tohoto jevu je celá řada. Tento článek je věnován matematickým a ekonomickým nástrojům, které je možné použít k tomu, aby se tato tendence k neefektivnosti projevovala pokud možno co nejméně.

Veřejná správa se při plánování a uskutečňování svých výdajů setkává v podstatě se zcela stejnými otázkami jako každý jiný ekonomický subjekt. Co je potřeba nejvíce? Co nám přinese největší užitek? Co pořídit hned a s čím ještě počkat? Kam umístit volné finanční prostředky? Na rozdíl od běžných domácností, soukromých investorů či podnikatelů, tedy těch, kdo disponují soukromými finančními prostředky, rozhodují zástupci státu a místní samosprávy o užití prostředků pocházejících z veřejných příjmů a odpovídají za účelnost a hospodárnost jejich využití občanům. Při rozhodování o výběru mezi různými variantami veřejných projektů je faktor nejistoty úplně stejný, avšak informace o "ceně" vstupů a výstupů jsou často nedostatečné. Převládající netržní prostředí, přítomnost "veřejného zájmu", potřeba zohledňovat sociální aspekty daných rozhodnutí, dopady externalit, nezbytnost zachování demokratičnosti procesu přijímání rozhodnutí, výrazná limitovanost dostupných zdrojů - to vše vytváří z rozhodování ve veřejném sektoru složitý proces.

2 Rozhodovací situace hodnocení veřejných projektů

Protože do kategorie analýzy (hodnocení) veřejných projektů spadají různorodé úlohy, není dost dobře možné předložit univerzální teorii a z ní vyplývající rozhodovací algoritmus, vhodný pro všechny typy rozhodování. Pro výběr vhodné metody hodnocení projektů je tedy nutné uvést klasifikaci rozhodovacích situací vyplývajících z identifikace souboru veřejných projektů.

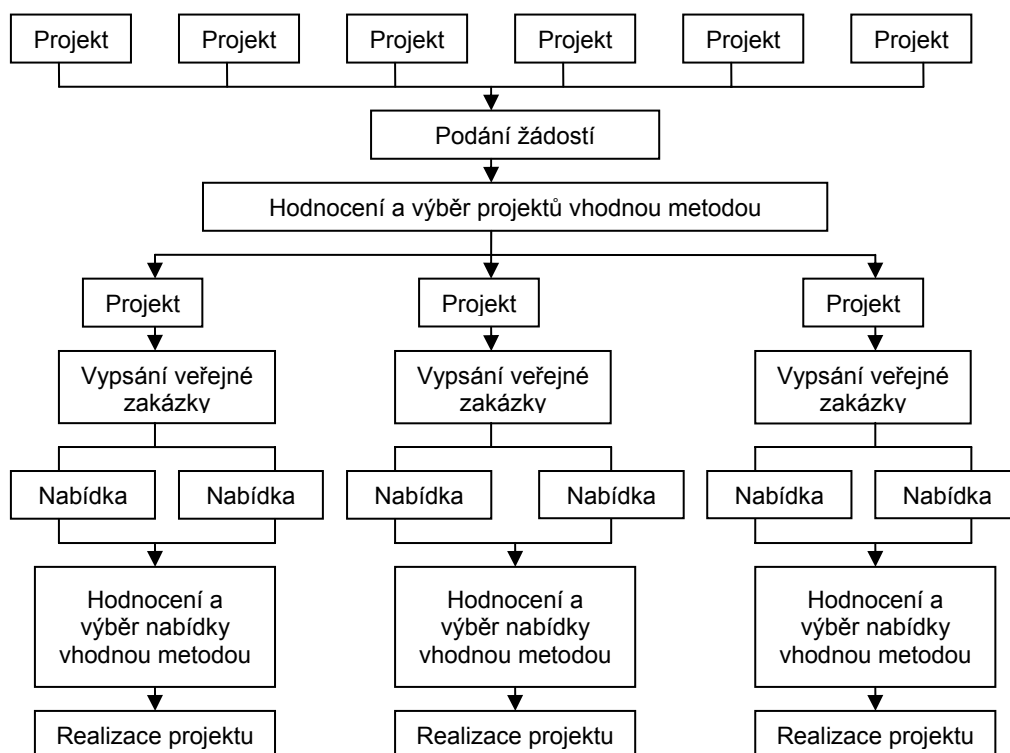
Oblast analýzy veřejných projektů v ČR se dá zjednodušit na následující dvě rozhodovací situace:

1. *Hodnocení nezávislých a vzájemně se vylučujících projektů.*

Volba mezi takovými projekty se vždy provádí způsobem „všechno nebo nic“, neboť většinou není možné realizovat z každého jen část. Uvedená situace platí také pro hodnocení jednotlivých nabídek veřejné zakázky.

2. *Hodnocení nezávislých, ale vzájemně se nevylučujících projektů.*

Jedná se např. o hodnocení projektů ze Strukturálních fondů EU. Cílem je nalézt optimální kombinaci projektů nebo jejich optimální velikost odpovídající disponibilním zdrojům, popř. i vybrat jeden „nejlepší“ projekt. Proces zadávání a hodnocení v tomto případě ukazuje obr. 1.



Obr. 1: Proces hodnocení veřejných projektů financovaných ze SF EU (Soukopová 2005)

Modelování je účelové zobrazování vyšetřovaných vlastností zkoumané skutečnosti pomocí vhodně zvolených vlastností modelu. Jedná se tedy o reprodukci vybraných vlastností studovaného objektu na modelu, tj. analogickém objektu simulujícím chování a vlastnosti původního studovaného objektu. Modely jsou tříděny podle zobrazovaných vlastností zkoumané skutečnosti na:

- substancionální (shodují se s originálem ve všech nebo v některých základních vlastnostech),
- strukturní (zobrazují především vnitřní uspořádání prvků z nichž se sestává zkoumaná skutečnost),
- funkční (zobrazují různé funkce) a

- smíšené (kombinace předchozích případů).

Z hlediska formálně konstruktivního třídíme modely na matematické (vytvořené pomocí matematických výrazových prostředků) a schematické (všechny nematematické modely). (více např. Beltrami 1998).

3 Matematický strukturní model hodnocení veřejných projektů

Při tvorbě strukturního matematického modelu rozhodovací situace při hodnocení veřejných projektů je nutné vycházet z uvedených dvou typů rozhodování a definovat pro ně různé rozhodovací modely:

- matematický strukturní model rozhodovací situace pro hodnocení nezávislých a vzájemně se vylučujících projektů;
- matematický strukturní model rozhodovací situace pro hodnocení nezávislých, ale vzájemně se nevylučujících projektů.

3.1 Matematický strukturní model rozhodovací situace pro hodnocení nezávislých a vzájemně se vylučujících projektů

Pro matematický popis tohoto modelu rozhodovací situace je třeba definovat:

$X = \{x_1, x_2, \dots, x_n\}$	množinu n nezávislých a vzájemně se vylučujících projektů (popř. nabídek veřejné zakázky),
$E = \{e_1, e_2, \dots, e_q\}$	množinu q hodnotitelů (expertů),
$K = \{k_1, k_2, \dots, k_k\}$	množinu k kritérií,
$G = \{g_{jr}\}$	matici funkcí ohodnocení projektu x_i , ($i = 1, \dots, n$) podle kritéria k_j , ($j = 1, \dots, k$) r -tým hodnotitelem e_r , ($r = 1, \dots, q$), kde pro jednoduchost uvažujeme, že $g_{jr}(x_i) = S_{ijr}$ a
$d = d(S_{ijr})$	metodu (funkci či algoritmus) hodnocení a výběru projektu vycházející z hodnot trojrozměrné matice $S = (S_{ijr})$ hodnocení. Jejím výsledkem je alespoň jeden projekt např. x_{opt} .

Matematický strukturní model rozhodovací situace pak lze obecně popsat jako matematickou strukturu:

$$R = [X, E, K, G, d, x_{opt}]. \quad (3.1)$$

3.2 Matematický strukturní model rozhodovací situace pro hodnocení nezávislých, ale vzájemně se nevylučujících projektů

Pro matematický strukturní model této rozhodovací situace je nutné na rozdíl od modelu (3.1) zadat odlišnou množinu $Y = \{y_1, y_2, \dots, y_n\}$ n nezávislých, ale vzájemně se nevylučujících projektů a matici funkcí $P = \{p_{jr}\}$ pro hodnocení projektů y_i ($i = 1, \dots, n$) podle kritéria k_j ($j = 1, \dots, k$) r -tým hodnotitelem e_r , ($r = 1, \dots, q$), kde funkce hodnocení $p_{jr}(y_i) = Q_{ijr}$ definují trojrozměrnou matici rozhodnutí $Q = (Q_{ijr})$. Metoda (funkce) hodnocení a výběru projektu $d(Q_{ijr})$ má za závislé proměnné hodnoty matice Q a jejím výsledkem je množina m vybraných projektů $D = \{y_{i1}, y_{i2}, \dots, y_{im}\}$, kde $1 < i_1 < \dots < i_m$, $1 < i, j < n$, $j = 1, \dots, m$.

Matematický strukturní model rozhodovací situace pak lze popsat následovně

$$R = [Y, E, K, P, d, D]. \quad (3.2)$$

Je důležité zdůraznit, že pro hodnocení nezávislých a vzájemně se vylučujících projektů (alternativ veřejného projektu, nebo nabídek veřejné zakázky) je možné použít všechny metody hodnocení a záleží pouze na tom, jaká použijeme hodnotící kritéria (zda peněžní (obecné finanční metody) či nepeněžní (nákladově-výstupové metody)) a zda jedno (jednokritériální metody) či více kritérií

(vícekriteriální metody). Podle výběru hodnotících kritérií pak budeme vybírat metodu hodnocení (více např. Soukopová 2005).

Při hodnocení nezávislých, ale vzájemně se nevylučujících veřejných projektů je situace složitější. Je možné použít všechny jednokriteriální metody a podle charakteru vstupů a výstupů zvolit vhodnou metodu. Použití vícekriteriálního hodnocení však naráží na určité problémy. Pokud jsou vzájemně se vylučující projekty zaměřeny na různé oblasti veřejného sektoru, není téměř možné pro jejich hodnocení použít vícekriteriální metody. Je to z toho důvodu, že se obtížně stanovují kritéria, která by vyhovovala rozdílným oblastem veřejného sektoru a přitom byla pro hodnocení smysluplná. Použití vícekriteriálního hodnocení je možné u vzájemně se nevylučujících projektů, které jsou podobně zaměřeny (pokud jde například o hodnocení environmentálních veřejných projektů, můžeme se při stanovení kritérií opřít o hledisko specializace projektu a následně volit environmentální kritéria.).

V současné době je tento problém vyřešen rozdělením veřejných projektů žádající o prostředky z veřejných rozpočtů nebo rozpočtů nadnárodních institucí na základě specializací programů, ze kterých jsou veřejné prostředky financovány. V praxi jsou tedy hodnoceny srovnatelné projekty. Proto tedy dále uvádím zobecněný matematický model rozhodovací situace hodnocení a výběru veřejných projektů.

4 Zobecněný matematický strukturní model rozhodovací situace hodnocení a výběru veřejných projektů

Vzhledem k tomu, že v praxi jsou hodnoceny projekty pouze v rámci určité oblasti veřejného sektoru (životní prostředí, doprava, rozvoj regionů apod.), tedy specializované veřejné projekty, je tyto možné hodnotit všemi hodnotícími metodami. Díky tomu je možné uvedené rozhodovací situace sloučit do jedné. Pro účely tohoto článku tedy sloučím rozhodovací situace a zavedu pojem **hodnocení variant veřejných projektů**, kdy **variantou** budu myslet nezávislé vzájemně vylučitelné i vzájemně nevylučitelné projekty.

4.1 Matematický strukturní model rozhodovací situace pro hodnocení a výběr variant veřejných projektů

Obecný matematický strukturní model rozhodovací situace pro hodnocení variant veřejných projektů pak zapíšeme

$$R = [A, E, K, F, d, D] \quad (4.1)$$

kde

$A = \{a_1, a_2, \dots, a_n\}$	je množina n variant projektů,
$K = \{k_1, k_2, \dots, k_k\}$	je množina k kritérií,
$E = \{e_1, e_2, \dots, e_q\}$	je množina q hodnotitelů (expertů),
$F = \{f_{ir}\}$	je matice funkcí hodnocení variant a_i ($i = 1, \dots, n$) projektů podle kritéria k_j ($j = 1, \dots, k$) r -tým hodnotitelem e_r ($r = 1, \dots, q$), kde $f_{jr}(a_i) = S_{ijr}$,
$d = d(S_{ijr})$	je metoda (funkce) výběru projektu z hodnot trojrozměrné matice $S = (S_{ijr})$ hodnocení, jejímž výsledkem je množina D vybraných projektů,
$D = \{a_{i1}, a_{i2}, \dots, a_{im}\}$	je množina m vybraných variant projektu, kde $1 < \dots < i_1 < i_2 < \dots < i_m$, $1 < i, j < n, j = 1, \dots, m$.

4.2 Matematický strukturní model rozhodovací situace při jednokriteriálním hodnocení

Vzhledem k tomu, že při jednokriteriálním rozhodování jsou veřejné projekty hodnoceny na základě jediného kritéria, je matematický strukturní model rozhodovací situace v tomto případě následující:

$$R = [A, E, k, F, d, D] \quad (4.2)$$

kde

$A = \{a_1, a_2, \dots, a_n\}$	je množina n variant projektů,
k	je hodnotící kritérium,
$E = \{e_1, e_2, \dots, e_q\}$	je množina q hodnotitelů (expertů),
$F = \{f_1, f_2, \dots, f_q\}$	je vektor metod (funkcí) hodnocení variant a_i ($i = 1, \dots, n$) projektů podle kritéria k r -tým hodnotitelem e_r ($r = 1, \dots, q$), kde $f_r(a_i) = S_{ir}$.
$d = d(S_{ir})$	je metoda (funkce) výběru projektu z hodnot matice $S = (S_{ir})$ hodnocení jejímž výsledkem je množina D vybraných projektů,
$D = \{a_{i_1}, a_{i_2}, \dots, a_{i_m}\}$	je množina m vybraných variant projektu, kde $1 < i_1 < i_2 < \dots < i_m$, $1 < i_j < n, j = 1, \dots, m$.

4.3 Matematický strukturní model rozhodovací situace při vícekritériálním hodnocení

Matematický strukturní model rozhodovací úlohy vícekritériálního hodnocení variant pro hodnocení variant veřejných projektů jsem zúžila na případ hodnocení jedním rozhodovatelem e , který lze zapsat jako:

$$R = [A, e, K, F, d, D] \quad (4.3)$$

kde

$A = \{a_1, a_2, \dots, a_n\}$	je množina n variant projektů,
$K = \{k_1, k_2, \dots, k_k\}$	je množina k kritérií,
$F = \{f_1, f_2, \dots, f_k\}$	je vektorová funkce hodnocení projektu a_i ($i = 1, \dots, n$) podle kritéria k_j ($j = 1, \dots, k$) rozhodovatelem (hodnotitelem) e , kde $f_j(a_i) = y_{ij}$, vyjadřuje hodnocení i -té varianty podle j -tého kritéria.
$d = d(y_{ij})$	je metoda (funkce) výběru projektu z hodnot kritériální matice $Y = (y_{ij})$ hodnocení, kterou vybíráme množiny vícekritériálních metod hodnocení podle toho, zda provádíme dílčí hodnocení variant či párové srovnávání variant veřejných projektů (více např. Fiala, Jablonský a Maňas, 1994)
$D = \{a_{i_1}, a_{i_2}, \dots, a_{i_m}\}$	je množina m vybraných variant projektu, kde $1 < i_1 < i_2 < \dots < i_m$, $1 < i_j < n$, $j = 1, \dots, m$.

5 Závěr

Rozhodování o výběru nejvhodnějšího veřejného projektu je složitou úlohou při které hraje roli řada faktorů. V klasickém ekonomickém modelu lze jednoznačně určit potřebné vstupy, definovat jejich optimální využití a změřit výsledek celého projektu poměřením vstupů a výstupů. Pro většinu veřejných projektů a hlavně projektů z oblasti ochrany životního prostředí nelze takovou analýzu provést. Je tedy velmi obtížné hodnotit efektivnost veřejných projektů z určitých oblastí veřejného sektoru jako je školství, zdravotnictví a speciálně životní prostředí. Díky absenci kritéria zisku je daleko obtížnější odhadnout, zda je daný projekt efektivní. Obtíže při hodnocení činnosti veřejných projektů jsou násobeny tím, že tyto projekty často sledují více cílů najednou.

Matematický strukturní model rozhodovací situace hodnocení a výběru veřejných projektů přispívá k řešení této složité situace. Je základem pro vytvoření pro informačního systému na podporu rozhodování ve veřejných projektech a je prvním krokem v životním cyklu tvorby takového systému.

6 Literatura

- [1] Beltrami, E.: Mathematics for Dynamic Modeling. Second Edition, Academic Press, 1998.
- [2] Fiala, P., Jablonský, J., Maňas, M.: Vícekriteriální rozhodování, VŠE, Praha, 1994.
- [3] Ochrana, F.: Veřejné zakázky – Metody a metodika efektivního hodnocení a výběru. Ekopress, Praha, 2004.
- [4] Soukopová, J.: 2005, Metody hodnocení veřejných projektů. Provozně ekonomická fakulta, Mendelova zemědělská a lesnická univerzita v Brně, Brno, 2005. Disertační práce.

Matematický model růstu populací mšic a afidofágů na kukuřičném poli

Michal Škrdla

Masarykova univerzita, Fakulta informatiky
Botanická 68a, 602 00 Brno
xskrdla@fi.muni.cz

Abstrakt

Tento článek se zabývá aktuálním vývojem v oblasti modelování růstu populací mšic a afidofágů na kukuřičném poli, což by se mělo stát tématem disertační práce autora. Po zhrnutém uvedení do problému modelování a cíle práce je prezentován aktuálně vyvíjený software, naměřená data a představen plánovaný směr dalšího výzkumu.

Abstract

This paper is focused on actual state of growth model of aphides's and aphidophagus's population in maize. This theme should be subject of author's thesis. After skim through problems and goals of work, there is presented currently developing software, measurements and is introduced future trend of the next research.

Klíčová slova

Růstový model, matematické modelování, J2EE, EJB 3.0, analýza rozptylu, entomologie, ochrana rostlin, přirození nepřátelé, mšice.

Keywords

Growth model, mathematical modelling, J2EE, EJB 3.0, entomology, plant protection, maize, natural predators, aphids.

1 Úvod

V době, kdy většina států EU bojuje s nadměrnou kontaminací přírody, je nezbytné dále životní prostředí nezatěžovat, ale vrátit se k biologicky přirozeným procesům pěstování rostlin a nevytvářet tak další ekologickou zátěž používáním agrochemikálií, zejména pesticidů a umělých hnojiv. Použité chemikálie většinou nevyhubí jen škůdce, ale zároveň i další druhy živočichů, kteří mohou být přirozenými nepřáteli dalších škůdců a tím oslabují obranyschopnost celého společenství. Proto se upouští od využívání velkého počtu chemických prostředků při ošetřování polních plodin a zemědělci se zaměřují na integrované zemědělství, které je alternativou k ekologickému zemědělství a není snadné splnit veškeré jeho přísné limity a standardy. Integrované zemědělství (integrované systémy rostlinné produkce) je k přírodě šetrné právě díky používaným prostředkům, návodům a konceptům (ICM – integrované obhospodařování plodin, IPM – integrovaná ochrana před škůdci a plevely) zpracovaných několika národními a evropskými organizacemi. Zároveň je ale i citlivé na různé ovlivňující rizikové faktory jako vývoj počasí, výskyt chorob, rozšíření a přemnožení parazitů, a proto musí velmi rychle a správně reagovat na nastalé situace.

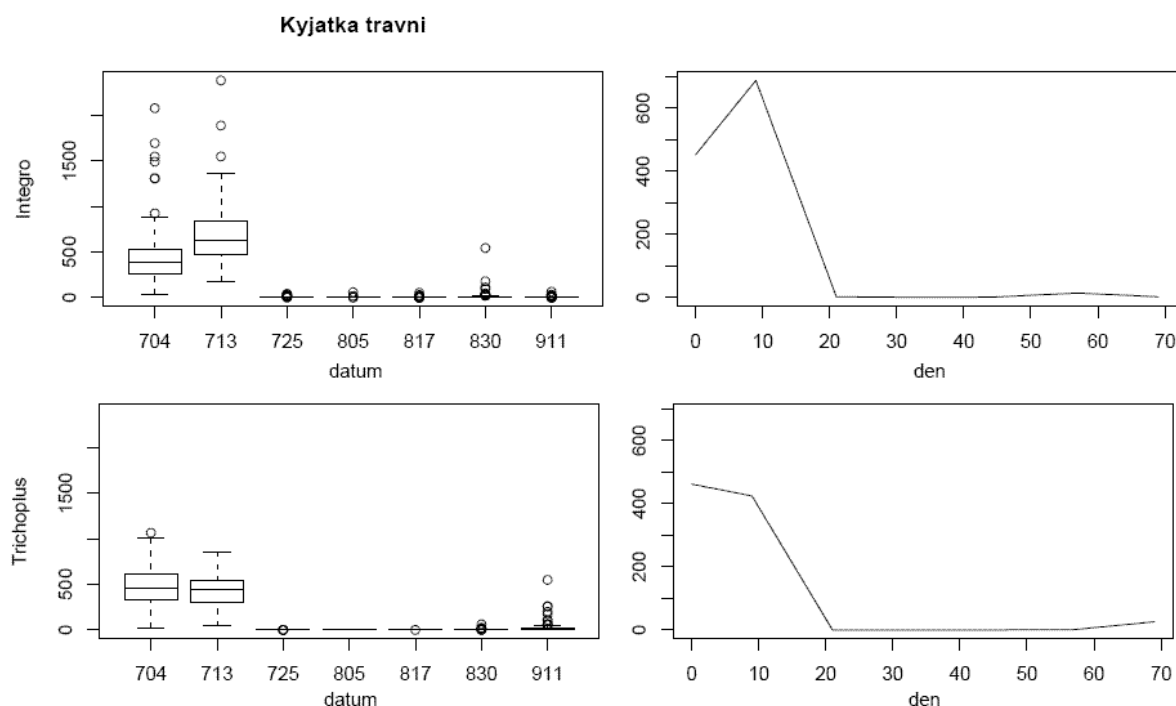
Kukuřice se celosvětově pěstuje zhruba na 150 miliónech hektarů, celková produkce za posledních 30 let vzrostla z 266 mil. tun na 640 mil tun. V České republice lze též pozorovat tento trend. Od počátku 90. let vzrostla výměra kukuřice z přibližně 30 000 ha na 85 000 ha v roce 2003. Rovněž hektarové výnosy u kukuřice na zrno vzrostly za stejné období z přibližně 4,5 t/ha na 5,5 t/ha v roce 2003 a dokonce na 6,4 t v roce 2004. Ochrana této plodiny tak nabývá vysoké důležitosti obzvláště kvůli problémům se škůdci, které se vyskytly například v Itálii, kde následně použili chemické prostředky a ty následně znehodnotili produkci mléka v celém regionu.

Z tohoto důvodu vyvíjíme model, který by dokázal zohlednit všechny rizikové faktory a zároveň dokázal predikovat vývoj populací škůdců na kukuřici a jejich regulaci za použití metod integrovaného či ekologického zemědělství.

2 Tvorba modelu

2.1 Sběr a analýza dat

Sběr dat probíhá již od roku 2004, kdy proběhlo monitorování populačních hustot mšic a jejich antagonistů v porostech kukuřice po biologickém ošetření proti zavíječi (*Trichogramma*) prostředkem *Trichoplus* a po aplikaci chemického insekticidu Marsal. V minulém roce pak proběhlo další měření opět po biologickém ošetření proti zavíječi a po použití chemického insekticidu Integro. Data byla naměřena ve 100 lokalitách pro každý použitý přípravek v sedmi termínech (4.7., 13.7., 25.7., 5.8., 17.8., 30.8. a 11.9.) a obsahují počet jednotlivých druhů mšic v těchto termínech. Tyto výsledky byly vyhodnoceny panem docentem Zdeňkem Pospíšilem a popsány v bakalářské práci Václava Psoty [1]. Příklad vyhodnocení nasbíraných dat je uveden na obrázku 1. Sběr dat podporuje firma Biocont Laboratory s.r.o., která pokračuje i v letošním roce a zajišťuje na pokusných lokalitách měření základních meteorologických dat (teploty, vlhkosti, ...).



Obr. 1: Příklad vyhodnocení nasbíraných dat

2.2 Nástroj pro management a analýzu dat

Data z minulých let byla dodána ve formátu Microsoft Excel a byla ručně zpracovávána, proto je aktuálně vyvíjen program pro správu těchto dat s plánovaným rozšířením o možnosti statistické analýzy.

Po zvážení požadavků na systém jsme se rozhodli pro architekturu J2EE včetně nové specifikace EJB 3.0. To nám umožní vyvíjet komponenty nezávisle na cílovém aplikačním serveru. Pro naše účely zatím využíváme aplikační server JBoss. Celý systém je komponován pomocí návrhového vzoru Model-View-Controller a řešen modulárně, což by mělo umožnit rozvoj tohoto software o další vlastnosti – například implementaci automatického sběru dalších informací například o teplotě, srážkách v průběhu mezi součtem velikosti populací, rozšíření o další technologické možnosti – web services, JMS messaging apod nebo integraci navrženého modelu, který by mohl aktuálně předpovídat možný vývoj a navrhnout řešení pro redukci populací škůdců.

Model

Model je funkčním a datovým základem celé aplikace. Pro jeho realizaci jsou využity „entity beans“ – entitní Java objekty jsou mapovány přímo na tabulky v relační databázi. Díky použití EJB 3.0

a obsažené specifikaci Java Persistence jsme získali databázovou nezávislost. Třídy, označené anotací `@Entity`, jsme implementovali jako serializovatelné, aby je bylo možné posílat i na klientskou aplikaci. Samotná implementace těchto entitních objektů je velmi snadná, jak ukazuje příklad 1, protože jsme nebyli svázáni s nějakým již existujícím návrhem.

```
package cz.muni.fi.corn;

import java.util.Date;

import javax.persistence.Entity;
import javax.persistence.Table;
import javax.persistence.Column;
import javax.persistence.Id;
import javax.persistence.Temporal;
import javax.persistence.TemporalType;

@Entity
public class Measurement implements java.io.Serializable
{
    private int id;
    private Date date;

    @Id @GeneratedValue
    public int getId() { return id; }
    public void setId(int pk) { id = pk; }

    @Temporal(TemporalType.DATE)
    public Date getDate() { return date; }
    public void setDate(Date d) { date = d; }
}
```

Příklad 1: Ukázka entity bean v EJB 3.0

Controller

Další částí jsou „session beans“ a „message-driven beans“, které zajišťují základní manipulaci s daty – načítání, ukládání atd. Využití EJB 3.0 v tomto směru ulehčuje práci vývojáři, který je například oproštěn od přístupu k databázové vrstvě (JDBC), složité správě přístupových práv nebo řízení transakcí. Dále pak obhospodařuje klientské programy a požadavky view vrstvy.

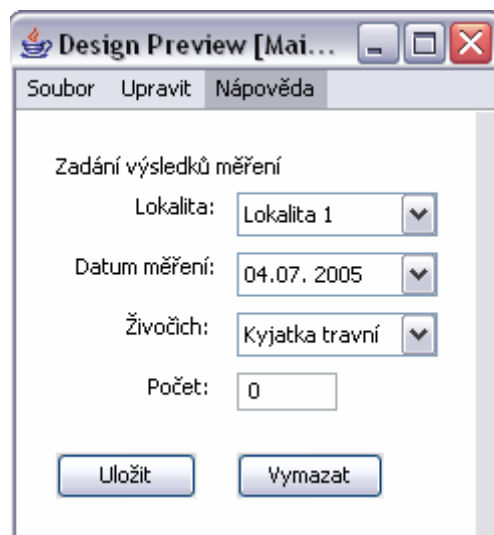
View

Tato část bude obsahovat klientskou aplikaci pro ukládání dat a zároveň webové rozhraní pro snazší přístup. Veškerá sesbíraná data budou dostupná na webových stránkách a bude možné vybírat různé podмноžiny těchto dat a s nimi pracovat při analýzách.

Součástí by se měl stát i program pro statistické vyhodnocování dat. V současné době uvažujeme o integraci systému počítačové algebry Maple skrz jeho rozhraní OpenMaple případně jiném ekvivalentním řešení.

Klientská aplikace

Klientská aplikace se připojuje pomocí JNDI Contextu k aplikačnímu serveru (JBoss) a může zadávat data, stahovat nasbíraná data nebo jejich podмноžiny. V budoucnu by mohla být rozšířena o spolupráci se systémem Maple, případně o další požadovanou funkcionalitu. Možnost zadávání dat v klientské aplikaci je uvedena na obrázku 2.



Obr. 2: Klientská aplikace

2.3 Matematický model

Matematický model obvykle popisuje systém, nebo jeho relevantní část, s pomocí množiny proměnných a množiny rovnic, které určují vztahy mezi nimi. Hodnoty proměnných mohou být např. reálná nebo celá čísla, booleovské hodnoty nebo textové řetězce. Proměnné reprezentují nějaké vlastnosti systému, např. rychlost růstu populace, vnitrodruhovou konkurenci, schopnost přežití při aplikaci některého přípravku apod.

V případě tohoto modelovaného problému se budeme snažit vyvinout dynamický, deskriptivní model, který bude reflektovat potenciální faktory ovlivňující růst populací mšic a afidofágů na kukuřici. Mezi tyto faktory patří hodnocení úhrnu srážek, velikosti teplot, ovlhčení listů, použití biologických a chemických prostředků pro ošetření rostlin.

Hlavním cílem modelovaného společenství je udržení jeho velikosti pod prahem škodlivosti, aby nedošlo ke katastrofickému poškození plodiny a jejímu znehodnocení. Dalším hledaným cílem může být snížení nákladů za využití ošetřovací prostředky.

V současné době ještě není rozhodnuto zda bude využit některý již existující model kalibrací parametrů, dojde k rozšíření některého známého modelu nebo bude vytvořen model zcela nový. Proto je důležité nejprve rozhodnout, které faktory budou v modelu zahrnuty a které zanedbáme:

- Růst populace: může populace růst neomezeně nebo její růst bude nějak limitován?
- Vyskytuje se mezi druhy nebo přímo uvnitř jednoho druhu nějaká konkurence?
- Lze kombinovat více chemických a biologických přípravků? Ovlivňují se navzájem v kladném nebo záporném smyslu?
- Jaký je vliv klimatu (teploty, srážky) na vývoj populace? Existuje například nějaká limitní teplota, po které je již prostředí pro škůdce nevyhovující? Respektive jaký teplotních rozsah je nejvhodnější pro jejich množení?

Na tyto otázky ještě neznáme přesnou odpověď a jejich ověření je součástí výzkumu z nasbíraných dat.

Dalším závažným problémem, který se obecně v matematickém modelování biologických systémů objevuje, je neurčitost biologických dat, která dosahuje odchylku až 20%. Proto musíme velmi obezřetně postupovat při valuaci parametrů a testování celého modelu. Na získané parametry bychom též měli aplikovat analýzu citlivosti, jejímž cílem je měnit parametry modelu a vyhodnotit tyto změny ve výsledcích modelu. Tato metoda je obzvláště užitečná k identifikování slabých míst modelu. Tato

místa pak mohou být zvýrazněna experimentováním nebo jednoduše popsána a vyznačena při aplikaci modelu, aby se předešlo špatné interpretaci výsledků.

Pokud je model obzvláště jednoduchý, je možné rozdělit výsledky řešení pro každý parametr v několika krocích. Odvozením lze získat přesné poměry změn prognózy vázané k jednotlivým parametrům. Ve složitějších modelech je nutné se dělení úplně vyhnout a použít některé numerické metody.

Velikost změny každého parametru závisí na tom, jak přesnou hodnotu parametru je potřeba získat. Z toho důvodu parametry, které jsou odvozeny ze vstupních dat, mohou obsahovat mnoho standardních chyb. Některé parametry mohou být odhadnuty, ale v tomto případě je nutné mít procentuální odhad spolehlivosti tohoto odhadnutého parametru. Obzvláště by měla být kladena zvýšená pozornost na parametry ve vzájemném vztahu, protože změna jednoho z nich může vést k případné změně ostatních závislých parametrů.

3 Závěr

Matematické modelování proniklo do různých oborů přírodních, technických, ekonomických i sociálních věd a stalo se důležitým pomocníkem při modelování a simulacích systémů, analýzách a předvídání různých procesů, jevů, chování druhů a stavů společenstev. Matematické modely poskytují srozumitelný popis všech relevantních faktorů dané situace a umožňují odhalit podstatné vztahy mezi prvky studovaného systému, proto není překvapením, že jejich využití se stále rozšiřuje i na běžné aplikace v reálném světě – tedy například v zemědělství.

Tento příspěvek se zabývá jedním potenciálním využitím matematického modelování při ochraně kukuřice před populacemi škůdců. Ty se mohou na základě různých faktorů šířit, ale jejich velikost může být také regulována různými metodami a prostředky. V příspěvku jsou tyto faktory popsány a na jejich základě se uvažuje o využití potenciálního modelu. Cílem modelu je tedy ochránit porost kukuřice před dosažením prahu škodlivosti škůdci.

V další části je popsán aktuální stav nasbíraných dat, způsob jejich sběru a vyhodnocování. Jelikož je tento systém (uchování dat a jejich vyhodnocení) nevyhovující je navržen software pro správu dat, vyhodnocení různých podmnožin naměřených dat a uveden i možný další rozvoj tohoto systému při dalších požadavcích na něj kladených.

4 Literatura

- [1] Psota, V.: Výskyt mšic (Aphidoidea) a afidofágů v porostech kukuřice při různých typech ochrany porostů. Bakalářská práce. 2006.
- [2] Pospíšil, Z.: Víno a roztoči (o jednom ekologickém modelu). In Biometrické metody a modely v současné vědě a výzkumu. ÚKZÚZ, Brno, 1998.
- [3] Burke, B., Monson-Haefel, R.: Enterprise JavaBeans 3.0. O'Reilly, 2006.
- [4] Kalas, J., Pospíšil, Z.: Spojité modely v biologii. Masarykova univerzita, Brno, 2001.
- [5] Pospíšil, Z.: Odhad parametrů logistické rovnice. In Biometrické metody a modely v současné vědě a výzkumu. ÚKZÚZ, Brno, 2002.
- [6] Murray, J.D.: Mathematical Biology: I. Springer, 2004.
- [7] Larsen, R. J., Marx, M. L.: An Introduction to Mathematical Statistics and Its Applications. Prentice Hall, 2001.
- [8] Barnes, B., Fulford, G.: Mathematical Modelling with Case Studies: A Differential Equation Approach Using Maple. T&F STM, 2002.
- [9] Jones, D.S., Sleeman, B.D.: Differential Equations and Mathematical Biology. Chapman & Hall/CRC, 2003.

Komunikační rámec pro výměnu environmentálních informací

Matěj Štefaník, Karel Kiswa

Masarykova univerzita, Institut biostatistiky a analýz,
Divize environmentální informatiky a modelování
Jana Uhra 10, 602 00 Brno
stefanik@iba.muni.cz, kiswa@iba.muni.cz

Abstrakt

V příspěvku je prezentován projekt Masarkovy University (Analýza a návrh komunikačního rámce s mezinárodními informačními systémy v životním prostředí). Projekt je financován z prostředků Ministerstva životního prostředí a tématicky se vztahuje k evropskému projektu DEMO-net. Náplní úvodní fáze výzkumu je sledovat environmentální standardy v komunikaci informačních systémů a prostředky pro zvýšení stupně systémové interoperability v národním ale především mezinárodním měřítku. Výzkum se zaměřuje na čtyři cílové skupiny (veřejná administrativa, akademická sféra, občané, podniky). Obsahem příspěvku je popis metodologie a aktuální výsledky vztahující se k interoperabilitě a komunikačnímu rámci pro národní a mezinárodní výměnu informací. Dalším tématem je znovupoužití informací veřejného sektoru s ohledem na požadavky eParticipation a záměrem EU podpořit interoperabilitu evropských služeb eGovernmentu.

Abstract

The Masaryk University Research Project (Analysis and Design of Communication Framework within International Environmental Information Systems) is presented. It is funded by the Ministry of Environment of the Czech Republic and related to the European DEMO-net project (The Democracy Network) funded by the Commission. Research activities of Masaryk University monitor environmental communication activities and current interoperability with four target groups (public administration bodies, academicians, citizens and businesses) of the Czech Republic within national and pan-European environmental information systems to support eDemocracy and eParticipation. Methodology and prospective results relating to an interoperability and communication framework for national and international environmental information exchange are being discussed. Another theme is the reuse of public sector information with respect to eParticipation and objectives of interoperability for pan-European eGovernment services.

Klíčová slova

eParticipation, DEMO-net, komunikační rámec, environmentální informační systémy, interoperabilita

Keywords

eParticipation, DEMO-net, communication framework, environmental information systems, interoperability

1 Legislativní základy

V minulosti byl nárok na informace o životním prostředí rovněž definován, ale jeho definice byla poměrně nepřesná. To umožňovalo úřadům interpretovat tyto pravidla poměrně libovolným způsobem, což zapříčinilo, že ani služby poskytované úřady nedosahovaly dostatečně kvalitní a požadované úrovně. Dalším významným problémem byly značné rozdíly v pravidlech pro přístup k informacím v rámci zemí Evropské unie. Rozdíly v právních normách a předpisech byly tak velké, že bylo potřeba vytvořit iniciativu pro jejich sjednocení. Pro vynucení určitého standardu v kontextu celé EU a pro sjednocení právních norem pro přístup k informacím o životním prostředí byly přijaty tři základní dokumenty.

Nejstarším z těchto dokumentů byla *Convention on Access to Information, Public Participation in Decision-making and Access to Justice in Environmental Matters*, která byla uzavřena v roce 1998 v dánském Aarhusu. Smlouva rozšiřovala původní dokument o přístupu k environmentálním

informacím (směrnice 90/313/EEC) o zásadní aspekt, který umožňoval veřejnosti participovat se na rozhodování v otázkách životního prostředí (účastnit se a zasahovat do řízení týkající se environmentální problematiky).

Dalším významným krokem, který zlepšil dostupnost informací v oblasti životního prostředí, bylo schválení směrnice 2003/4/EC *on public access to environmental information*, která nastavila jasná a transparentní pravidla pro zveřejňování a zpřístupňování informací o životním prostředí v držení veřejných a soukromých subjektů.

Snadnější přístup veřejnosti k informacím o životním prostředí a šíření těchto informací přispívají k větší osvětě občanů v otázkách životního prostředí, k volné výměně názorů, k účinnější účasti veřejnosti při rozhodování o otázkách životního prostředí a k lepšímu životnímu prostředí. Rozdíly v zákonech mezi jednotlivými členskými státy, které se týkají přístupu k informacím o životním prostředí by mohly v rámci Evropské unie vést k rozdílnému přístupu k těmto informacím nebo k rozdílným podmínkám hospodářské soutěže.

Orgány veřejné správy mají nyní jasně definované postupy, jakým způsobem zpřístupnit informace o životním prostředí široké veřejnosti a jakým způsobem podpořit šíření těchto informací a to zejména prostřednictvím informačních a komunikačních technologií. (klíčovým bodem je rovněž nutnost periodické revize, které usměrní tyto postupy v návaznosti na aktuální technické možnosti).

Posledním významným dokumentem je směrnice 2003/98/EC *on the re-use of public sector information*, která definovala pravidla, jakým způsobem nakládat s dokumenty v držení veřejné správy. Veřejný sektor shromažďuje, vytváří, reprodukuje a šíří široké spektrum informací z mnoha oblastí, např. sociální, hospodářské, zeměpisné, meteorologické, turistické informace, informace o podnikání, patentech a vzdělávání. Subjekty veřejného sektoru provádí tyto funkce při plnění svých veřejných úkolů. Využívání těchto dokumentů k jiným účelům znamená opakované použití. V pravidlech a postupech členských států týkajících se využívání zdrojů informací veřejného sektoru existují značné rozdíly, které vytvářejí překážky pro plné využití hospodářského potenciálu tohoto základního zdroje dokumentů (situace byla podobná jako u poskytování informací o životním prostředí). Proto bylo třeba provést minimální harmonizaci pravidel a postupů pro opakované použití dokumentů veřejného sektoru, aby se potenciál těchto dokumentů mohl plně využít.

Existuje celá řada dalších dokumentů, ale ty se zabývají pouze dílčími nebo konkrétními aspekty environmentálních informací. Tři výše zmíněné dokumenty tvoří společně ucelený základ pro poskytování libovolných informací z oblasti životního prostředí.

Dokumenty ale nenařizují použití nějaké konkrétní technologie. Obsahuje pouze popis funkcí, které by měla státní správa zahrnout do své agendy a které by měla respektovat při jednání s občany.

Hlavním cílem těchto právních předpisů je na jedné straně zjednodušení a sjednocení přístupu k dokumentům a na straně druhé podpoření participace občanů v kontextu životního prostředí a to ve všech třech klíčových vztazích:

- Informační: pouze jednosměrná komunikace, kdy veřejné instituce vytváří a dodávají informace občanům (podpora a zjednodušení přístupu k dokumentům)
- Konzultační: dvousměrná komunikace, kdy občané poskytují státní správě zpětnou vazbu. Státní správa nastavuje pravidla a řídí proces.
- Aktivní participace: partnerství se státní správou, kdy jsou občané aktivně angažováni a podílí se na definici procesu a obsahu rozhodování.

Cílem je tedy zapojit co největší počet občanů do procesu rozhodování. Nejvhodnějším nástrojem k těmto účelům jsou bez pochyby ICT.

2 Klíčové otázky

Klíčová otázka tedy zní, jaké technologie a postupy použít, aby se zjednodušila a zefektivnila komunikace mezi úřady a občanem. Jakým způsobem navrhnout služby, aby měl každý občan možnost participovat na rozhodovacím procesu a aby měl veškeré dostupné informace (potřebné

k svému rozhodování) na dosah ruky. Při hledání optimálního řešení je nutné uvažovat nejenom nad otázkami technického charakteru, ale také nad otázkami sociálními.

- Pro dosažení tohoto cíle je potřeba v prvním kroku analyzovat možnosti, které nám poskytují současné technologie. Množství technologií používaných v současnosti je opravdu velké a často se ve velmi podobných situacích nebo aplikacích používají zcela rozdílné standardy a postupy. Analýzou v současnosti nepoužívanějších technologií je možno nalézt nejenom nejrozšířenější technologie, ale také kritická místa, na kterých jsou současné technologie již na pokraji svých možností. Tyto místa je potřeba sledovat se zvýšenou pozorností, aby se předešlo možnému selhání systému v budoucnosti.
- Další krok je sledování nově objevených metod a technologií. V současné době probíhá v rámci ICT bouřlivý rozvoj v řadě technologických odvětví. Dopad nových technologií na občany EU může mít velmi výrazný. Podpora ze strany sémantického webu nebo systému rozpoznávání řeči zcela jistě změní přístup, jakým lidé komunikují a tím pádem i způsob jakým komunikují se státní správou. Ta bude muset změnit způsob komunikace tak, aby vyhověla novým požadavkům ze strany občanů.
- Nové možnosti, které přinese technologický rozvoj, přinesou i nové příležitosti, jakými bude občan moci participovat na rozhodovacím procesu. Proto je nezbytné sledovat a hodnotit technologie také z pohledu jejich přínosu pro zvýšení participace občanů v otázkách veřejných záležitostí a ne pouze z pohledu jejich technologické přínosnosti. Cílem je tedy definovat taková technická řešení, která maximalizují spolupráci občanů s institucemi ve veřejném sektoru.
- Nezanedbatelnou částí je vývoj nástrojů a metod pro hodnocení technologií vzhledem k jejich přínosnosti jak v oblasti technické, tak také v oblasti participace. V obou případech je nezbytné srovnávat u jednotlivých technologií celou řadu atributů, které jsou velmi často zcela odlišného charakteru. Tato nesourodost činí vzájemné porovnání technologických alternativ velmi obtížné.

3 Hlavní cíle

Projekt Komunikační rámec je rozdělen do několika fází:

1. Analýza cílových skupin

- Cílem 1. etapy je analytická fáze projektu zahrnující specifikaci dlouhodobého rámce příslušných spoluprací. V této fázi výzkumu jsou identifikovány informační potřeby jednotlivých cílových skupin (veřejná správa, podniky, vědeckovýzkumné a vzdělávací organizace, veřejnost a nevládní organizace) vzhledem ke komunikaci s relevantními informačními systémy v zahraničí.

2. Analýza interoperabilitního rámce

- Cílem 2. etapy výzkumu je analyzovat komunikaci s relevantními informačními systémy identifikovanými v první etapě. Analýza zahrnuje posouzení komunikace se systémy JSIŽP a mezinárodními systémy v environmentální oblasti, dále použité informační a komunikační technologie, legislativní mandát, organizační rámec a technologické a komunikační standardy, na jejichž základě systémy fungují a komunikují.

3. Návrh komunikačního rámce pro Českou republiku

- Cílem 3. etapy výzkumu je návrh efektivních metod komunikace českých informačních systémů s mezinárodními komunikačními a informačními systémy v environmentální oblasti identifikovaných a analyzovaných v předchozích etapách a specifikace dlouhodobého rámce příslušných spoluprací.

4 Metodologie

Z velké části se náš výzkum týká zpracování velkého množství různých dokumentů. Tyto dokumenty se mohou dotýkat například legislativní sféry (směrnice, nařízení, doporučení EU) nebo standardizačních dokumentů (standards, normy, doporučení pro dodržování standardů). U legislativních dokumentů je třeba myslet i na zákony, které jsou teprve ve fázi přípravy a v platnost vstoupí například až za rok či dva. Mezi dokumenty mohou být zařazeny například i HTML stránky jednotlivých organizací, ve kterých jsou popsány jejich interní postupy a standardy (především EEA, CEN, atd.). Dalšími dokumenty jsou studie nasazení a použití technologií a analýza jejich přínosu pro danou oblast. Dále hledáme dokumenty, obsahující analýzy technologických novinek, které by měly nahradit aktuální způsoby výměny informací a komunikace v horizontu několika dalších let (technologie, které se stanou každodenní realitou). V neposlední řadě je potřeba vyhledat co největší počet dokumentů, které se částečně týkají technické stránky, ale z velké části reagují na sociální problémy (doporučení pro vývoj aplikací pro lidi s postižením nebo pro seniory). Uvedené typy dokumentů tématicky pojmu většinu dat potřebných pro náš výzkum.

Pro kvalitní analýzu je potřeba vyhledat co největší množství relevantních dokumentů. Je nezbytné vypracovat metodologie, jakým způsobem tyto dokumenty hodnotit a jak dokumenty kategorizovat.

Při analýze nalezených standardů, je klíčová právě jejich kategorizace. Vzhledem k tomu, že některé dokumenty mají mnohem závažnější dopad než jiné, je dalším důležitým prvkem jejich hodnocení priorit. (přidělení určité hodnoty (čísla) danému dokumentu, která představuje jeho závažnost vzhledem ke všem ostatním dokumentům). Vysokou váhu tedy lze přidělit například dokumentům z oblasti legislativy a nízkou některým doplňkovým dokumentům.

Prvním úkolem ale je, zařadit každý dokument do správné kategorie v závislosti na tom, které tématické oblasti se dotýká. Tématických oblastí lze definovat několik:

- Používané technologie – technologie, které jsou používány v některé instituci nebo libovolnou státní správou
- Technologie ve vývoji – technologie, které se teprve vyvíjí a jejich praktické využití zatím neexistuje (nebo jenom v omezené míře)
- Podpůrné standardy – standardy ne přímo technického charakteru, ale standardy, které určitým způsobem nastavují pravidla či mantinely pro již existující postupy, aby je učinili více užitečnými nebo více dostupnými.

Takto shromážděné a kategorizované dokumenty je potřeba dále zpracovat. Každý dokument se může dotýkat velmi široké palety tématických oblastí (nejedná se o rozdělení podle způsobu použití dané technologie jako v předchozím případě). Těchto tématických oblastí je celá řada. Od zcela obecných jako je např. datum vydání, doba platnosti, autor a název dokumentu až po zcela konkrétní atributy např. relevantnost dokumentu ke specifické oblasti, typ standardu ke kterému se vztahuje nebo předpokládaná doba nasazení (u teprve vyvíjených technologií).

Cílem projektu není pouze shromáždění standardů a norem, které jsou potřebné pro uskutečnění výměny informací po stránce technické, ale také analýza stávající situace a nalezení překážek a problémů, které v současnosti nejvíce brání občanům v možnosti participovat na rozhodování v otázkách životního prostředí.

Z tohoto důvodu je součástí projektu dotazník, který měl nalézt odpověď na otázku: „*V čem vidí občané největší překážky v komunikaci a co by se podle jejich názoru mělo změnit, aby se situace zlepšila?*“.

Hlavním požadavkem při návrhu dotazníku bylo zahrnout do výzkumu co nejširší skupinu občanů. Sběr odpovědí od respondentů probíhal elektronicky (speciální webová stránka). Důvodem bylo jednoduché zpracování dotazníků (nebylo potřeba další konverze do elektronické formy), ale také množství obdeslaných respondentů. Použití elektronického dotazníku omezilo rozsah respondentů pouze na lidi, kteří mají přístup k internetu. Pozitivem tohoto řešení byl podstatný nárůst objemu respondentů. Vynechání skupiny obyvatel, která přístup k internetu nemá, by se mohlo jevit jako

zásadní problém. Nicméně je potřeba si uvědomit, že otázky pokládané v dotazníku byly otázky zaměřené spíše na oblast dostupnosti environmentálních informací pomocí moderních ICT. Respondenti, kteří nepracují s internetem by ve většině otázek nemohli vhodně odpovědět.

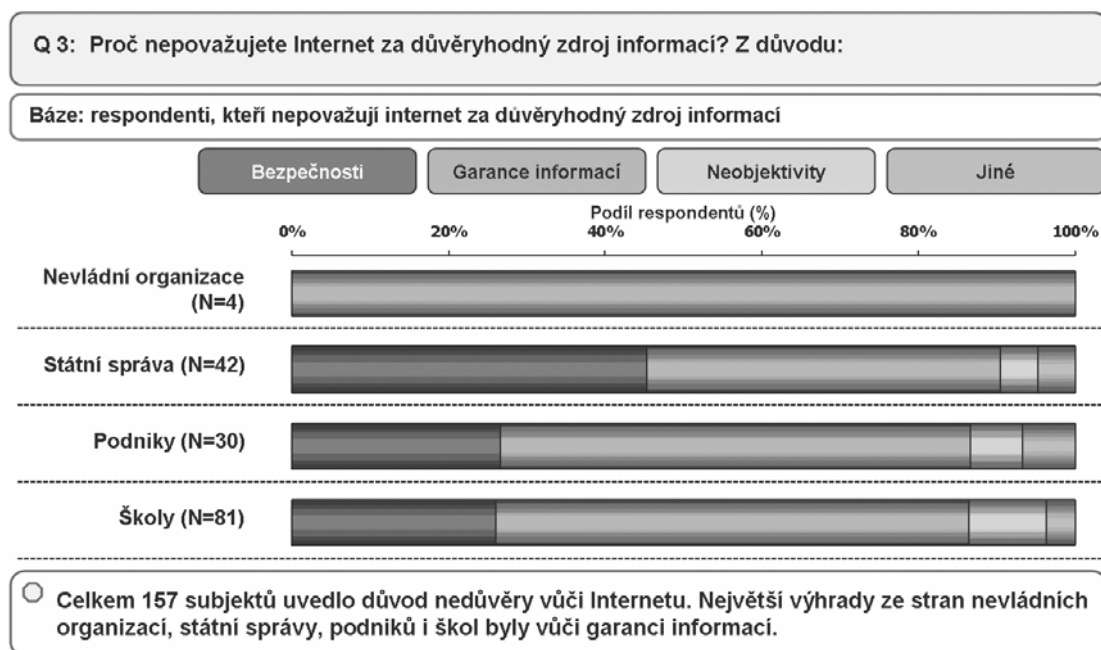
Při vývoji dotazníku jsme nejprve navrhli jeho kostru, ve které jsme nastínili pravděpodobné směřování otázek. Poté následovalo několik konzultací s vybranými zástupci jednotlivých skupin v krajích (Vysočina, Jihomoravský, Moravskoslezský), při kterých se dotazník dopracovával do výsledného stavu. Následovalo rozeslání dotazníku na e-mailové adresy. Tyto adresy byly shromažďovány paralelně s tvorbou dotazníku. Všechny e-mailové adresy pochází z veřejně dostupných zdrojů informací (např. portály veřejné správy, nebo weby nevládních organizací). Podobně se získávali kontaktní informace podniků a škol. Některé výsledky, které přinesla analýza dotazníku budou předvedeny níže.

5 Předběžné výsledky

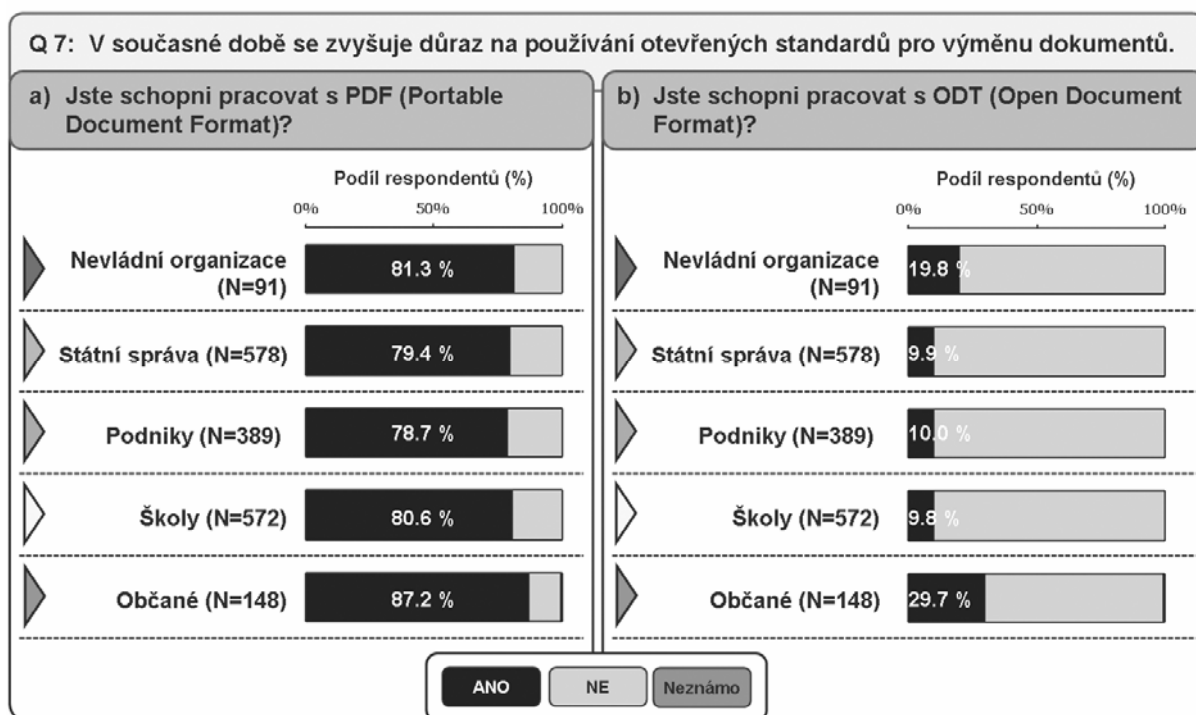
Vzhledem k možnostem, které pro komunikaci poskytuje internet a obecně ICT jsme mohli zahrnout do průzkumu poměrně širokou vrstvu respondentů. Osloveno bylo přes 6000 občanů a organizací. Počet vyplněných dotazníků byl nakonec nižší (asi 30%). Cílové skupiny, které byly osloveny respektují původní rozdělení (státní správa, podniky, nevládní organizace, školy). K těmto skupinám byl přidán ještě občan jako samostatná kategorie. Největší podíl z odpovědí činily školy a instituce státní správy.

Jaké jsou předběžné výsledky z našeho dotazníkového průzkumu? Z průzkumu je patrné, že občané stále cítí značnou nedůvěru k informacím nalezeným na internetu. Největší problémy vidí občané ve věrohodnosti prezentovaných informací, což je v současné době poměrně palčivý problém. Hned na druhém místě byla bezpečnost a zabezpečení soukromí.

Dalším zajímavým výsledkem byla připravenost pro používání Open Document Format. Jak se ukázalo, největší schopnost adaptovat se na nově používané technologie mají především občané společně s nevládními organizacemi. Naopak poměrně špatně dopadly státní instituce společně s podniky a školami. Zvláště velké mínus je to v kontextu státních institucí vzhledem k tomu, že právě instituce by se měli snažit být maximálně otevřené vůči svým občanům a používat co nejotevřenější možnou formu komunikace. Podstatně lépe dopadla schopnost využívat v současnosti hojně používaný Portable Document Format, který bez problémů zvládne používat většina respondentů.



Obr. 1: Nedůvěra vůči Internetu



Obr. 2: Ovládání PDF a ODF

Nicméně rozdíl mezi schopnostmi občanů a všech ostatních skupin je stále značně velký. Tento rozdíl ve schopnosti používat PDF a ODF je s největší pravděpodobností zapříčiněn faktem, že pro občana jako jednotlivce je mnohem snazší přejít na novou technologii než pro libovolnou jinou instituci. Navíc instituce reflektují požadavky občanů a dokud se nová technologie nestane výrazněji využívaná, bylo by její nasazení v rámci veřejné instituce zbytečné.

Další otázka zdůraznila důležitost zpracování presentovaných informací v několika jazykových mutacích. Většina respondentů preferuje při vyhledávání informací národní jazyk (v tomto případě češtinu). Nevládní organizace, podniky a školy mají mnohem výraznější schopnost vyhledávat i materiály napsané v cizím jazyce. V tomto směru státní správa opět zaostává (tady poměrně výrazně). Možným vysvětlením je fakt, že státní instituce využívají jako zdroj informací především další státní instituce (a tím pádem nemusí některé informace vyhledávat v zahraničí) nebo pracují pouze na základě českého právního systému a tak zahraniční informace vůbec nepoužívají.

6 IDABC & Interoperabilita, IF, EIF

Úvod

Jak již bylo řečeno, cílem projektu VaV SM/10/99/05 je nalezení efektivních metod komunikace českých informačních systémů v rámci JISŽP s mezinárodními komunikačními systémy v oblasti životního prostředí.

Problematikou komunikace informačních systémů se mimo jiné organizace zabývá IDABC. Je to program řízený Generálním ředitelstvím pro podnikání a průmysl Evropské komise. Podporuje poskytování veřejných služeb občanům i podnikům v rámci Evropské Unie s využitím nejnovějších informačních a komunikačních technologií. Pro dosažení těchto cílů vydává různá doporučení, technické specifikace, vyvíjí řešení, ..., která umožňují jednotlivým členským státům (státním správám, občanům a podnikům) komunikovat elektronicky efektivněji jednotným způsobem.

Nyní je program IDABC upraven rozhodnutím Evropského parlamentu a Rady č. 2004/387/ES ze dne 21. dubna 2004 o interoperabilním dodávání pan-evropských služeb eGovernmentu správním orgánům, podnikům a občanům (IDABC).

Tímto rozhodnutím se na období 2005–2009 zřizuje Program pro interoperabilní poskytování celoevropských služeb elektronické správy evropským orgánům veřejné správy, orgánům Společenství a jiným subjektům a evropským podnikům a občanům ("program IDABC").

Účelem programu IDABC je identifikovat, podpořit a prosazovat rozvoj a zřizování celoevropských služeb elektronické správy a interoperabilních telematických sítí, které je podporují, podporujících členské státy a Společenství v rámci jejich příslušných oblastí působnosti při provádění politik a činností Společenství a dosahování podstatného přínosu pro orgány veřejné správy, podniky a občany. Pro své účely definuje pojmy: "telematická síť", "celoevropské služby elektronické správy", "projektem společného zájmu", "infrastrukturními službami", "horizontálním opatřením", "interoperabilitou". Doporučení obsahuje přílohy:

- Příloha I. – oblasti politiky pro projekty společného zájmu
- Příloha II. – horizontální opatření

Analýza

Po analýze portálu IDABC (<http://europa.eu.int/idabc/>) je zřejmé, že některé zde definované obecné principy komunikace by mohly sloužit (po přizpůsobení pro environmentální oblast) jako základní principy pro projekt Návrh komunikačního rámce s mezinárodními informačními systémy v ŽP. Jsou to zejména principy tvorby a fungování Interoperability Framework a European Interoperability Framework.

- Interoperability Framework (IF) – množina standardů a směrnic, která popisuje způsob, který organizace musí dodržet (nebo by měly dodržet) při vzájemné interakci. Z toho vyplývá, že se IF může v čase měnit (s příchodem nových technologií vzniknou nová doporučení).
- European Interoperability Framework (EIF) – definuje množinu doporučení a směrnic pro eGovernment služby takových, aby veřejná správa, podniky a občané mohly komunikovat na mezinárodní úrovni.

Pro zavedení těchto frameworků je nutné ještě definovat pojem "interoperabilita":

- Interoperabilita – schopnost informačních a komunikačních systémů a řídicích procesů podporovat výměnu dat a umožnit sdílení informací a znalostí.

Interoperabilita

V nejzákladnější formě interoperability se všechny eGovernment služby dají rozdělit do následujících tří typů:

- přímá interakce mezi občanem, nebo podnikem z jednoho členského státu se správou jiného členského státu, a (nebo) Evropské instituce
- výměna dat mezi jednotlivými členskými státy za účelem řešení případů, kdy občané, nebo podniky mohou vznést požadavek (ohledně jiné členské země) přímo na správu své země.
- výměna dat mezi různými EU institucemi, nebo agenturami a EU institucí, nebo agenturou a jednou, nebo více správami členských států.

Formáty interoperability

Existují tři hlavní stránky interoperability a spolu s každou je uveden seznam obecných zásad a doporučení, který jsou zde zaměřeny na eGovernment služby, ale mohou sloužit jako základní principy pro jakoukoliv obecnou (environmentální) komunikaci (výměnu dat) :

organizační interoperabilita – stránka interoperability, která je zodpovědná za definování prioritních cílů, modelování řídicích procesů, informování o spolupráci správ, které si přejí vyměňovat informace a mohou mít rozdílné vnitřní struktury a procesy.

- Požadavky na mezinárodní služby by měly být vymezeny společně, všemi účastníky. To by mělo vést k identifikaci a určení prioritních služeb, které mají být poskytovány.

- Pro realizaci mezinárodní služby je nutné nejdříve analyzovat příslušné řídicí procesy a aktéry, kteří jí budou ovlivněni a také by se mělo dohodnout nezbytné rozhraní, které budou její řídicí procesy využívat pro komunikaci a měla by definovat základní standardy.

sémantická interoperabilita - stránka interoperability, která zaručuje, že existuje přesný způsob výměny informací a je srozumitelný pro jakoukoliv aplikaci. Základním požadavkem pro výměnu informací je jednotný jazyk, který umožní popis významu a struktury výchozích dat – například nějaký značkovací jazyk (XML).

- Pro každou službu na mezinárodní úrovni by datové elementy pro výměnu měly být interoperabilní tím, že budou vyžadovat alespoň odpovědnost jednotlivých účastníků (správ členských států) za publikování informací na národní úrovni
- Iniciativa na mezinárodní úrovni pro vývoj základní sémantiky na bázi XML by měla být prováděna koordinovaně a měla by spolupracovat s již existujícími standardy.

technická interoperabilita - stránka interoperability, která pokrývá technické problémy spojené s propojením počítačových systémů a služeb. Nejpoužívanější jsou tato řešení:

- Přímé propojení aplikace s příslušnou middleware komponentou. Například statistická aplikace v podniku, která automaticky posílá požadovaná data Národnímu statistickému úřadu
- Založení portálu ovládaného řídicí aplikací. Portál ovládá komunikaci s uživateli (identifikace uživatele, autentizace uživatele, poskytnutí přehledu příslušných služeb, komunikace s řídicí aplikací, ...)

Interoperability Framework (IF), obecný přístup

Při implementaci národního interoperabilitního rámce lze požadované interoperability dosáhnout dvěma způsoby: standardizací technologií a harmonizací legislativy

Další doporučení jsou:

- používání otevřených standardů
- začlenění existujících standardů
- přepracování administrativních procesů a co nejlepší využití dostupné technologie. To je také možnost jak služby více přizpůsobit uživatelům.
- udržovat jednotlivé systémy nezávislé na patentovaných (vlastnických) technologiích
- centrálně dohodnutá XML shemata by měla být poskytována bezplatně. Tato forma znovupoužití snižuje náklady a potřebu odděleně vyvíjet mechanismy pro vzájemnou výměnu dat
- sledovat vývoj v širších souvislostech. Například změna legislativy může vynutit požadavek na poskytnutí nějaké služby
- redukce množství sbíraných dat pomocí dobře nadefinovaných datových slovníků a datových struktur
- zajištění bezpečnosti informací prevencí neautorizovaných přístupů do systému, a v případě vysoce citlivých informací zajištění bezpečnost každého záznamu (nebo dokonce každé jednotlivé části)
- umožnit liberální přístup (uživatelsky příjemný interface, přístup pro postižené, podpora cizích jazyků, ...).

European Interoperability Framework (EIF), obecný přístup

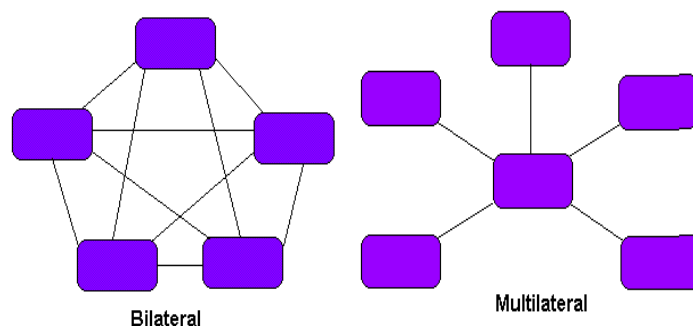
Správy členských států a EU instituce a Agentury by měly používat doporučení poskytované Evropským rámcem pro interoperabilitu (EIF) k zavedení mezinárodního rozměru do svých vlastních

interoperabilitních rámců a správné infrastruktury, aby mohly poskytovat mezinárodní služby, které budou schopny spolupráce. Dodržování EIF by mělo být také zmíněno v příslušném národním interoperabilním rámci.

Implementace následujících hlavních principů by měla být uvažována pro jakoukoliv mezinárodní službu:

- Přístupnost (accessibility)
- Mnohojazyčnost (multilingualism)
- Bezpečnost (security)
- Soukromí (privacy)
- Subsidiarita (subsidiarity)
- Používání otevřených standardů
- Posouzení prospěchu z open-source software
- Použití “multilateral” řešení

“multilateral” řešení je systém, kde každý člen komunikuje s ostatními přes jedno centrum. V tomto centru podporuje dohodnuté způsoby komunikace a dohodnuté standardy a specifikace. Oproti “bilateral” řešení, kdy každý člen systému je propojen s ostatními členy přímo a mezi různými dvojicemi členů může být jiný způsob komunikace založený na jiném standardu (popř. bez standardu).



Obr. 3: „Bilateral“ a „Multilateral“ řešení

Závěr

Kostra každé služby měla být postavena na obecném principu tak, aby bylo možné ji parametrizovat přímo pro danou problematiku. Vzhledem k rychlému vývoji informačních komunikačních technologií by tento základ měl být co nejméně závislý na použitých technologiích. Veškerá řešení (technická, sémantická, ...) a modely (infrastruktura, ...) by měly být dány spíše jako logická struktura, tak aby je každý účastník mohl realizovat podle svých možností (legislativa jednotlivých členských států, finanční stránka, ...).

IDABC nabízí některá obecná řešení (modely) a také ty, která jsou již zaměřena na určitou problematiku, např. eGovernment. V rámci projektu “Návrh komunikačního rámce s mezinárodními informačními systémy v ŽP” může sloužit jako zdroj informací pro analýzu a následnou tvorbu základních principů a metodologie pro environmentální oblast.

7 Shrnutí

Vzrůstající nasazení nových ICT přináší nové možnosti veřejné správě a širší veřejnosti. Rozvoj technologií je v současnosti velmi prudký a mezi nejčastěji zmiňované technologie bezesporu patří například ODF (Open Document Format) či OSS (Open Source Software), které využívají otevřené standardy a mají potenciál podpořit vyšší interoperabilitu mezi systémy a také zjednodušit komunikaci směrem k občanům. Nové technologie mohou nabídnout originální řešení problémů, které nemohou být

řešeny klasickým softwarem a mají potenciál pro významné úspory ve výdajích veřejné správy na ICT.

Pokud se zaměříme pouze na Open Source a ODF zjistíme, že ačkoliv je tento princip znám již několik let stále ještě není jeho vývoj u konce a je nutné vyvinout identifikaci ODF a následně způsob implementace v rámci veřejné správy např. v oblasti kancelářských aplikací a při řešení formátu výměny dokumentů. V těchto oblastech již vývojáři OSS produktů (kancelářských balíků) přistupují k implementaci standardu OASIS OpenDocument do svých produktů.

Cílem ale není pouze prověřovat možnosti OSS, ačkoliv je jeho přínos v současnosti velmi důležitý. V další práci se budeme zaměřovat i na zcela nové technologie, které by mohly přinést nové přístupy pro výměnu informací, pro podporu komunikace mezi občanem a státní správou a pro podporu aktivní účasti občanů v rozhodování. Mezi tyto budoucí a perspektivní technologie lze zcela jistě zařadit například:

- Rozpoznávání řeči
- Vyhledávání znalostí
- Ontologie
- Sémantický web

Právě tyto a další technologie v blízké budoucnosti velmi pravděpodobně změní principy, na kterých stojí současný model komunikace. Otázka stále zní, jak velký tento dopad bude a jaké změny ve způsobu komunikace bude vyžadovat.

8 Literatura

- [1] Hřebíček, J., Pitner, T., Benko, V. (2003): Standardization of Environmental Data and Information Management in the Czech and Slovak Republic. Proceedings of 17. International Conference Informatics for Environment Protection. EnviroInfo 2003. The Information Society and Enlargement of the European Union, September 2003, Cottbus, Germany, pp. 117-124.
- [2] Hřebíček, J. (et al), (2005): VaV SM 99/10/05 Analysis and Design of Communication Framework with International Environmental Information Systems – Annual report. (in Czech). Brno: Masaryk University, 2005.
- [3] Hřebíček J. (2005): Environmental Communication Framework in e-Government of the Czech Republic. Proceedings of 19th International Conference Informatics for Environment Protection, EnviroInfo 2005. Networking Environmental Information. September 2005, Brno, Czech Republic, pp. 841-847.
- [4] IDABC (2004): European Interoperability Framework for pan-European eGovernment Services. Version 1.0. Luxembourg: Office for Official Publications of the European Communities.
- [5] Directive 2003/4/EC of the European Parliament and of the Council of 28 January 2003 on public access to environmental information and repealing Council Directive 90/313/EEC, Official Journal L 041 , 14/02/2003 P. 0026 – 0032
- [6] Directive 2003/98/EC of the European Parliament and of the Council of 17 November 2003 on the re-use of public sector information, Official Journal L 345 , 31/12/2003 P. 0090 - 0096
- [7] Aarhus Convention; Convention on Access to Information, Public Participation in Decision-making and Access to Justice in Environmental Matters.
- [8] EURES - European-wide job search portal: (<http://europa.eu.int/eures>)
- [9] PLOTEUS, training opportunities database:(<http://europa.eu.int/ploteus>)
- [10] SOLVIT solving administrative obstacles in cross-border procedures:
- [11] (http://europa.eu.int/comm/internal_market/solvit)
- [12] TRIS 98/34 information site concerning national technical regulations: (http://europa.eu.int/comm/enterprise/tris/index_en.htm)

[13] Your Europe: (<http://europa.eu.int/public-services/>)

Příspěvek vznikl za podpory grantu Ministerstva životního prostředí České republiky č. SM/10/99/05 „Návrh komunikačního rámce s mezinárodními informačními systémy v ŽP“.

Rychlé grafické algoritmy a jejich využití se systémem Maple

Vladimír Žák

Vysoké učení technické v Brně, Fakulta strojního inženýrství
Technická 2, 616 69 Brno
zakyn@centrum.cz

Abstrakt

Příspěvek se zabývá využitím rychlých grafických algoritmů se systémem Maple. Jsou zde popsány základní pojmy matematické teorie, která je potřebná pro tvorbu efektivních grafických algoritmů. Tato teorie umožní vytvořit velmi rychlé algoritmy, které pak mohou pomocí technologie OpenMaple přímo přistupovat k jádru systému Maple. V příspěvku bude uveden nejjednodušší příklad tohoto využití.

Abstract

This paper deals with fast graphics algorithms to use them with Maple system. It describes goal terms of mathematical theory that are needed to create effective graphics algorithms. This theory allows us to create very fast algorithms that are able to access to the kernel of Maple system via OpenMaple technology. In this paper there is the simplest example of this approach.

Klíčová slova

Rychlé grafické algoritmy, matematická teorie, systém Maple, technologie OpenMaple.

Keywords

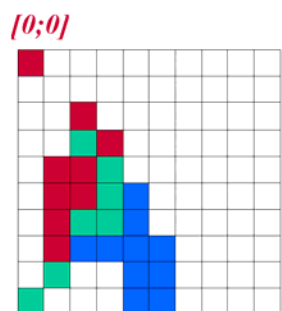
Fast graphics algorithms, mathematical theory, system Maple, OpenMaple technology.

1 Úvod

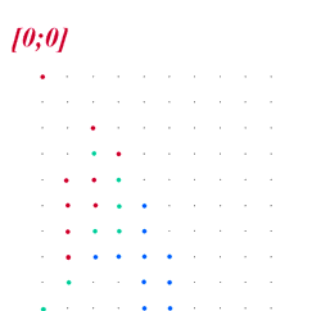
V počítačové grafice jsou data uložena buď jako vektorová data, který popisuje sekvenci grafických operací, jež jsou později vykresleny (např. metafile), nebo ve formě matice, která reprezentuje pixely obrázku (např. bitmapa).

Data bitmapy jsou uložena jako matice, která reprezentuje souřadnice bodů, jež mají být vykresleny. Tyto body jsou obvykle označovány jako pixely. V matematickém modelování jsou dva druhy pixelů: logický a fyzický. Fyzické pixely jsou chápány jako množina malých plošek (viz. Obr. 1). Logické pixely jsou reprezentovány množinou izolovaných euklidovských bodů (viz. Obr. 2).

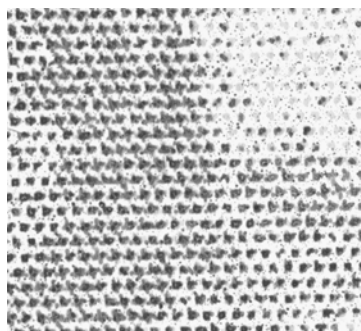
Výstup je poté zobrazován jako šachovnice, která je množinou čtverců různých barev (chápány jako otevřené disjunktní množiny). V této souvislosti se však ještě mluví o fyzických pixelech ovládaných přímo hardwarem, které jsou samozřejmě závislé na konkrétním výstupním zařízení. Fyzické pixely kontrolované přímo hardwarem mohou vypadat např. jako na Obr. 3.



Obr. 1: Fyzický pixel

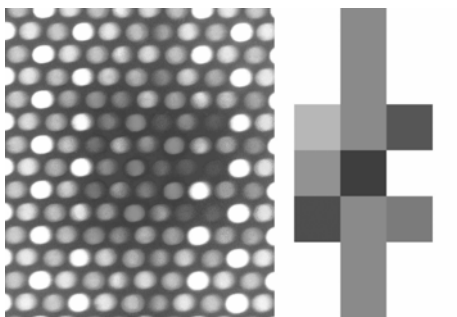


Obr. 2: Logický pixel



Obr. 3: Fyzické pixely tiskárny Canon

Fyzické pixely chápané jako čtverce jsou nezávislé na výstupním zařízení, a tedy nemohou být přímo kontrolovány hardwarem. Je tedy zřejmé, že nejde o fyzické pixely v pravém slova smyslu, ale pouze o jejich matematické modely (viz. Obr. 4).



Obr. 4: Fyzické pixely a jejich matematické modely

Pojem „fyzický pixel“ (chápan jako šachovnice) je velmi rozšířen a často používán. Proto se přizpůsobíme této běžné terminologii. Jde tedy o pixely, které jsou nezávislé na výstupním zařízení, protože jde o matematické modely.

Právě těchto rozdílů mezi chápáním pojmu pixel, lze s výhodou využít pro vytvoření velmi efektivních grafických algoritmů.

2 Teoretická východiska

V tomto odstavci jsou uvedeny základní pojmy matematické teorie potřebné k vytvoření již zmíněných rychlých grafických algoritmů.

Definice (Multiindex): Necht' $I = \{0, 1, \dots, i_k, \dots, m_k\}, k = 1, 2, \dots, n$ jsou indexové množiny. Pak množinu $I^{(n)} = \prod_{k=1}^n I$ nazýváme multiindexem.

Definice (nosič prostoru): Necht' $J = \langle a_k, b_k \rangle, k = 1, 2, \dots, n$ jsou intervaly. Množinu $J^{(n)} = \prod_{k=1}^n J$ nazýváme n-rozměrným nosičem digitálního prostoru.

Poznámka: V dalším textu bude $k = \{1, 2, 3, \dots, n\}$.

Definice (Multidělení): Necht' ${}_k D = \{ {}_k x_0, {}_k x_1, \dots, {}_k x_{i_k}, \dots, {}_k x_{m_k} \}$ jsou ekvidistantní dělení intervalů ${}_k J = \langle a_k, b_k \rangle, k = 1, 2, \dots, n$. Množinu $D^{(n)} = \prod_{k=1}^n {}_k D$ nazýváme ekvidistantním multidělením nosiče $J^{(n)}$.

Definice (digitální prostor, rozlišení): Necht' $J^{(n)}$ je nosič digitálního prostoru, $D^{(n)}$ jeho ekvidistantní multidělení. Uspořádanou dvojici $\mathbf{D}^{(n)} = (J^{(n)}, D^{(n)})$ nazýváme n -rozměrným digitálním prostorem. Uspořádanou n -tici $\mathbf{r} = (m_1, \dots, m_n)$ nazýváme rozlišením digitálního prostoru.

Definice (fyzická doména): Podmnožinu $F^{(n)} \subset J^{(n)}$ nosiče $J^{(n)}$ digitálního prostoru $\mathbf{D}^{(n)} = (J^{(n)}, D^{(n)})$ nazýváme fyzickou n -D doménou právě tehdy když

$$F^{(n)} = \langle {}_1 x_{i_1}, {}_1 x_{i_1+1} \rangle \times \langle {}_2 x_{i_2}, {}_2 x_{i_2+1} \rangle \times \dots \times \langle {}_k x_{i_k}, {}_k x_{i_k+1} \rangle \times \dots \times \langle {}_n x_{i_n}, {}_n x_{i_n+1} \rangle = \prod_{k=1}^n \langle {}_k x_{i_k}, {}_k x_{i_k+1} \rangle.$$

$$\text{Zapisujeme } \prod_{k=1}^n \langle {}_k x_{i_k}, {}_k x_{i_k+1} \rangle = F_{[i_1, \dots, i_k, \dots, i_n]}^{(n)} = F_i^{(n)}.$$

Poznámka: V odborné literatuře se velmi často používá pojem „pixel“ jako „nejmenší obrazový element zobrazitelný na daném výstupním zařízení“. „Nejmenší objemový element“ se často nazývá „voxel“. Tyto dva elementy jsou speciálními případy fyzických domén, pixel je fyzickou 2-D doménou a voxel je fyzickou 3-D doménou.

Dále je nutné definovat pojmy fyzický prostor, logický prostor, logický pixel, mapování a další. Tyto pojmy lze nalézt např. v [1] popř. [2].

3 Technologie .Net framework

Technologie zvaná .Net framework od společnosti Microsoft je tvořena běhovým prostředím (common language runtime - CLR) a základní knihovnou tříd (basic class library - BCL). Toto prostředí a tyto knihovny umožňují vývojáři vytvářet software jednodušeji a rychleji. Velkou výhodou užití této technologie je možnost použití libovolného podporovaného programovacího jazyka. Podporovány jsou jazyky Visual Basic, C++, C#, J#, Fortran a mnoho dalších.

Pro spuštění programu napsaného pomocí BCL je nutné na klientském počítači nainstalovat běhové prostředí CLR. Toto prostředí je primárně určeno pod systém Windows, ale společnost Microsoft velmi podporuje projekt Mono, který má za cíl umožnit běh .Net Frameworku v prostředí systému Linux.

Vzhledem k velkým výhodám využití této technologie zejména pro vytváření webových aplikací je zvolen příklad, který je právě napsán pro .Net framework.

4 Spolupráce se systémem Maple

Systém Maple je systém určený pro matematické výpočty a simulace. Velmi často je využíván v inženýrské praxi pro své velmi široké možnosti výpočtů a zejména pro simulace. Aktuální verze již obsahuje pokročilé nástroje pro modelování velmi složitých situací, bohužel některé grafické výstupy nejsou zcela správné a dostatečně rychlé.

Systém Maple poskytuje API rozhraní pro externí spolupráci. Toto API je označováno jako OpenMaple (více naleznete např. v [5]) a je poskytováno pro několik programovacích jazyků, např. C, Java, Visual Basic, Fortran. V následujícím odstavci je uveden příklad využití systému Maple pomocí webových technologií, který je však omezen licencí použití OpenMaple od výrobce systému Maple. Příklad je postaven na technologii .Net framework, o které jsme se zmínili v předchozím odstavci.

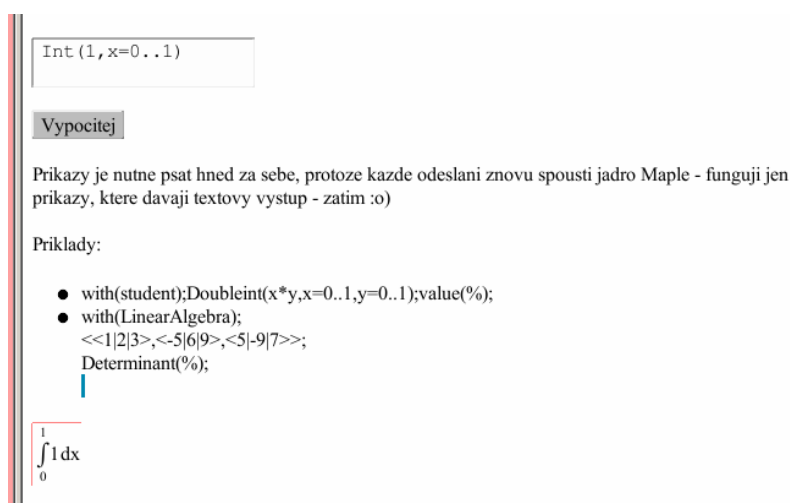
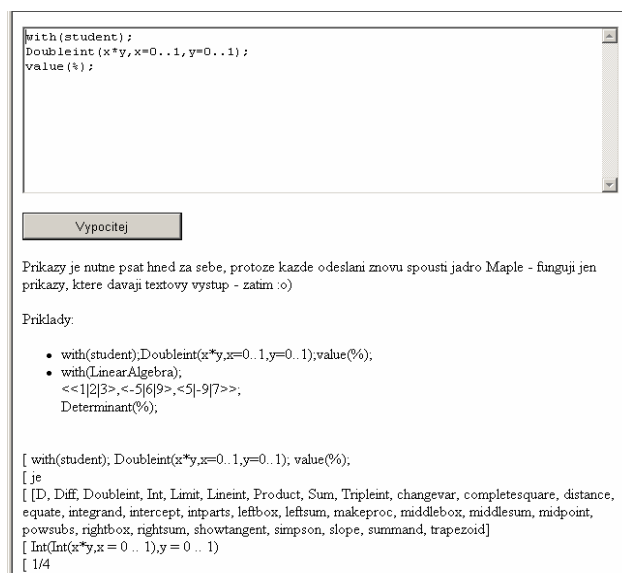
Na Obrázku 5 vlevo je vidět výstup pro webový prohlížeč, který podporuje technologii MathML, vpravo je prohlížeč, který ji nepodporuje. Neumí tedy zobrazit matematické symboly jinak než jako prostý text.

Z výše uvedeného je zřejmé, že je nutné vytvořit dynamickou knihovnu, která by zpřístupňovala technologii OpenMaple jakémukoliv jazyku podporovaného technologií .Net framework. V případě vytvoření této knihovny se stává OpenMaple tedy systém Maple přístupný z jakékoliv aplikace napsané pro CLR, protože zde nezáleží na zvoleném programovacím jazyku.

5 Zobrazování funkcí daných implicitně

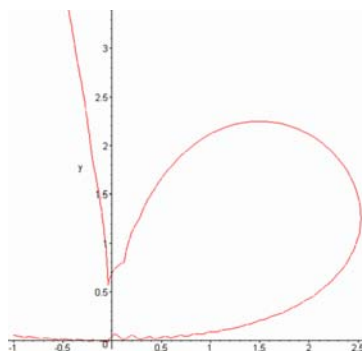
Jak již název příspěvku napovídá, zabývá se I rychlými grafickými algoritmy. Tyto algoritmy jsou postaveny na teorii, která je velmi zkráceně uvedena výše. Tato teorii umožňuje vytvořit velmi efektivní a rychlé algoritmy pro zobrazované křivky a ploch (více viz. [3]).

Zaměřme se na zobrazení funkcí, které jsou zadány rovnicí $f(x, y) = 0$. Tyto funkce se obvykle nazývají funkce zadané implicitně. Grafická konstrukce těchto křivek je poměrně obtížná, ale je jasné, že tyto křivky lze v podobných systémech jako je Maple zobrazit.



Obr. 5: Příklad webové aplikace užívající technologie OpenMaple a Net framework

Zobrazme tedy funkci $\left(\frac{x}{2} + \frac{y}{3}\right)^3 - x^2y = 0$ nejprve v systému Maple pomocí příkazu `plot` bez jakéhokoliv parametru. Výstup bude jako na Obrázku 6.



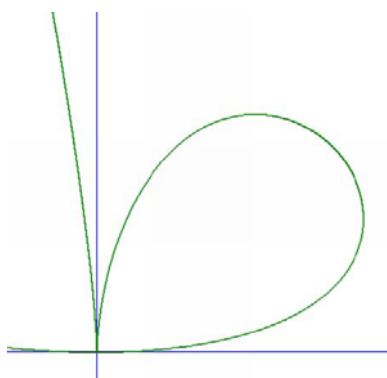
Obr. 6: Křivka o rovnici $(x/2 + y/3)^4 - x^2y = 0$ (špatně)

Již na první pohled je zřejmé, že tato křivka asi nebude správná. Po přidání parametru `numpoints` a změně jeho hodnoty z 50 na 36670 dostaneme konečně správný graf dané funkce. Tento obrázek je totožný s výstupem z vlastního programu.

V čem je tedy zásadní problém? Jde o to, že uživatel nevím, jakou má zadat hodnotu parametru `numpoints`, protože jak rozhodne, že již je graf správný. Dále je zde problém s rychlostí vykreslování. Každá změna tohoto parametru znamená velký nárůst času potřebného ke konstrukci této křivky.

Vytvořil jsem program, který implementoval nejpřesnější možný a rychlý algoritmus pro vykreslování implicitně zadaných křivek. Toto řešení dává výsledky do jedné sekundy, zatímco správné vykreslení křivky v Maple trvá asi osmkrát déle.

Je tedy zřejmé, že vytvoření speciálního programu je někdy velmi užitečné, ale práce v tomto programu je limitována právě jeho zaměřením. Pomocí technologie OpenMaple však je k dispozici celá výpočetní síla takového systému jako je Maple. Jde tedy o velmi silné řešení daného problému, protože lze nejen volat externí aplikaci pro vykreslení implicitně zadané křivky přímo z Maple, ale i naopak.



Obr. 7: Křivka o rovnici $(x/2 + y/3)^4 - x^2y = 0$ (správně)

6 Závěr

Rychlé grafické algoritmy jsou základem jakékoliv dnešní práce na počítači, neboť i běžní uživatelé dnes denně pracují s různými programy pro zobrazování obrazových dat a vlastně jakékoliv grafické prostředí je postaveno na rychlých grafických algoritmech. Výše byl uveden příklad propojení webové aplikace se systémem Maple, který může sloužit jako základ pro propojení jakéhokoliv grafického systému, který by nahradil některé velmi pomalé a nepřesné funkce pro zobrazování v systému Maple. Dále byl uveden příklad rychlého algoritmu pro vykreslení implicitně zadané funkce a jeho srovnání

s vykreslením téže funkce v systému Maple. Z toho lze vidět, že specializované programy mohou být daleko lepší a externí spojení se systémem Maple může být výhodné nejen pro práci v Maple, ale i mimo něj, a to voláním jeho funkcí.

7 Literatura

- [1] Žák, V.: Efektivní grafické algoritmy. 2006. Pojednání k disertační práci.
- [2] Martišek, D.: Matematické principy grafických systémů, Littera, Brno, 2002.
- [3] Žára, J., Beneš, B., Felkel P., Sochor., J.: Moderní počítačová grafika, Computer Press, Brno, 2004.
- [4] Webová prezentace firmy Maplesoft, www.maplesoft.cz, 2006.
- [5] Webová prezentace firmy Microsoft, www.microsoft.com, 2006.

Sborník příspěvků 3. letní školy aplikované informatiky

Bedřichov, 25. – 27. srpna 2006

Divize environmentální informatiky a modelování

Institut biostatistiky a analýz

Masarykova univerzita

Editoři

prof. RNDr. Jiří Hřebíček, CSc.

RNDr. Jaroslav Ráček, Ph.D.

Vydala Masarykova univerzita, Brno, 2006

Tisk **XXXXXX**

ISBN: **XXXXXXXXXX**